THE EFFECT OF TEACHER-TEAM GENERATED FORMATIVE ASSESSMENTS ON STUDENT WRITING ACHIEVEMENT TEST LEVELS

by

Justin Calhoun

Liberty University

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Liberty University
2020
THE EFFECT OF TEACHER-TEAM GENERATED FORMATIVE ASSESSMENTS ON STUDENT WRITING ACHIEVEMENT TEST LEVELS

by Justin Calhoun

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Liberty University, Lynchburg, VA

2020

APPROVED BY:

Lisa Reason, Ph.D., Committee Chair

Jonathon Fields, Ed.D., Committee Member

Glenn Holzman, Ph.D., Committee Member
ABSTRACT

Formative assessment has been studied for the last 2 decades. Increased high-stakes accountability requirements have intensified the focus on preparing students to demonstrate increased achievement on end-of-course summative tests. Districts have turned to common formative assessments as a means to increase student achievement on end-of-course testing. Therefore, more study was needed to determine if common formative assessments lead to increased students’ achievement on summative writing assessments. This causal-comparative study examined two groups of eleventh-grade students to determine if any differences existed between student achievement scores on teacher-created formative assessments and common formative assessments created by teacher teams. Data were gathered from eight high schools that represented the entire population of Northeast Tennessee. A two-way analysis of variance (ANOVA) was used to analyze data, and findings indicated that statistically significant differences existed when common formative assessments were administered to eleventh-grade students.

Keywords: Formative assessments, common formative assessment, achievement scores, end-of-course assessment.
Dedication

I would like to dedicate this dissertation to my family. My wife Amy and my son Hudson are truly the reason why I pursued this degree. I will be forever grateful for their support in undertaking this work.

Because of Jesus Christ my life has been changed. My prayer every day is that I would be the leader He instructs me to be through His word. I want Him to be glorified and exalted in all I do. I am thankful for the help He provided during the work contained in this study.
Acknowledgments

I am grateful for my committee for their support and wisdom during this process. I am especially thankful for meeting Dr. Lisa Reason, my chair, who has been a calming presence and great ally in this education pursuit.

Jonathon Fields, I am thankful you served on this committee, but I am even more thankful for your friendship. I have always been able to count on you for encouragement, insight, and support. Good friends are hard to find. I am thankful I found you.

I would also like to acknowledge my parents who instilled a strong work ethic and provided me with the skills necessary to achieve this goal. My mother who taught me the value of education. My dad would taught me the value of hard work. I will be forever grateful for your support and wisdom.
# Table of Contents

ABSTRACT .............................................................................................................. 3

Dedication ............................................................................................................... 4

List of Tables ......................................................................................................... 9

List of Figures .......................................................................................................10

List of Abbreviations ............................................................................................11

CHAPTER ONE: INTRODUCTION ..............................................................................12

  Overview ........................................................................................................... 12

  Background ....................................................................................................... 12

  Problem Statement ............................................................................................15

  Purpose Statement .............................................................................................16

  Significance of the Study ..................................................................................17

  Research Questions ..........................................................................................18

  Definitions .........................................................................................................18

CHAPTER TWO: LITERATURE REVIEW ..................................................................20

  Overview ........................................................................................................... 20

  Standards ..........................................................................................................20

  Conceptual or Theoretical Framework ............................................................25

  Related Literature .............................................................................................28

  Summary ............................................................................................................50

CHAPTER THREE: METHODS ..................................................................................52

  Overview ........................................................................................................... 52

  Design ................................................................................................................52
Research Question(s) ................................................................. 53
Hypotheses ........................................................................... 53
Participants and Setting . ......................................................... 55
Instrumentation ..................................................................... 56
Procedures ............................................................................... 60
Data Analysis .......................................................................... 61

CHAPTER FOUR: FINDINGS ....................................................... 62
Overview ................................................................................ 62
Research Questions .................................................................. 62
Null Hypotheses ..................................................................... 63
Descriptive Statistics ................................................................. 64
Results .................................................................................. 66
Analysis .................................................................................. 67
Summary ................................................................................ 76

CHAPTER FIVE: CONCLUSIONS .................................................. 77
Overview ................................................................................ 77
Discussion .............................................................................. 77
Implications ........................................................................... 82
Limitations ............................................................................. 84
Recommendations for Future Research ................................. 85
Summary ................................................................................ 86

REFERENCES ........................................................................... 87

APPENDIX A: INTERNAL REVIEW BOARD APPROVAL .......... 99
List of Tables

Table 1. Scores by Category of Assessment .................................................................65
Table 2. Gender Scores by Category of Assessment ......................................................66
Table 3. Levene’s Test of Equality of Error Variance by Category of Assessment ..........67
Table 4. Summary Table of Two-Way ANOVA: Category of Assessment Score ..........69
Table 5. Summary Table of Two-Way ANOVA: Type of Assessment by Gender ..........72
List of Figures

Figure 1. Bar graph with TNReady writing score averages by type of formative assessment

                          .................................................................................................................. 70

Figure 2. Box plot with of the development category of TNready scores by gender. ............. 73
List of Abbreviations

Race to The Top (RTTT)

Common Core State Standards (CCSS)

National Assessment of Educational Progress (NAEP)

Tennessee Comprehensive Assessment Program (TCAP)

Every Student Succeeds Act (ESSA)

No Child Left Behind (NCLB)

English Language Arts (ELA)
CHAPTER ONE: INTRODUCTION

Overview

Public education students are accustomed to year-end achievement testing. The data derived from these assessments serve as accountability measures for local districts and provide the basis for district decisions (Black and Wiliam, 2009). In the state of Tennessee, students in grades 3–12 are required to take end-of-course achievement tests in all content areas. It has been recommended that districts incorporate frequent formative assessments into their curriculums to adequately prepare students for these summative assessments (Tennessee Department of Education, 2015). This research study investigated whether common formative writing assessments can accurately predict student academic achievement on a standardized criterion-referenced test. This first chapter of this dissertation discusses the study.

Background

The No Child Left Behind (NCLB) Act of 2001 was reauthorized in 2013. This reauthorization of the Every Student Succeeds Act (ESSA) has instituted accountability measures for public schools. Local education agencies are required to demonstrate achievement across student populations. Local education agencies are held accountable for all student populations and are expected to meet state accountability measures. States are mandated to include criterion-referenced achievement tests to measure these goals as they are related to high school English scores. Student achievement data from these tests are included to determine if schools met, exceeded, or fell below their pre-determined goals for achievement (Darling-Hammond et al., 2016). While the ESSA (2013) reauthorization includes a focus on “fewer, smarter assessments” it has also encouraged states to develop formative assessments as a source
of student data (ESSA, 2013; Shepard, Penuel, & Davidson, 2013). Reducing testing time and including a new focus on writing expression requires local agencies to continue to meet high standards of achievement while testing students less often, in addition to including writing as a measurement.

The reauthorization of ESSA in 2013 was accompanied by the Common Core State Standards (CCSS). These new standards—created through the partnership of the Governors Association and the Council of Chief State School Officers—laid out a more rigorous pathway to college and career readiness. Forty-six states initially agreed to the standards in 2010, and the standards were fully implemented by the 2014–2015 school year. In the CCSS, writing has been moved from an instructional tool to a standard of learning. Writing is addressed in multiple standards, content areas, and has subsequently been included on criterion-referenced end-of-year testing. The writing standards address multiple modes of writing, outline higher expectations for student production of writing, and specifically address what writing looks like across content areas. The focus has been shifted to include task, audience, and purpose as opposed to the mere process of writing (CCSS Initiative, 2010). ESSA and CCSS have been combined to increase the college and career readiness of K–12 students. American College Testing (ACT) reports, colleges, and universities have indicated that incoming freshman are not adequately prepared for the rigorous demands of collegiate-level writing (American College Test, 2015). The CCSS aimed to address the writing demands required in colleges and careers to address this influx of remedial students.

The introduction of the CCSS and the shift it requires towards equipping graduates to be college- and career-ready has led to an increased reliance on formative assessment. Schools and districts are required to make decisions regarding teacher effectiveness and student achievement
based on summative achievement data that are gathered at the end of the course (Darling-Hammond et al., 2016). Using a single assessment’s data for multiple purposes is a difficult challenge (Popham et al., 2014); teachers cannot adjust instruction for current students based on the previous cohort’s test scores. The CCSS require schools to test often enough to obtain reliable data that adequately supports decision-making (Shute & Kim, 2014). Both states and local education agencies are now focusing on formative assessments due to the increased demands of standards and reduced amount of testing (Tennessee Department of Education, 2015). Summative end-of-course testing does not provide data demonstrating students’ trends of learning (Filsecker & Kerres, 2012). Teachers have begun to place high value on formative assessment data that equips teachers with the data needed to adjust instruction and maintain high achievement scores (Wormeli, 2017). Guidelines and requirements for state summative testing have been provided by states, but detailed support for developing effective formative assessments have been widely left to local education agencies to create. The focus on formative assessments has been influenced by the increased reliance on summative writing assessments. Summative assessments continue to remain “one of the least understood constructs in assessment” (Olinghouse, Zheng, & Morlock, 2012, p. 98).

The State of Tennessee’s adoption of the CCSS has led to changes in testing requirements. Tennessee’s former criterion-referenced test, Tennessee Comprehensive Achievement Program (TCAP), has been replaced by TNReady. Students in grades 3–12 have taken the TNReady since the 2015–2016 school year (Tennessee Department of Education, 2015). All students in grades 9–12 now take a writing assessment based on multiple modes (informational and expository, opinion and argument, and narrative). In previous years, students’ writing was only tested on a pre-determined mode of writing; conversely, TNReady requires
multiple modes of writing that are not predetermined. Students respond to tasks that require them to demonstrate their skills in expository, argumentative, or narrative writing modes. The State of Tennessee has increased its focus on formative assessments to prepare for the new testing platform. However, the frequency and variety of formative assessments vary across the state. The Tennessee Task Force on Student Testing and Assessment recommended “providing expectations to districts regarding formative assessment usage and communication” (Tennessee Department of Education, 2015, p. 5) to develop a more uniform set of formative assessment guidelines. Furthermore, the task force recommended that formative assessments be locally created and encouraged teachers to create their own formative assessments; teacher-created formative assessments are typically more aligned to Tennessee state standards than vendor-provided assessments. Through creating common formative assessments, teachers will be provided with vital data that will guide daily lesson planning and address students’ needs.

**Problem Statement**

Summative achievement testing is a challenge facing teachers, administrators, and public school supervisors. Increased accountability and decreased testing opportunities also create challenges for students. American public schools place a high value on what is tested (Black et al., 2009). Educators have pushed formative assessment to the forefront in order to adequately prepare students for high-stakes achievement tests. Districts and schools have been left questioning their ability to create appropriately rigorous formative assessments to gauge student needs. The decrease in summative, end-of-year assessments has led to an increase in formative assessments, which have become more valuable in gauging student needs (Black, 2009; Schneider, Egan, & Julian, 2013).
Many districts and schools have created their own common formative writing assessments that are used to adjust instruction to meet student needs. Other districts focus solely on instruction and trust the curriculum to be enough to meet the demands of the summative assessment. Teachers are seeking to obtain enough reliable data to make decisions and predictions concerning end-of-course testing. Prior studies have suggested that teachers should regularly incorporate formative assessment into their instruction; however, other studies have suggested that teachers do not use formative assessment effectively (Curry, Mwavita, Holter, & Harris, 2016; Scheider & Andrade, 2016). This misuse of formative assessment can be attributed to teachers’ lack of confidence on how to use assessment data to adjust instruction and increase student achievement (Andrade & Heritage, 2017). The use of formative assessment has mainly been limited to traditional avenues of instruction. Teachers and districts often overlook various approaches for how to impact students’ end of course scores using formative assessments (Stull, Varnum, Ducette, & Schiller, 2011). This study examined whether common formative assessments—which are defined as assessments created by grade or course-level teams of educators that measure student success criteria in focus within a curricular unit of study—were accurate predictors of student achievement on end-of-course writing assessments (Ainsworth & Viegut, 2014). Limited research existed regarding how common formative assessments affect student achievement as measured by end-of-course tests (Schneider et al., 2013).

**Purpose Statement**

The purpose of this study was to examine the differences in TNReady writing scores among eleventh-grade students who took team-created common formative writing assessments at least twice a semester and students who took teacher-created formative assessments. This casual-comparative or ex post facto design worked well within the school setting, as neither of the
classes nor variables could have been organized into experimental groups during the school year (Ary, Jacobs, Razavieh, & Sorenson, 2006; Gall, Gall, & Borg, 2007). The common formative writing assessments administered to both groups was a district curricular decision; therefore, an experimental study could not be conducted. The study utilized end-of-year TNReady writing scores for the classes from two groups of eleventh-grade students.

**Significance of the Study**

All public school students in grades 3–12 are required to take end-of-course achievement tests in the state of Tennessee. These exams—referred to as TNReady tests—are aligned with the Tennessee State Academic Standards which align with the CCSS. Students in grades 3–12 are required to write in a response to a task as part of the TNReady assessment. The scores from TNReady tests are used as 35% of teachers’ overall evaluation score and 15% of the students’ final grade.

Eight high schools from three separate school systems were included in this study. These eight high schools developed a common formative assessment system to measure students’ writing ability prior to end-of-course testing. The common formative assessment is given in grades 9–12 at multiple points in each semester. The writing tasks are crafted in the style of practice writing tasks released by the Tennessee State Department of Education. Teachers administered the formative assessments at predetermined dates. The students’ samples were scored using the same rubric and scoring guidelines as the end-of-course assessment. The findings in this study may assist schools and districts in making more informed decisions regarding common formative assessments. The purpose of formative assessment is to provide teachers and schools with evidence that students are moving toward the learning targets prescribed by state standards. Ainsworth and Viegut (2014) termed these learning targets,
“sequential building blocks of instruction necessary for students to understand the larger learning intentions of the unit” (p. 15). Summative assessments report student achievement regarding state standards to stakeholders (Bonner, 2013). Formative assessment has varying effects based on grade level, frequency of assessment, and use of data. This study contributed to the literature by providing additional data on the use of common formative writing assessments with high school English students.

**Research Questions**

**RQ1:** What difference exists (if any) between writing achievement test (category) scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade students who took two formative assessments created by individual teachers?

**RQ2:** What difference exists (if any) between writing achievement test (category) scores of eleventh-grade male students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade female students who took two formative writing assessments created by teacher teams (common formative assessments)?

**Definitions**


2. *Formative assessment* - Formative assessments are conducted in the classroom and provide assessment data that informs instructional decisions and provides feedback opportunities (Stull et al., 2011).
3. **Summative assessment** - Summative assessment is the state criterion-referenced achievement test given at the end of the school year to determine student growth and achievement (Black et al., 2012).

4. **Common formative assessment** - Common formative assessments are collaboratively designed by a grade- or course-level team of educators to assess students’ understanding of the particular learning intentions and student success criteria currently in focus within a curricular unit of study (Ainsworth & Viegut, 2014).
CHAPTER TWO: LITERATURE REVIEW

Overview

Today’s schools assess what is valued and writing is at the forefront of that testing. In spite of the statistics on the writing deficiencies of today’s students, writing is still widely believed to be the most neglected of the “Three R’s” of education: reading, writing and arithmetic (Harris, Graham, Friedlander, & Laud, 2013; Lacina & Block, 2012; Scheiber, Reynolds, Hajovsky, & Kaufman, 2015). Federal guidelines for testing and state initiatives place a high value on data produced from high-stakes assessments. Federal, state, and local decisions are impacted by the use of student data from these yearly assessments. Most states, including Tennessee, require on-demand writing assessments for their students and base their decisions on these summative scores (Brookhart, 2013). A majority of studies have focused on writing in the early grades, but few have focused on writing at the upper-secondary and secondary level (Lacina & Block, 2012). In their meta-analysis of writing, Graham and Herbert (2011) found that studies examining writing have steadily declined over the last 3 decades. Graham and Herbert observed that writing research has decreased 41% in the 1980s, to 24% in the 1990s, to 15% thus far in the 21st century. Few studies have examined the effect of common formative assessments on summative writing achievement tests.

Standards

Communication is paramount for students. Writing, as the key foundation of communication, affects students’ ability to succeed in K–12 education and the workforce (Butler & Britt, 2011). The importance of writing extends beyond the classroom and into other areas of students’ lives, including psychosocially and financially (Troia & Olinghouse, 2013).
The National Assessment of Educational Progress (NAEP; 2011) characterizes writing as, “A complex, multifaceted and purposeful act of communication that is accomplished in a variety of environments, under various constraints of time, and with a variety of language resources” (p. 17). Writing is a complex cognitive activity that requires formalized style (Harris et al., 2013). Additionally, writing is positively connected to reading comprehension, grade point averages, and college success (ACT, 2018; Graham & Herbert, 2011). Students who do not write at a proficient level may not score well in content area classes. ACT’s College Readiness Report indicated that 86% of high school graduates planned to obtain a postsecondary education; however, 35% did not meet any of the test’s four benchmarks (English, reading, math, science reasoning). Only 38% met three or four benchmarks and 60% met the English benchmark (American College Testing, 2015). NAEP (2011) reported that 54% of America’s students scored basic in writing. College students who struggle to write also struggle to graduate.

Employers expect that employees can complete a certain level of writing in both blue- and white-collar jobs (Butler & Britt, 2011; Graham & Herbert, 2011; Graham et al., 2011; Olinghouse et al., 2012). Up to 80% of industries require effective written communication as part of their job requirements; 90% of white-collar jobs and 80% of blue-collar jobs include writing as a key component of job success (Graham et al., 2011; NAEP, 2011). Students must be able to communicate through writing to be productive members of the community; social media and emails have become primary means of communication in the 21st century. Therefore, academic writing has become an integral part of education (Olinghouse et al., 2012).

The NCLB Act was signed into law in 2002. The NCLB Act sought to identify why the nation’s students were failing and attempted to remedy underachievement issues. The NCLB Act attempted to diagnose and correct the problem through standardized achievement testing in
grades K–12 (Jorgensen & Hoffmann, 2003). Thirteen years later, President Obama signed the ESSA, which promised “fewer, better assessments” which would lift the burden of “ineffective testing” (Executive Department of the President, 2015, p. 4) off state and local decision-makers. While having fewer and more meaningful assessments, ESSA aimed to “preserve annual assessments” and help “empower state and local decision-makers to develop their own systems of school improvement based on evidence” (Executive Department of the President, 2015, p. 1). The topic of writing was specifically mentioned in the report, and the report stated that assessments would include fewer multiple-choice items and more authentic assessment measures.

To help states fulfill the promise of empowerment in school-based decision-making, the government initiated Race to the Top (RTTT), a federal grant program aimed at improving education through improvements in assessments, programs, and curricula (Executive Department of the President, 2015). CCSS were instrumental in RTTT’s popularity among states. The goal of CCSS was to raise student achievement scores through implementing more rigorous standards (Troia & Olinghouse, 2013). The National Governors Association and the Council of Chief State School Officers created the CCSS in 2010. Forty-six states initially accepted and implemented those standards, and these states had fully implemented the CCSS into instruction by 2014–2015 (Applebee, 2013; Harris et al., 2013). In specific reference to writing, the CCSS initiative (2010) states:

To be college and career ready writers, students must take task, purpose, and audience into careful consideration, choosing words, information, structures, and formats deliberately. They need to know how to combine elements of different kinds of writing—for examples, to use narrative strategies within argument and explanation within
narrative- to produce complex and nuanced writing. They have to become adept at gathering information, evaluating sources, and citing material accurately, reporting findings from their research and analysis of sources in a clear and cogent manner. They much have the flexibility, concentration, and fluency to produce high-quality first draft text under a tight deadline as well as the capacity to revisit and make improvements to a piece of writing over multiple drafts when circumstances encourage or require it. (p. 41)

Standards provide a point of reference and a goal to achieve, but as Troia and Olinghouse (2013) observed, “academic standards do not specify the ‘how’ of instruction, but the ‘what’” (p. 348). The CCSS sought to be specific about the demands of writing to help inform the instruction needed for students to achieve the standards through assessment. The CCSS require a deeper level of student understanding, are more vertically aligned across grade bands, and provide a more specific focus (Bennett, 2015). Gillespie, Graham, Kiuhara, and Herbert (2013) conducted a study of 211 randomly selected high school teachers from various content areas to investigate teachers’ use of writing in the classroom. Participating teachers were asked what type of writing they used in their classrooms, how often they engaged in those activities, and in what context the activities occurred. Eighty-two percent of participating teachers reported regularly using writing in the classroom and 53% of teachers explicitly taught writing strategy and skills. However, when reporting their findings, Gillespie et al. (2013) identified the discrepancy between students’ writing more frequently and students’ ability to write effectively as defined by the CCSS: “Findings from the current study make it clear that for many of the teachers participating in this study there is a considerable distance between what they do in their classrooms and the expectations embodied in the CCSS” (p. 1071).
Due to the promise of fewer assessments in the ESSA and the increased focus on rigor in writing, the CCSS encouraged the use of formative assessments in K–12 classrooms to increase the value of the teaching and learning process (Popham et al., 2014). RTTT gave credence to formative assessment and recognized that formative assessment was a legitimate educational resource to aid student learning (Herman, Osmundson, Dai, Ringstaff, & Timms, 2015); however, the CCSS do not explicitly dictate how an assessment should be designed or what principles should influence assessment design (Bennett, 2015). More study was needed to determine which assessments enable students to meet the CCSS. When used as the main source of feedback in writing, summative, year-end, or semester assessments fail to adequately prepare students for their post-secondary endeavors and hinder the instructional process of teaching writing. Instruction ceases and students do not have the opportunity to reflect on their scores after summative assessments have been administered (Wormeli, 2017). Summative assessments cannot be the sole means of decision-making for districts and classroom teachers as state and federal guidelines shift away from multiple-choice tests to more authentic assessments of student writing ability. Despite the instatement of the NCLB Act, 72% of students are not writing more effectively than previous years (Lacina & Block, 2012).

Most states require some type of writing assessment from their students in grades 3–12; these assessments are used as accountability measures. States officials require high-stakes testing to create reports to stakeholders (Brookhart, 2013). Only one writing sample from each student is required in most cases. Teachers who teach in states that require writing assessments tend to only teach the genre of writing required for summative assessment (Brimi, 2012). Writing to only one specific genre encourages formulaic writing practices that impede students’ writing development (Cumming, 2013). Troia and Olinghouse (2013) described this practice as placing “instructional
emphasis” (p. 346) on the genre of writing that will be assessed during the summative writing assessment. Teaching and testing only one genre of writing greatly inhibits students’ ability to grow as writers.

Students’ writing careers do not stop at high school graduation. The expectations of communicating through written expression will more than likely lie outside one genre of writing (Washington, 2014). Frequent common formative assessments provide ample opportunities for teachers to provide feedback to students and require students to review their own writing. The feedback cycle between teacher and student reflects collaboration in a professional work environment, thus helping students be more prepared for a career after graduation (Pereira, Flores, & Niklasson, 2016). Lowering the number of summative assessments required of districts creates a vacuum in terms of assessments. Formative assessments must become more prevalent as evaluators of student achievement to reach more students and help close the achievement gap (Li, 2016). Increasing the frequency of formative writing assessments and requiring students to write in multiple modes has been proven to be statistically more effective than focusing on only one mode of writing (Brimi, 2012).

**Conceptual or Theoretical Framework**

Bandura’s (1989) social cognitive theory of academic self-regulation supports frequent formative assessment use in the classroom as a predictor of student achievement. Zimmerman and Bandura (1994) noted “in social cognitive theory of academic self-regulation students regulate the motivational, affective and social determinants of their intellectual functioning as well as the cognitive aspects” (p. 846). Bandura observed that students benefited from “modeled expertise” and “modeling helps to focus the attention on key linguistic features which promote knowledge acquisition” (p. 13). Zimmerman and Bandura also noted that modeling instructional
writing strategies and basing instruction on observed deficiencies helped improve student writing outcomes. Student cognitive development occurs at a rapid rate and requires support from teachers to ensure this influx of knowledge creates skills in the students. Teachers can provide specific and timely feedback when assessments are frequently administered to students.

Writing assignments are typically conducted independently and graded results and feedback are often provided at a much later date. Zimmerman and Bandura (1994) found that “instruction in writing strategies has been shown to enhance perceived self-efficacy and to improve the schematic structure and quality of composition” (p. 846). Assessments inform the curriculum and instruction of teachers. Reliance on summative assessments can limit the appropriate development of student writers. Students who receive frequent feedback are provided with the opportunity to analyze their own work. Students’ self-assessment improves self-efficacy and allows students to own their learning, which can lead to increased student achievement (Wylie & Lyon, 2015).

Camp (2012) observed that limiting assessment to a single test or mode of writing failed to “represent the multidimensional nature of writing growth” (p. 93). When applying the stage models of development, assessing multiple samples of student writing in multiple modes can provide evidence of students’ writing maturity. Relying on only one piece of writing at the end of the course cannot adequately provide evidence for diagnosis of writing deficiencies or exhibit student cognitive development. Diagnosis of student writing “looking through present performance to the underlying developmental forces that will bring about a different future performance” (p. 96) can provide teachers with enough evidence to provide feedback and direction to students as they develop as writers. Thus, formative assessments are much more aligned to the social cognitive theory and may increase students’ self-regulation during the
writing process and provide more opportunities for teachers to observe developmentally appropriate growth in student writing. Self-regulation, in addition to planning and attention to task, are essential skills of successful writers (Kim, Al Otaiba, Wanzek, & Gatlin, 2015).

Elliot, Rupp, and Williamson (2015) noted that social cognitive models provide “high quality, specific information about both the writing construct and its boundaries” (p. 2). Formative assessments also provide frequent opportunities for feedback in multiple modes of writing; frequent feedback has been found to promote growth in students’ cognitive development (Bandura, 1989). Formative writing assessments have the potential to stimulate a cycle of growth in students through frequent use and may promote self-motivation in students. Writing is a complex skill and requires a great deal of self-regulation (Harris et al., 2013).

The motivational aspect of writing directly correlates with Bandura’s (1989) social cognitive theory. The writing process has five stages: (a) prewriting, (b) writing, (c) revision, (d) editing, and (e) publication. Students can become frustrated while going through the writing process and may stop writing altogether (Graham et al., 2013). Lower achieving students need more frequent experience with writing tasks to raise their cognitive level in terms of writing skill development. Summative assessments do not provide these opportunities; rather, summative assessments merely report what skills the student failed to gain. It is more likely for summative assessments to be linked to students’ level of achievement and comparisons made amongst groups or individual students. Students—especially lower performing students—are more likely to grow from self-comparative analysis than social-comparative (Bandura, 1989). Lower performing students are more likely to benefit from frequent common formative assessments with individualized feedback from teachers at multiple points throughout the course. Common formative assessments provide students with frequent feedback, which enables students to
monitor their own learning. Additionally, common formative assessments also enable the teacher to adjust instruction as identified by the class trends (Black & Wiliam, 2009).

Evidence-centered design (ECD) is a conceptual framework that helps define how formative writing assessments work within social cognitive theory. Social cognitive theory is the why and ECD is the how. Common formative assessments that follow a predetermined systematic approach for task creation provide more validity. ECD is described as a “principled framework for designing, producing, and delivering educational assessments and ensures that the way in which evidence is gathered and interpreted bears on the underlying knowledge and purposes the assessment is intended to address” ((Zieky 2014, p. 79). This framework assists school-based assessment teams as they design, deliver, and interpret curriculum-embedded classroom assessments (Elliot et al., 2015). ECD details a 4-process model for assessment delivery: (1) activity selection, (2) presentation, (3) response processing, and (4) summary scoring (Zieky, 2014). These 4-process model steps in ECD complement the elements of effective writing instruction (Graham et al, 2013); effective instruction must be administered through an effective design. ECD considers effective writing instructional assessments that are conducted at multiple times throughout the writing course for multiple genres of writing. This design will provide a more accurate evaluation of students than relying on summative assessments alone.

**Related Literature**

**Effective Writing Instruction**

The development of proficient writers takes place at the K–12 level. Teachers’ awareness of what effective writing instruction entails is necessary to developing proficient writers. Students are expected to be proficient at written communication during their K–12 education, in
college, and in their career. State and national standards have explicitly detailed writing expectations and set high standards for the production of writing, such as citing textual evidence to support responses to both literary and informational texts (CCSS Initiative, 2010). Therefore, colleges have the expectation that students will demonstrate writing proficiency in multiple genres of writing. Troia and Olinghouse (2013) studied writing in the Common Core classroom and their findings differed from the ideology in the CCSS. Similarly, Graham and Kiuhara (2012) found a clear discrepancy between the level of expectation in the CCSS standards and the quality of writing instruction. Graham and Kiuhara noted that a significant discrepancy existed between the writing instruction teachers reported and the expectations of the CCSS. The secondary classrooms in their national study spent less than 3 hours per semester in explicit writing instruction. Additionally, teachers frequently modified whole group instruction to meet the needs of the lower performing students in the class and a great percentage of teachers did not differentiate instruction for struggling writers (Troia & Olinghouse, 2013).

In 2017, students in Tennessee were required to take the TNReady in English I, II, and III. These English courses were taken by students in grade 9 (English I), 10 (English II), and 11 (English III). Students’ reading and writing were assessed and reported in student, school, and district summaries (Tennessee Department of Education, 2015). Writing and reading are two interrelated skill sets that are reciprocal in nature; thus, the skills are tested together on a single assessment in the state of Tennessee (Graham and Herbert, 2011; Tennessee Department of Education, 2015). It would be detrimental to ignore the relationship between reading and writing; this relationship must not be ignored by scholars. Students’ writing scores are declining and in need of intervention; thus fact has led to state and national initiatives to improve writing in the K–12 setting (NAEP, 2011; ACT, 2015; CCSS initiative, 2010).
CCSS standards have continued to clarify the relationship between writing and reading. This relationship between English language arts (ELA) and literacy in the content areas is important, as students “have to become adept at gathering information, evaluating sources, and citing material accurately, reporting findings from their research and analysis of sources in a clear and cogent manner” (CCSS initiative, 2010, p. 41). The language of this standard specifies that students should be writing about what they have been reading. Writing helps students process information and make the connection between text and knowledge; however, writing is often overlooked when teachers list reading comprehension strategies (Herbert, Simpson, & Graham, 2013).

Graham and Herbert (2011) noted that students who write about the material they read have an enhanced comprehension of text. Additionally, writing skills instruction strengthens students’ reading skills; increasing the amount that students write will improve how well students read. Previous literature focused on other influencers of reading comprehension, such as vocabulary, reading programs, and teaching strategies. Researchers focused solely of the improvement of reading comprehension through writing activities.

Teachers are incorporating writing into their instruction in response to the CCSS. Gillespie et al. (2013) studied 172 teachers in multiple content areas and found that 82% of participating teachers were incorporating writing into their classrooms; however, only 52% explicitly taught writing skills. Regardless of this gap in rigor, an increase in reading comprehension was observed through the mere act of writing. In Gillespie et al.’s previously mentioned meta-analysis, 45 of the time teachers engaged in explicit writing instruction; however, a positive correlation was still shown in reading comprehension. Gillespie et al.’s findings indicated that teachers who provide explicit writing instruction with a purpose could see
gains in student achievement data on high-stakes writing achievement tests (Graham & Herbert, 2011).

The level of writing instruction in secondary ELA classrooms directly correlates with students’ writing abilities. Students approach writing at various levels of proficiency; however, all students benefit from the scaffolding of writing instruction. Scaffolding is a teacher instructional practice that provides students with the learning supports necessary to produce the desired results (Mulatsih, 2011). The scaffolding framework of effective writing instruction leads to growth and achievement by developing writers regardless of their current proficiency level. The five stages of writing are generally referred to as prewriting, drafting, revision, editing, and publication (Tuttle, 2013). Novice—or basic performing—students may be hesitant to begin the writing process and may not reach the publication phase of the writing process without effective scaffolding. Students nearing the advanced level can also experience growth in their writing. Scaffolding stimulates growth and achievement in students because it provides students with a more complete framework of writing instruction. Writing is a complex cognitive skill that involves formalized style and requirements (Bayat, 2014; Bruning, Dempsey, Kauffman, Mckim, & Zumbrunn, 2013). Scaffolding can provide the intricacies necessary to help writers develop writing skills by continually defining audience, purpose, and context in students’ writing through explicit instruction. Scaffolding also provides teachers with the structure necessary to help students develop a deep understanding of the standards, including the complexity of writing as a system of communication (Mulatsih, 2011). Timely feedback provided by the teacher allows students to monitor their writing and provides the teacher with an opportunity to adjust student learning and instruction through providing feedback. Timely feedback leads to increased student writing performance (Tuttle, 2013) and gains in achievement are more likely to occur when
students are given the opportunity to apply feedback from the formative assessment. Teachers in effective writing classrooms use a nonlinear writing process (Benjamin, 2013). Formative assessment in writing classrooms can be described as “The continual growth cycle of monitoring, diagnosing, giving feedback, allowing time for growth, and affirming growth” (Tuttle, 2013, p. 3). Frequent and timely feedback allows students to become better writers by continually reviewing and revising their work.

Some students suffer from writing anxiety; however, frequent writing experiences help most students overcome writing anxiety (Bayat, 2014). Writing anxiety is a type of situational anxiety and can be treated. Providing multiple opportunities for writing in multiple genres throughout a course are staples of effective writing instruction. The CCSS require students to write in multiple genres and also to “make improvements to a piece of writing over multiple drafts” (CCSS Initiative, 2010, p. 12). Students must spend a great deal of time writing to be proficient in multiple genres. Graham et al.’s (2011) meta-analysis noted that single writing samples from students did not provide an accurate picture of the students’ proficiency level. Graham et al. found that teachers who tracked student writing samples over the course of a semester saw statistically significant gains in the students’ writing.

Writing is a process that occurs over multiple writing experiences and requires teachers’ ongoing monitoring of student progress. Effective writing instruction cannot exist apart from frequent opportunities to write (Cumming, 2013). Cumming (2013) cited six accomplishments of sustained writing practice: (1) students improve their writing, (2) students improve their reading, (3) students become more fluent in the writing process, (4) student become more comfortable with the writing process, (5) students can transfer knowledge between contexts, and (6) students can think more deeply thinking about content and construct new knowledge. Teachers must
provide explicit writing instruction to promote effective writing classes. Teachers can use explicit instruction to support students in the multiple genres of required writing and use research-based effective writing strategy to inform scaffolding techniques that will develop students into better writers.

**Summative Assessments**

Summative assessments are most often given at the end of the course. Summative assessments are usually referred to as high-stakes achievement tests and are used at all levels of education (Popham et al., 2014). A test can be considered high-stakes when at least one stakeholder potentially faces some type of consequence due to the results (Stobart & Eggen, 2012). Summative assessments tend to consist of norm-referenced, prespecified criteria and generally take place at the end of the course. Norm-referenced tests provide a description of the general performance of large groups students at a specific point in time (Popham et al., 2014). Therefore, summative end-of-course testing is administered after the learning has occurred and requires students to interpret their own results (Fastré, Van der Klink, Sluijsmans, & van Merriënboer, 2013). In most U.S. states, summative writing assessments are stand-alone assessments characterized by their content and timing. On-demand writing assessments average about 1 hour and usually require students to write about a preassigned topic (Brookhart, 2013). These time and topic parameters tend to create test anxiety and elevated apprehension amongst students by removing the opportunity for students to work at their own pace (Olinghouse et al., 2012). The delay between the assessment date, the time the learning occurred, and the dissemination of scores provides little opportunity for meaningful data usage by either students or teachers (Frohbieter, Greenwald, Stecher, & Schwartz, 2011). End-of-course tests measure the learning that occurred over a semester; however, students have moved on to the next course by
the time the scores are reported. These circumstances surrounding summative assessment place students in the role of receiving scores and the teacher in the role of distributing scores (Fastré et al., 2013).

Summative assessment scores are often used for accountability purposes and impact districts, students, and individual teachers. Summative assessments are generally used to track achievement trends annually (Popham et al., 2014). Data derived from summative end of year testing is the most influential decision-making tool in accountability systems (Shepard, Penuel, & Davidson, 2017). Some teachers have referred to summative assessments as “summative judgments” (Wormeli, 2017, p. 40) due to the high-stakes accountability associated with these summative assessments. These high-stakes accountability measures lead teachers to be proactive in their response to testing by teaching to the test. The teaching-to-the-test approach places the focus on test preparation and narrows the curricular goals, which can make the curriculum less coherent and meaningful to student learning (Beaver & Weinbaum, 2015).

Summative assessments are often associated with (a) labeling, (b) placing high value on material that is easier to test, (c) restricting learning to only tested material, and (d) the lessening of student motivation to learn (Carless, 2012). Summative assessments are meant to be assessments of learning; however, due to the accountability attached to the assessments, teachers often feel pressure to prepare students for the test rather than teaching students how to become better writers. The struggle to balance classroom time spent on writing instruction and classroom time spent on test preparation has led some districts to narrow the curriculum to only focus on tested materials and skills (Clark, 2012; Olinghouse et al., 2012). Popham et al. (2014) observed that the data provided by these high-stakes tests should be greatly scrutinized. Reteaching and test preparation consistently appear in the review literature as instructional approaches in courses
requiring summative assessment; however, teachers rarely use end-of-course summative assessment data to monitor or adjust instruction because of the lag in test results (Filsecker & Kerres, 2012). Summative end-of-course testing data identify where students are in relation to state standards, while varied assessments—at the school and classroom levels combined with summative year end data—provide a more complete representation of student achievement (Beaver & Weinbaum, 2015; Brookhart, 2015).

**Formative Assessment**

By definition, summative assessment can occur only at the end of the course; therefore, formative assessments—assessments administered prior to the summative—have become a growing trend in education over the last 2 decades (Crossouard & Pryor, 2012). Black and Wiliam (1998) have been at the forefront of formative assessment literature for the last 20 years. Formative assessment first appeared in research literature in the 1960s, but it was not until Black and Wiliam began to study and review the related literature that formative assessment became a common term in education. Black and Wiliam focused the assessment debate toward formative assessment and are credited with the widespread use of formative assessment (Carless, 2009; Frohbieter et al., 2011). Black and Wiliam initially found formative assessment to be misunderstood and misused by teachers, which led to ineffective implementation; therefore, it was clear that teachers’ perceptions concerning formative assessment needed to change for these assessments to be effectively implemented.

Black and Wiliam followed up this initial finding with their 2009 seminal work *Inside the Black Box: Raising Standards Through Classroom Assessment*. The black box was used as a metaphor for the classroom. Black and Wiliam’s (2006) thesis stated that the inputs will influence the outputs, but the only way to receive the desired outputs is by examining what
happens in the box. This rationale is why many scholars have argued that effective formative assessment should be the center of classroom instruction and the standard of student learning. A consensus definition of formative assessment is lacking in the literature (Li, 2016). Ainsworth and Viegut (2014) summarized the definitions and ideas of formative assessment as an “assessment for learning,” ‘planned process of assessment,’ ‘loops of instruction,’ ‘assessment as feedback’” (p. 31). Black and Wiliam’s (2009) definition of formative assessment is the most commonly quoted. Black and Wiliam define formative assessment as the extent to which evidence is obtained and interpreted by both students and teachers to inform instructional decisions. This definition was derived from the basic concepts of teaching and learning: where the student is in his or her learning, where the student needs to go, and how the teacher gets the student there (Ramaprasad, 1983; Wiliam & Thompson, 2007). Therefore, formative assessment can be viewed as an integral connection between curriculum and end of course assessment; formative assessment is a way to support learning rather than a means to support summative assessment (Crossouard & Pryor, 2012). The emphasis in formative assessment should be on developing students’ skills and adjusting instruction as opposed to recording grades that may only identify student weaknesses (Wormeli, 2017). In practice, this theory and definition take the form of a formative assessment cycle of learning: gather the evidence, make inferences based on the evidence, and use the evidence to adjust instruction (Black & Wiliam, 1998, 2009).

Formative assessment as a classroom practice. The increase in formative assessment literature has uncovered assumptions made by teachers. These assumptions are due to teachers overlooking formative assessments and assuming that formative assessments were present in the curriculum without explicitly looking for it (Alkharusi, Aldhafri, Alnahani, & Alkalbani, 2014). Teachers who do cite specific examples of formative assessments in their classrooms are often
implementing those assessments incorrectly (Wormeli, 2017). If teachers misinterpret formative assessment, then students misinterpret the intention behind the assessment as well. Students may not understand why they are being required to learn the course material, what learning goals they are expected to meet, or how to interpret the results of the assessment (Wylie & Lyon, 2015). Teachers’ daily classroom practices involve some type of assessment; consequently, teachers often equate formative assessment as a test (Lam, 2013). However, formative assessment is not a test, single instrument, or instructional tool. Formative assessment is a process; it is a cycle of feedback that leads to student learning and instructional shifts (Filsecker & Kerres, 2009; Hudesman et al., 2013; Shute & Kim, 2014). Formative assessment has been proven to lead to increased student achievement and help close the achievement gap; however, this only occurs when formative assessment is used as a strategic tool embedded in daily instruction (Li, 2016; Wormeli, 2017). Lower achieving students benefit the most from increased frequency and increased quality of formative assessment (Black & Wiliam, 2006; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012; McMillan, Venable, & Varier, 2013); therefore, teachers may also assume that more formative assessment equates to more student learning.

Districts possess a basic knowledge of formative assessment but lack guidance on the creation, implementation, and use of formative assessment. Misunderstanding concerning implementation and administration of formative assessment has led to stand-alone, quarterly, or embedded assessment in the curriculum (Frohbieter et al, 2011). Hattie (2009) conducted a meta-analyses and found an effect size of .41 from frequent testing and an effect size of .62 when the testing was used in conjunction with effective formative assessment practices. Similarly, Black and Wiliam (2006) noted that the U.S. educational system generally assumes that more frequent testing alone will lead to increased student achievement. Black and Wiliam concluded that
formative assessment was ineffective in these instances. Black and Wiliam asserted that formative assessment increases student achievement effect size between .4 and .7 standard deviations; however, Kingston and Nash (2011) found the effect size to be less than .4 in their meta-analysis of formative assessment data. Kingston and Nash contributed the discrepancy of effect sizes to lack of clearly defined formative assessment implementation; therefore, more study was needed to determine to what extent formative assessment increases student achievement. Wylie and Lyon (2015) also identified teacher implementation of formative assessment as a topic of further study when reviewing formative assessment data. The goal of formative assessment is to raise student achievement and improve teacher instructional practices; formative assessment is most effective as a process that occurs over time as opposed to an instructional tool or single assessment (Stull et al., 2011). Effective formative assessment must be strategically and thoughtfully aligned with the curriculum rather than being impulsively included in a lesson (Wormeli, 2017). Assessments should be connected to the curriculum and instruction students are presently receiving. Horizontal alignment between the instruction of the teacher and the assessment of the curriculum goals is the cornerstone of effective student learning (Abrams, Varier, Jackson, 2016).

Herman et al. (2015) found a correlation between teachers’ content and pedagogical knowledge and student achievement through formative assessment. Herman et al. also found that teachers’ content and pedagogical knowledge increased due when teachers followed the formative assessment cycle. Effective formative assessment is dependent on teachers’ content and pedagogical knowledge. Haug and Ødegaard (2015) noted that formative assessment can come to be viewed as a “strategy focusing more on pedagogical skill than on the content to be taught” (p. 630). Data derived from formative assessments guide how the assessment is used and
ensure the assessment’s effectiveness (Ainsworth & Viegut, 2014). Teachers use formative assessment to provide feedback to students about their current performance as well as specific learning objectives and goals based on their assessment data. Additionally, teachers can use the data from formative assessments to adjust instruction and meet the needs of students (Crossouard & Pryor, 2012). Students tend to perceive classroom environments as either “test oriented” or “learning oriented” (Alkharusi et al., 2014, p. 95); this perception affects how likely students will be to adopt the goals of the classroom teacher and master the standards. Students’ who believe have high academic efficacy demonstrate increased motivation and increased student achievement (Zimmerman & Bandura, 1994). Learning-oriented classroom teachers are much more likely to convey assessment as learning to students.

**Formative assessment in writing classrooms.** Both formative and summative writing assessments have been debated for the last 60 years. From 1950–1970, writing assessments were objective, item-based tests. From 1970–1986, holistically scored essays were assessed. Presently, program-based assessments are used (Behizadeh & Engelhard, 2011). Theories of writing and composition grounded in classroom instructional practices have consistently “influenced assessment practices in the United States” (Behizadeh & Engelhard, 2011, p. 205). Brookhart (2013) noted that writing assessment has been a consistent measure in states’ accountability systems. Formative assessments have been a growing source of attention for both state and federal policymakers (Crossouard & Pryor, 2012); however, more data are needed to determine how formative assessment data gathered throughout the course of study affects summative achievement scores on writing assessments. As theories about classroom instruction evolve, so do theories about writing assessments. This evolution creates a widening gap between writing theory and practice of writing assessments, especially with high-stakes assessments. As these
authors noted, the evolution to “improve writing instruction starts with a revolution to improve writing assessments” (p. 206).

Instructional theories and practices influenced how students were assessed in writing. Writing assessments should encourage good instruction (Brimi, 2012). Through their historical examination of writing assessments, Behizadeh and Engelhard (2011) cited the lack of locally constructed writing assessments and called for this need to be addressed by federal and state educational leaders. The lack of locally constructed formative assessment practice places more significant weight on the summative assessment data results. Formative assessments are generally embedded in the curriculum and may be overlooked, as it is assumed that formative assessments are present without examining how these assessments can be used to produce student data (Crossouard & Pryor, 2012). Alkharusi et al. (2014) asserted that students must understand the clear connection between instruction and assessment in order to maximize learning; therefore, “assessment should be something done with and for students rather than something done to them” (p. 103).

Focusing on summative writing scores to inform local decisions has had a dismal effect on writing instruction. Summative assessments often have only one writing sample; thus, teachers are inadvertently encouraged to teach only one type of writing: the genre being assessed (Benjamin, 2013; Brimi, 2012; Popham et al., 2014). Due to the credence given to the results of summative assessments, teachers feel pressure to demonstrate student achievement and growth by preparing students for the summative assessment to the detriment of the other genres of writing students will encounter (Dixson & Worrell, 2016). Such “teaching to the test” and “formulaic writing” (Benjamin, 2013, p. 23) is viewed unfavorably by both stakeholders and teachers. Students graduating from high school and enrolling in college are assumed to have
adequate writing skills in multiple genres of writing. Schools in states that require an end-of-year summative assessment may see their students engaging in fewer multi-genre writing experiences as teachers are pressured to teach to the test to increase student achievement data (Brimi, 2012). Teachers who focus on the summative assessment instead of multiple formative writing assessments create a deficiency in students’ writing abilities.

Teachers who focus on summative writing assessments rely heavily on the rubric for the end-of-course summative assessment (Tuttle, 2013). Students did not understand the summative rubric language and viewed it as a type of “checklist” which provided “general descriptors” but offered little direction to actually improve their writing (Tuttle, 2013, p. 22). Rubrics are widely used in writing instruction; however, rubrics are limited in scope (Beck, Llosa, Black, & Anderson, 2018). Rubrics reflect students’ achievement level in reference to state standards; however, students do not always understand their scores on rubrics due to their lack of comprehension of the rubric language or because the rubric criteria are poorly written (Broadbent, Panadero, & Boud, 2018). Writing rubrics evaluate the finished writing product rather than provide guidance during the writing process. Teachers should uses assessments as a way to focus on student learning rather than assigning a grade (Beck et al., 2018). Butler and Britt (2011) studied recently enrolled community college students and found that expert writers revise their writing at twice the rate of novice writers: 24% to 12%, respectively. Butler and Britt found that high school students were revising their writing at 11%. Butler and Britt’s study participants were novice writers who only made revisions at the sentence level as opposed to revising large portions of their written text. Frequent formative assessment provides opportunities for students to receive feedback from their teacher and opportunities to apply teacher feedback to create more effective writing samples (Lang, Stanley, & Moore, 2013). Beck
et al. (2018) found that using formative assessments in the writing classroom helped teachers “construct learning trajectories for all students, plan instruction and set benchmarks for measuring student progress, focused their feedback to students, which leads to probable increased writing scores” (p. 69). However, more data were needed to determine how these formative writing assessments were implemented and how the data were used to adjust instruction and monitor progress. Formative assessment has multiple definitions and implications for implementation that could greatly influence the effect of populations (Haug & Ødegaard, 2015). Tuttle (2013) observed that summative assessments begin at the final stage of the writing process while formative assessment provide an opportunity for revision at every stage of the writing process, thus allowing teachers to provide instructional support for students at all five stages of the writing process. The five stages of writing from a formative assessment point of view include student writing, monitor, diagnose, feedback, and growth (Tuttle, 2013).

**Common formative assessments.** Common assessments are defined as assessments for learning that are collaboratively designed by a grade or course-level team of educators to assess student understanding of the learning intentions and student success criteria currently in focus within a curricular unit of study (Ainsworth & Viegut, 2014). Common formative assessments are developed at the school or district level for purposes specific to the school or district (Abrams et al., 2016). Common formative assessments are aligned with the curriculum and are given at predetermined points throughout the course.

Common formative assessments are given at strategic points to identify how students are progressing toward mastery of the state standards. In semester-long classes, common formative assessments are most often administered at the mid-point of each 9 week period (Brookhart, 2015). On the testing spectrum, common formative assessments generally occur between
summative end-of-course testing and daily classroom assessments, and on a more frequent basis (Abrams, McMillian, & Wetzel, 2015). Teachers who administer formative assessments are provided with multiple opportunities to obtain data to identify the learning trajectory of students. Frequently administered assessments allow for feedback to be given more frequently, thus resulting in higher student achievement; this is a conventional idea concerning common formative assessments (Konstantopoulos, Li, Miller, & van der Ploeg, 2017). The data provided by multiple common formative assessments help teachers determine students’ learning needs to help students progress towards mastery of the state standards. Teachers must modify instruction to help students make achievement and growth gains. The frequency of common formative assessments provides teachers with the data points necessary to measure student growth over the course of the semester (Ado, 2013). To adjust instruction based on common formative assessment data, teachers must identify students nearing mastery and students in need of remediation and guide student grouping decisions to determine possible reteaching of curriculum (Abrams et al., 2016). Frequent common formative assessments provide the basis for modification of learning (Popham, 2014). The sequence of the assessments is determined by the curriculum. The scope and sequence of common formative assessments are determined by the standards measured on the end-of-course summative achievement test (Gareis & Grant, 2015). Course-level teacher teams or individual classroom teachers generally create common formative assessments to assess students’ progress towards mastery of the standards (Brookhart, 2015). Many studies have focused on the creation of teacher teams and how teacher teams make data-based decisions; however, more data were needed on how these teacher teams create assessments that increase student learning (Shepard, Penuel, & Davidson, 2017. Common formative assessments are created at the local level; thus, the assessments are often not standardized and
teachers must apply sound pedagogical thinking to create these assessments. Common formative assessments are classified as low-stakes when compared to summative end-of-course assessments; however, common formative assessments can reflect the same level of rigor in reference to the writing tasks that will be assessed on the end-of-course summative assessment (Dixson & Worrell, 2016).

It is unclear if common formative assessments or individual teacher created formative assessments lead to increased student achievement. Abrams et al. (2015) observed that the results of common formative assessments were not consistent among the population of teachers within a school. Abrams et al. also observed that the majority of teachers noted that common formative assessments helped them “identify and correct gaps in their teaching,” but the majority of teachers reported “their own assessments provided more useful information” (p. 150) than common formative assessments.

Common formative assessments provide students with frequent opportunities to write in multiple modes—such as expository, informational, and narrative writing—and students are provided regular feedback on their writing. Common formative assessments are aligned to the standards and strategically placed in the curriculum; therefore, student performance can be compared (Brookhart, 2015). Students are more likely to take ownership of their learning and develop self-regulation characteristics when they reflect regularly on their writing. Students who are exposed to rigorous writing tasks at regular intervals better understand the expectations of the state standards and the learning targets in the curriculum (Ronfeldt, Farmer, McQueen & Grissom, 2015). Both teachers and students can better prepare for the end-of-course summative assessment when teachers provide students with feedback using rubrics that are aligned to the state assessment. The frequency of multiple common formative assessments provides teachers
with opportunities to reflect on student writing production. Common formative assessments provide opportunities to modify instruction based on student needs. Due to the current educational climate, educators are often confused about formative and summative assessment; this confusion can be attributed to school accountability factors and the increase in both testing and the reporting of testing results (Dixson & Worrell, 2016). Teachers can identify student and classroom deficiencies to guide instructional decisions, which could positively affect student achievement on end-of-course tests.

**Gender Difference in Writing Achievement**

Prior literature has been devoted to the perceived differences between males and females students in terms of academic achievement (Reynolds et al., 2015). Gender stereotypes are often used to explain the difference in male and female student achievement data (Alkharusi et al., 2014). Reading, mathematics, and science achievement scores are often cited and provide evidence to support theories concerning gender differences in achievement (Troia, Harbaugh, Shankland, Wolbers, & Lawrence, 2012). Gender differences are often less significant than expected despite assumptions about gender differences in academic achievement; however, writing achievement appears to be an exception (Reynolds et al., 2015; Scheiber et al., 2015). Kim et al., (2015) postulated that “gender appears to matter in writing achievement” (p. 85). All phases of writing reflect a difference between the sexes (Troia et al., 2012). This gender gap is evident across populations and research designs (Scheiber et al., 2015). Jones’ (2012) indicated that male students do not score as well as female students in reading or writing. Similarly, Pargulski and Reynolds (2017) found that female students demonstrated higher scores on writing assessments in both the 50th and 75th percentiles. In the 1998 and 2007 NAEP, female students scored higher than male students in grade 8 and 12 in all three modes of writing: informational,
narrative, and argumentative (Troia et al., 2012). This trend continued into the 2000s, when female students scored between 17 and 25 points higher than male students in grades 4, 8, and 12 (Kim et al., 2015; Reynolds et al., 2015). The female advantage in writing performance was introduced in the literature over 20 years ago but has received little attention (Scheiber et al., 2015). The achievement gap between male and female students has remained relatively unchanged over the last 2 decades (Jones, 2012). Minimal data exist regarding writing assessment score differences between the genders (Scheiber et al., 2015).

The gender gap in writing achievement begins early and has been proven to continue throughout the educational career of students. The gender gap begins to appear as early as age 6; furthermore, the female advantage increases with age (Scheiber et al., 2015). Pargulski and Reynolds’s (2017) measured writing achievement in terms of sentence composition, essay composition, and spelling to explain the gender gap in writing assessment. Pargulski and Reynolds’s found that female students scored higher at both the 50th and 75th percentile as well as at both tails of distribution, which showed an even greater difference between the genders. Additionally, male students are more likely to fail writing assessments than female students (Scheiber et al., 2015). The gap in the between male and female students’ writing success is most likely attributable to male students being more likely to engage in goal-oriented approaches to education when they perceive the assessment to be purposeful. The majority of male students are unable to see the purpose in the writing because it is a process approach to learning. Male students are much more likely to be successful in short-term, goal-oriented assignments (Alkharusi et al., 2014).
Tennessee State Testing

The state of Tennessee reconfigured its testing format in 2015. As a RTTT state, Tennessee enacted changes that cultivated a new assessment climate for the state. The changes in standards have created new goals for students that have led to a change in assessment practices (Camp, 2012). The TCAP was the standardized summative assessment previously used in the state. The Tennessee Department of Education (2015) defined summative or annual assessments as “measures of student learning at the end of the semester/year” (p. 4). All students in grades 3–11 took this annual exam in ELA, math, science, and social studies. Additionally, students in grades 3, 5, 8, and 11 took the TCAP Writing Test, and high school juniors are required to take either the ACT or SAT. When Tennessee was granted RTTT grant funding, they changed the summative testing format to better address the more rigorous standards of the CCSS. Beginning in 2015-2016 students in grades 3–12 will take TNReady, which replaces the ELA and writing assessment (Tennessee Department of Education, 2015). All students in grades 3–12 now take a writing assessment based on multiple genres: informational and expository, opinion and argument, and narrative. Teachers do not know which genre will be required of students prior to testing. This is a substantial change from previous TCAP assessments, when teachers for grades 3, 5, 8, and 11 were informed of the specified genres that students would be assessed on. Students are now required to complete two on-demand writing tasks. Each task is in response to a different writing genre, and both writing tasks are independent of each other. In an on-demand writing assessment, a student is given a prompt and asked to formulate a response within a specified time frame (Benjamin, 2013). Students have been taking TNReady since 2014. In subsequent pilot years, students were tested in the same assessment format and on the same platform and the scores were reported to the schools. Similar to the NAEP writing assessment,
Tennessee’s TNReady is an on-demand, timed writing assessment that provides detailed data on the proficiency level of students in grades 3–12.

Members of the Tennessee Task Force on Student Testing and Assessment were tasked with reviewing summative and formative assessment data and logistics to make recommendations for TNReady (Tennessee Department of Education, 2015). More study was needed to identify how formative assessments can support student achievement on the summative TNReady end-of-course assessment. States and district leaders aim to demonstrate widespread student achievement on the TNReady assessment. Student achievement can be increased through coherent alignment of assessments that are grounded in learning theory. Assessment systems must be coherent to students and vertically aligned in order to increase student achievement. Classroom, school, and district-level assessments must reflect the format, expectations, and requirements of the state assessment (Shepard et al., 2015).

TNReady is a criterion referenced test. This criterion referenced assessment uses cut scores to determine students’ level of mastery. Students’ classifications are identified as below, approaching, on track, or mastered. The purpose of criterion referenced testing is to identify where students scored in terms of the state standards (Wilson, 2013). The members of the Tennessee Task Force on Student Testing and Assessment recommended that the following assessment requirements be eliminated: (a) the explore and plan testing requirements in the grades 8 and 10, (b) both summative assessments created by ACT, and (c) mandatory state testing in K–2 (Tennessee Department of Education, 2015). The recommendations to limit the amount of summative assessments were the result of reviewing survey data provided by teachers.

During the 2014–2015 school year, Tennessee teachers were asked to respond to the statement, “I understand how to use results from statewide standardized exams to improve my
teaching;” 76.6% responded in the affirmative (Tennessee Department of Education, 2015, p. 13). Conversely, when asked if the “results of statewide standardized exams help me determine if my students are gaining skills and knowledge to meet state standards,” 53.5% of teachers responded in the negative (Tennessee Department of Education, 2015, p. 13). This discrepancy highlighted the concern that summative assessments likely did not strengthen student learning or teacher instruction. In the same survey, Tennessee teachers raised complaints about too much time being spent on summative assessments. Sixty-two percent of teachers felt they were spending too much instructional time preparing for the assessment, and 69% of teachers believed they spent too much time on actual testing. These survey responses from Tennessee educators led the task force to issue guidelines for formative assessment usage and implementation.

Formative assessments were defined as “measures of student learning throughout the year that enable educators to determine if students are making progress and how best to adjust instruction” (Tennessee Department of Education, 2015, p. 4). Currently, Tennessee students spend 11–12 hours, or 1%, of the school year being tested on summative exams; no guidelines for formative assessment had been provided prior to this report. Summative assessments may highlight gaps in achievement, but only formative assessments can inform instruction by providing teachers with the data necessary to modify their instructional practices. The Tennessee Department of Education (2015) concluded that “TNReady will directly change how formative assessments are used and implemented” (p. 10). The frequency and variety of formative assessments vary across the state of Tennessee. Members of the Tennessee Task Force on Student Testing and Assessment recommended “providing expectations to districts regarding formative assessment usage and communication” (p. 5) to develop a more uniform set of formative assessment guidelines. Furthermore, task force members recommended that formative
assessments be locally created and encouraged teachers to create their own formative assessments; teacher-created formative assessments are more aligned to Tennessee state standards than vendor-provided assessments. Teachers can create and use common formative assessments to gather data during the instructional cycle that will guide daily lesson planning and address students’ needs.

**Summary**

This review of the literature summarized much of the data related to common formative assessments and their impact on summative test scores. Summative year-end writing assessments remain a requirement of schools and serve as a form of accountability. Common formative assessments are used in the classroom as part of instruction and help prepare students to score well on the end-of-course writing assessment. Therefore, it is crucial to find evidence-based strategies for improving student achievement on writing summative assessments and develop a coherent system in which all three types of assessment are present and provide data to increase student achievement (Shepard, Penuel, & Davidson, 2016). When used as a process consisting of timely student feedback and adjusted teacher instructional approaches, common formative assessment has been shown to increase scores and prepare students to be more successful writers. Common formative assessments have been shown to increase student achievement when integrated into the curriculum, administered to groups of students at multiple, pre-determined intervals, and when used to adjust teacher instruction (Black & Wiliam, 1996; Graham et al., 2011). Nevertheless, a gap remains in the current literature regarding the use of formative assessments to state accountability tests in writing. Few studies have examined the effect of teacher-created (formative assessments) and team-created (common formative assessments) on
student writing achievement scores. The current study provided data that contributed to the existing gap in the literature.
CHAPTER THREE: METHODS

Overview

This quantitative study examined whether common formative writing assessments affected academic achievement on end-of-course writing assessments. The formative assessment scores were analyzed to determine if a causal comparative relationship existed between formative and summative achievement data. The purpose Chapter 3 is to describe the: (a) sample population selected for this study, (b) instruments that were administered for data collection, (c) methods, materials and procedures utilized to implement and collect the data for the study, and (d) selection and use of statistical procedures employed in the analysis of the collected data.

Design

The purpose of this non-experimental, causal-comparative study was to examine the possible effects of the use of common formative assessments on student writing achievement scores. For this study, one group of eleventh-grade students used a school system’s curriculum plan and participated in common formative assessments throughout the semester. The second group of eleventh-grade students only used teacher-designed formative assessments. This causal-comparative design worked well within the school setting, as neither of the classes nor variables could have been organized into experimental groups during the school year (Ary et al., 2006; Gall et al., 2007). The common formative writing assessments administered to both groups were a district curricular decision; therefore, an experimental design was not used. This study used end-of-year TNReady writing scores from two groups of eleventh-grade students for the classes.
Research Question(s)

RQ1: What difference exists (if any) between writing achievement test (category) scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade students who took two formative assessments created by individual teachers?

RQ2: What difference exists (if any) between writing achievement test (category) scores of eleventh-grade male students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade female students who took two formative writing assessments created by teacher teams (common formative assessments)?

Hypotheses

The independent variable in this study was the common formative writing assessments. In Group 1—the treatment group—these assessments were comprised of multiple modes of writing (expository, argumentative, and narrative). Students in Group 1 were given these assessments throughout the semester of the block schedule English III class. Students in Group 2 were not given any common formative assessments. Group 2 only took teacher-created formative assessments during the semester. Both groups were held accountable to the same Tennessee state standards for English III courses. The dependent variable was the TNReady writing assessment that students took during the 2016–2017 year.

The null hypotheses for this study were as follows.

H₀₁: There will be no statistically significant difference between the 2017 TNReady writing assessment development achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and
the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H₀₂:** There will be no statistically significant difference between the 2017 TNReady writing assessment conventions achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H₀₃:** There will be no statistically significant difference between the 2017 TNReady writing assessment language achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H₀₄:** There will be no statistically significant difference between the 2017 TNReady writing assessment focus and organization achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H₀₅:** There will be no statistically significant difference between the 2017 TNReady writing assessment development achievement scores of eleventh-grade male and female students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by teacher teams (common formative assessments).
H_06: There will be no statistically significant difference between the 2017 TNReady writing assessment conventions achievement scores of eleventh-grade male and female students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by teacher teams (common formative assessments).

H_07: There will be no statistically significant difference between the 2017 TNReady writing assessment language achievement scores of eleventh-grade male and female students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by teacher teams (common formative assessments).

H_08: There will be no statistically significant difference between the 2017 TNReady writing assessment focus and organization achievement scores of eleventh-grade male and female students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by teacher teams (common formative assessments).

Participants and Setting

The study participants were drawn from the eleventh-grade population of eight public high schools located in northeastern Tennessee during the 2016–2017 school year. The residents in school districts had lower-to-middle class income and the district was predominately rural. All eight schools shared similar demographics, block schedules, and were generally comparable to one another. Eleventh-grade students are typically juniors and take English III classes, which is otherwise known as junior English.
The number of participants sampled in this study was approximately 1600, which exceeded the required minimum for a medium effect size (Gall et al., 2007). The sample came from eight different high schools and consisted of 776 male students and 822 female students. The gender ratio was 48% male students to 51% female students. The study population comprised of 7862 White students, 135 Black students, and 158 Hispanic students. Within the study population, 26.5% of students were economically disadvantaged. Students were selected from three junior English classes within each school. Teachers of junior English classes in these eight schools followed similar curricula that aligned to the Tennessee State Standards. Students in Group 1 participated in multiple common formative writing assessments throughout the semester. Students in Group 2 did not require or participate in common formative writing assessments.

Assessments for the treatment group were comprised of expository, argumentative, and narrative writing and were given at specific points during the semester of the block schedule English III class. Per the district curriculum, teachers were required to administer at least three common formative writing assessments during the semester. The common assessments were placed at the end of each unit of study; each unit of study was 9 weeks long.

Teachers of the comparison group did not require or provide common formative writing assessments to students. These teachers could design and administer teacher-created assessments throughout the academic term.

**Instrumentation**

The measurement scores for this study were derived from Tennessee’s state achievement test, TNReady. The state of Tennessee contracted Questar as the vendor of TNReady. To ensure the content validity of the TNReady, Questar and the Tennessee Department of Education
worked in conjunction to align the summative assessment with the Tennessee State Standards set forth by the Tennessee Board Of Education. Tennessee teachers were given blueprints for all ELA courses and were notified of the genres of writing that could possibly be required of students. Reliability is one of the two cornerstones of technical quality in testing and measurement. The total assessment reliabilities ranged from .85 to .89 for ELA in the 2016–2017 TNReady writing assessment. TNReady items were written by qualified, professional content specialists (Tennessee Department of Education, 2015). The TNReady writing assessment used a rubric to score students on a 1 to 4 scale in each category. The individual scores in each category—development, focus and organization, and language and convention—were averaged to generate the student and overall writing scores in addition to the individual scores of each category.

Curriculum specialists and committees of Tennessee educators reviewed the items. Writing tasks were evaluated for overall quality and clarity, content coverage and appropriateness, alignment to the curriculum, and grade appropriate stimuli with an emphasis on higher-order thinking skills. Writing tasks should be considered free from bias toward or against any group. To ensure the assessment’s reliability, all exemplary sample papers and training materials were selected from Tennessee student responses from 2013–2016 school years. A committee comprised of Tennessee teachers and Questar trained scorers established baseline score ranges by reviewing and scoring responses. Every student essay was examined twice to ensure accurate scoring (Tennessee Department of Education, 2018).

The team-generated assessments from eight high school English III classes were included in this study. Each common formative assessment included the three separate writing genres: expository, informational, and narrative writing. The assessments were sampled across state

The common formative writing assessments were developed at the district level by either curriculum supervisors or ELA instructional coaches leading English III teacher teams. The instructional coach who oversaw the creation of these assessments contributed to the ELA standards at the state level and now works for the Tennessee Department of Education. The assessment content was aligned to both the Tennessee State Standards and the district curriculum guide (Tennessee Department of Education, 2015). The assessments provided data to English III teachers that informed their instructional practices, identified students in need of additional support, and also helped prepare students for the end-of-course summative assessment.

Items included in the development of these assessments were selected from prompts and writing tasks were identified from exemplar examples provided by Questar. Teachers reviewed the students’ writing samples after the administration of each common formative writing assessment. English III teachers met at both the school and district level to review trends and proficiency levels. A comprehensive review of both the prompt and student work samples was conducted to review the quality of the assessments. English III teachers provided feedback to each student based on the state’s writing rubric; the state writing rubric addressed the standards
of both the curriculum and the common formative assessment. Student writing samples were classified according to proficiency level: approaching, on-track, or mastered. These proficiency levels were also assigned by the state on the summative end-of-course assessment.

Individual teacher-created formative writing assessments were created in isolation based on individual teachers’ pedagogical knowledge as content experts. Teacher-creative formative assessments were aligned with the objectives and standards included in the curriculum to prepare students to take the summative end-of-course writing assessment. These formative assessments were considered to be an assessment for learning and were embedded into the daily classroom instruction (Herman et al., 2015). Some local educational agencies allowed classroom teachers the autonomy to create and administer classroom formative assessments according to their preferences (Tennessee Department of Education, 2015). Teacher-created formative assessments were used by individual teachers in their own classroom for instructional purposes; thus, the assessments were not vetted as thoroughly for validity as team-created common assessments. Teachers administered formative assessments throughout the semester and used the data to provide feedback to students. Teachers adjusted their instruction to meet the needs of students based on the data provided by the assessment (Shepard et al., 2016). The writing tasks included on formative assessments resembled the summative end-of-course writing assessment in terms of verbiage and expectations for each genre of writing (Tennessee Department of Education, 2015). Teachers created their formative assessments based on item-release documents of previously administered assessments from the Tennessee Department Of Education. Teacher-created formative assessments referenced the state writing rubric—which was used to score the summative end-of-course assessment—when providing feedback to students.
Writing is subjective in nature and generally studied from a qualitative point of reference (Gillespie et al., 2014). Student achievement scores and proficiency levels are objective in nature and are reported quantitatively. The assessment items used to prepare students for summative end-of-course writing generally reflect teachers’ views of writing instruction and include some type of formative assessment (Abrams et al., 2015). Teacher-created formative assessments are much more subjective and have a lesser degree of validity than common formative assessments created by teams of teachers (Ainsworth & Viegut, 2014; Black & Wiliam, 2009). Common formative assessments have greater levels of validity in both the creation of the assessment and the reporting of the assessment data. Teacher-created and common formative assessments are used to prepare students for the summative end-of-course writing assessment; however, these assessment approaches varied in frequency, administration, creation, and data reporting (Dolin, Black, Harlen, & Tiberghien, 2018). The purpose of this study was to examine whether differences existed in TNReady writing scores among eleventh-grade students who took team-created common formative writing assessments at least twice a semester and students who took teacher-created formative assessments.

**Procedures**

This study was conducted with the approval of three East Tennessee public school systems. Internal Review Board approval was granted (see Appendix A) before any data collection or analysis took place for this study. This study was ex post facto; therefore, only archival data were be used. The data provided for this study were supplied through the Tennessee Department of Education and the local school systems.

Data were collected from the TNReady state achievement tests for English III classes in eight high schools during the 2016–2017 school year. The data were recorded in a spreadsheet.
and categorized by school, gender, and independent variable. The Statistical Package for the Social Sciences software (SPSS) was used to analyze the data and test the hypotheses.

**Data Analysis**

This study data were used to measure the possible effects of common formative writing assessments on student achievement scores for students enrolled in eleventh-grade English classes over the 2016–2017 school year. A causal-comparative study was conducted to test the null hypotheses concerning the relationship between common formative writing assessments and student achievement in each category (subtest) of the TNReady summative writing assessment. The use of a two-way analysis of covariance (ANOVA) was appropriate for this study because it compared the variance of both groups’ TNReady achievement scores to the individual students’ scores. The two-way ANOVA compared the data to determine if the difference in mean scores between the treatment and comparison group was statistically significant. The two-way ANOVA was used to establish if the observed difference between mean scores was due to chance or through true differences (Gall et al., 2007).
CHAPTER FOUR: FINDINGS

Overview

The purpose of this quantitative causal-comparative study was to examine the possible effects of the use of common formative assessments on student writing achievement scores. In this study, one group of eleventh-grade students were administered at least two teacher team-created common formative assessments throughout the semester while the other group was administered at least two teacher-created formative assessments. The study used end-of-year TNReady writing scores for the classes from two groups of eleventh-grade students from three districts. The writing achievement scores were reported in four categories, as referenced in the state writing rubric: development, focus and organization, conventions, and language. A two-way ANOVA was used to determine if statistically significant differences existed between the two types of assessments administered to students over the course of a semester.

Research Questions

RQ1: What difference exists (if any) between writing achievement test (category) scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade students who took two formative assessments created by individual teachers?

RQ2: What difference exists (if any) between writing achievement test (category) scores of eleventh-grade male students who took two formative writing assessments created by teacher teams (common formative assessments) and the test scores of eleventh-grade female students who took two formative writing assessments created by teacher teams (common formative assessments)?
Null Hypotheses

**H_{01}:** There will be no statistically significant difference between the 2017 TNReady writing assessment development achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H_{02}:** There will be no statistically significant difference between the 2017 TNReady writing assessment conventions achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H_{03}:** There will be no statistically significant difference between the 2017 TNReady writing assessment language achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H_{04}:** There will be no statistically significant difference between the 2017 TNReady writing assessment focus and organization achievement scores of eleventh-grade students who took two formative writing assessments created by teacher teams (common formative assessments) and the scores of eleventh-grade students who took formative assessments created by individual teachers.

**H_{05}:** There will be no statistically significant difference between the 2017 TNReady writing assessment development achievement scores of eleventh-grade male and female students
who took two formative writing assessments created by teacher teams (common formative
assessments) and the scores of eleventh-grade students who took formative assessments created
by teacher teams (common formative assessments).

**H_{06}:** There will be no statistically significant difference between the 2017 TNReady
writing assessment conventions achievement scores of eleventh-grade male and female students
who took two formative writing assessments created by teacher teams (common formative
assessments) and the scores of eleventh-grade students who took formative assessments created
by teacher teams (common formative assessments).

**H_{07}:** There will be no statistically significant difference between the 2017 TNReady
writing assessment language achievement scores of eleventh-grade male and female students
who took two formative writing assessments created by teacher teams (common formative
assessments) and the scores of eleventh-grade students who took formative assessments created
by teacher teams (common formative assessments).

**H_{08}:** There will be no statistically significant difference between the 2017 TNReady
writing assessment focus and organization achievement scores of eleventh-grade male and
female students who took two formative writing assessments created by teacher teams (common
formative assessments) and the scores of eleventh-grade students who took formative
assessments created by teacher teams (common formative assessments).

**Descriptive Statistics**

Data used for this research study were gathered from eight high schools in the northeast
region of Tennessee. These eight schools represented three county public school systems. Four
of these high schools used common assessments (team-created assessments) in their English III
classrooms and four of the high schools used individual teacher created formative assessments.
The total population for the study was 1,598 eleventh-grade English III students who took the TNReady assessment in the 2016–2017 school year.

The scores were analyzed based on the indicators of the Tennessee state writing rubric: development, focus and organization, language, and conventions. The scores reported in these indicator categories are considered the writing achievement scores for this study. Of the 1,598 eleventh-grade students who took the TNReady assessment in 2016–2017, 48% \((n = 780)\) were administered teacher team-created common writing assessments and 51\% \((n = 818)\) were administered individual teacher created formative writing assessments. The independent variable for this study was the category of formative writing assessment administered to students. Table 1 presents a summary of the mean scores and significance levels for each category.

Table 1

*Scores by Category of Assessment*

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean Scores</th>
<th>N</th>
<th>Significance Level Between Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Individual</td>
<td>2.65</td>
<td>780</td>
<td>.031</td>
</tr>
<tr>
<td>Individual</td>
<td>2.55</td>
<td>818</td>
<td></td>
</tr>
<tr>
<td>Focus and Organization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Individual</td>
<td>2.95</td>
<td>780</td>
<td>.799</td>
</tr>
<tr>
<td>Individual</td>
<td>2.96</td>
<td>818</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Individual</td>
<td>3.01</td>
<td>780</td>
<td>.856</td>
</tr>
<tr>
<td>Individual</td>
<td>3.01</td>
<td>818</td>
<td></td>
</tr>
<tr>
<td>Conventions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Individual</td>
<td>5.96</td>
<td>780</td>
<td>.001</td>
</tr>
<tr>
<td>Individual</td>
<td>5.64</td>
<td>818</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Development, focus and organization, language, and conventions are individual TNReady writing assessment categories on the state writing rubric.

The gender of the students who participated in the study was also reported and analyzed. In this study, 51\% \((n = 822)\) of participants were female and 48\% \((n = 776)\) of participants were male. Gender was an independent variable in this study. The mean averages and the significance level based on gender is shown in Table 2.
Table 2

*Gender Scores by Category of Assessment*

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean Scores</th>
<th>N</th>
<th>Significance Level Between Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.53</td>
<td>822</td>
<td>.001</td>
</tr>
<tr>
<td>Female</td>
<td>2.67</td>
<td>776</td>
<td></td>
</tr>
<tr>
<td>Focus and Organization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.90</td>
<td>822</td>
<td>.091</td>
</tr>
<tr>
<td>Female</td>
<td>3.01</td>
<td>776</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.97</td>
<td>822</td>
<td>.174</td>
</tr>
<tr>
<td>Female</td>
<td>3.05</td>
<td>776</td>
<td></td>
</tr>
<tr>
<td>Conventions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5.72</td>
<td>822</td>
<td>.066</td>
</tr>
<tr>
<td>Female</td>
<td>5.87</td>
<td>776</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \((N = 1598)\).

### Results

#### Data Screening

The data were reviewed for missing student scores and no responses were excluded from the study. When the data screening was completed, 1,598 individual student scores were used in analysis \((n = 1598)\). The sample size of 1,598 met requirements for generalizability with a large effect size (Gall et al., 2007).

#### Assumptions Testing

Levene’s test of equality of error variances was used to test for the homogeneity of this data. Levene’s test was violated for the language category and the focus and organization category. Levene’s test for equality of variances was not found to be violated for the language category analysis, \(F(3,1594) = 1.04, p = 0.37\), and the focus and organization category, \(F(3,
This assumptions test demonstrated that the error of variance was not equal across groups. Two-way ANOVAs are robust to violating this assumption (Gall et al., 2007); however, the assumption was violated for the development and conventions category: $F(3, 1594) = 4.38, p = .004$ and $F(3, 1594) = 4.58, p = .003$. Table 3 illustrates the results of the Levene’s test of equality of error variances used for the data of this study.

Table 3

**Levene’s Test of Equality of Error Variance by Category of Assessment**

<table>
<thead>
<tr>
<th>Category</th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>4.38</td>
<td>3</td>
<td>1594</td>
<td>.004</td>
</tr>
<tr>
<td>Focus and Organization</td>
<td>2.43</td>
<td>3</td>
<td>1594</td>
<td>.063</td>
</tr>
<tr>
<td>Language</td>
<td>1.04</td>
<td>3</td>
<td>1594</td>
<td>371</td>
</tr>
<tr>
<td>Conventions</td>
<td>4.58</td>
<td>3</td>
<td>1594</td>
<td>.003</td>
</tr>
</tbody>
</table>

*Note.* Tests the null hypothesis that the error variance of the dependent variables is equal across groups.

**Analysis**

**Null Hypothesis 1**

A two-way ANOVA was used to determine whether a significant difference existed between 2016–2017 TNReady writing achievement scores of students who were administered only common formative assessments and students who received formative assessments designed by a single teacher. The mean scores were evenly distributed. A statistically significant relationship existed between the achievement test levels in development category scores on the
TNReady writing assessment: \( p < .05 \). The mean scores of the remaining dependent variables were evenly distributed and were greater than or equal to the common assessment mean score.

**Null Hypothesis 2**

A two-way ANOVA was used to determine whether a significant difference existed between 2016–2017 TNReady writing achievement scores of students who were administered only common formative assessments and students who received formative assessments designed by a single teacher. The mean scores were evenly distributed. A statistically significant relationship existed between the achievement test levels in the conventions category scores on the TNReady writing assessment: \( p < .05 \). The mean scores of the remaining dependent variables were evenly distributed and were greater than or equal to the common assessment mean score.

**Null Hypothesis 3**

A two-way ANOVA was used to determine whether a significant difference existed between 2016–2017 TNReady writing achievement scores of students who were administered only common formative assessments and students who received formative assessments designed by a single teacher. The mean scores were evenly distributed. A statistically significant relationship did not exist between the achievement test levels in the language category scores on the TNReady writing assessment: \( p > .05 \). The mean scores of the remaining dependent variables were evenly distributed and were greater than or equal to the common assessment mean score.
Null Hypothesis 4

A two-way ANOVA was used to determine whether a significant difference existed between 2016–2017 TNReady writing achievement scores of students who were administered only common formative assessments and students who received formative assessments designed by a single teacher. The mean scores were evenly distributed. A statistically significant relationship did not exist between the achievement test levels in the focus and organization categories scores on the TNReady writing assessment: \( p > .05 \). The mean scores of the remaining dependent variables were evenly distributed and were greater than or equal to the common assessment mean score.

Table 4 summarizes the two-way ANOVA category of assessment scores. The table represents the summary of the team-created and teacher-created scores on the TNReady summative assessment.

Table 4

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>3.54</td>
<td>1</td>
<td>3.54</td>
<td>4.65</td>
<td>.031</td>
</tr>
<tr>
<td>Focus and Organization</td>
<td>.127</td>
<td>1</td>
<td>.127</td>
<td>.065</td>
<td>.799</td>
</tr>
<tr>
<td>Language</td>
<td>.054</td>
<td>1</td>
<td>.054</td>
<td>.033</td>
<td>.856</td>
</tr>
<tr>
<td>Conventions</td>
<td>37.75</td>
<td>1</td>
<td>37.75</td>
<td>10.82</td>
<td>.001</td>
</tr>
</tbody>
</table>

*Note.* Tests the null hypothesis that a statistical significance exists between eleventh-grade students who took either team-created or individual teacher-created formative assessments.
Figure 1 represents the mean scores of each indicator of the rubric on the 2016–2017 TNReady writing assessment. The mean scores are grouped by the type of formative assessment administered during the semester. The mean scores did not vary greatly between team-created and teacher-created formative assessment scores except for the indicators of development and conventions.

Figure 1. Bar graph with TNReady writing score averages by type of formative assessment.

**Null Hypothesis 5**

A two-way ANOVA was used to test for statistical significance between male and female writing scores on the 2016–2017 TNReady writing assessment. The mean scores were evenly distributed. A statistically significant relationship exists between the development achievement test levels between male and female students who were administered team-created formative assessments on the TNReady writing assessment: \( p < .05 \). The mean scores of the remaining dependent variables (categories) were evenly distributed.
Null Hypothesis 6

A two-way ANOVA was used to test for statistical significance between male and female writing scores on the 2016–2017 TNReady writing assessment. The mean scores were evenly distributed. A statistically significant relationship did not exist between the conventions achievement test levels between male and female students who were administered team-created formative assessments on the TNReady writing assessment: \( p > .05 \). The mean scores of the remaining dependent variables were evenly distributed.

Null Hypothesis 7

A two-way ANOVA was used to test for statistical significance between male and female writing scores on the 2016–2017 TNReady writing assessment. The mean scores were evenly distributed. A statistically significant relationship did not exist between the language achievement test levels between male and female students who were administered team-created formative assessments on the TNReady writing assessment: \( p > .05 \). The mean scores of the remaining dependent variables were evenly distributed.

Null Hypothesis 8

A two-way ANOVA was used to test for statistical significance between male and female writing scores on the 2016–2017 TNReady writing assessment. The mean scores were evenly distributed. A statistically significant relationship did not exist between the focus and organization achievement test levels between male and female students who were administered team-created formative assessments on the TNReady writing assessment: \( p > .05 \). The mean scores of the remaining dependent variables were evenly distributed.
Table 5 summarizes the two-way ANOVA assessment by gender. The table represents the summary of the team-created scores of both male and female students on the TNReady summative assessment.

Table 5

Summary Table of Two-Way ANOVA: Type of Assessment by Gender

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>8.37</td>
<td>1</td>
<td>8.37</td>
<td>10.97</td>
<td>.001</td>
</tr>
<tr>
<td>Focus and Organization</td>
<td>5.65</td>
<td>1</td>
<td>5.65</td>
<td>2.86</td>
<td>.091</td>
</tr>
<tr>
<td>Language</td>
<td>3.00</td>
<td>1</td>
<td>3.00</td>
<td>1.84</td>
<td>.174</td>
</tr>
<tr>
<td>Conventions</td>
<td>11.76</td>
<td>1</td>
<td>11.76</td>
<td>3.37</td>
<td>.066</td>
</tr>
</tbody>
</table>

*Note.* Tests the null hypothesis that a statistical significance exists between male and female eleventh-grade students who took either team-created or individual teacher-created formative assessments.

Figure 2 represents the difference in mean scores between male and female scores for the development category of the rubric. Female students who were administered team-created formative assessment scored significantly higher in the development category than male students.
Figure 2. Box plot with of the development category of TNReady scores by gender.

Results for Null Hypothesis 1

Null hypothesis 1 was rejected for the development achievement score on the 2016–2017 TNReady writing assessment. A statistically significant difference existed at the \( p = .031 \) level in TNReady scores for the development indicator on the TNReady writing assessment: \( F(1, 1594) = 10.97, p = .031 \). The development mean scores on the common assessment \( (M = 2.67, SD = .875) \) were higher than the mean scores on the individual teacher-created assessment \( (M = 2.55, SD = .885) \).

Results for Null Hypothesis 2

Null hypothesis 2 was rejected for the conventions achievement score on the 2016–2017 TNReady writing assessment. A statistically significant difference existed at the \( p = .001 \) level in TNReady scores for the conventions indicator on the TNReady writing assessment: \( F(1, 1594) \).
= 10.82, p = .001. The conventions mean scores on the common assessment (M = 5.96, SD = 1.80) were higher than the mean scores on the individual teacher-created assessment (M = 5.64, SD = 1.95).

**Results for Null Hypothesis 3**

Null hypothesis 3 was accepted for the language achievement score on the 2016–2017 TNReady writing assessment. A statistically significant difference did not exist at the (p = .856) level in TNReady scores for the Language indicator on the TNReady writing assessment: F(1, 1594) = .033, p = .799. The language mean scores on the common assessment (M = 3.01, SD = 1.22) were equal to the mean scores on the individual teacher-created assessment (M = 3.01, SD = 1.32).

**Results for Null Hypothesis 4**

Null hypothesis 4 was accepted for the focus and organization achievement score on the 2016–2017 TNReady writing assessment. A statistically significant difference did not exist at the (p = .799) level in TNReady scores for the focus and organization indicator on the TNReady writing assessment: F(1, 1594) = .065, p = .799. The focus and organization mean scores on the common assessment (M = 2.95, SD = 1.34) were lower than the mean scores on the individual teacher-created assessment (M = 2.96, SD = 1.46).

**Results for Null Hypothesis 5**

Null hypothesis 5 was rejected for the development achievement on the 2016–2017 TNReady writing assessment. A statistically significant difference existed at the (p = .001) level in TNReady scores for the Development indicator on the TNReady writing assessment: F(1, 1594) = 10.97, p = .001. Female students’ development mean scores on the common assessment
were higher than male students’ mean scores on the common team-created assessment \((M = 2.49, SD = .922)\).

**Results for Null Hypothesis 6**

Null hypothesis 6 was accepted for the conventions achievement on the 2016–2017 TNReady writing assessment. A statistically significant difference did not exist at the \((p = .066)\) level in TNReady scores for the conventions indicator on the TNReady writing assessment: \(F(1, 1594) = 3.37, p = .066\). The conventions mean scores for females on the common assessment \((M = 6.27, SD = 1.65)\) were higher than the male mean scores on the common team-created assessment \((M = 5.64, SD = 1.89)\).

**Results for Null Hypothesis 7**

Null hypothesis 7 was accepted for the language achievement on the 2016–2017 TNReady writing assessment. A statistically significant difference did not exist at the \((p = .174)\) level in TNReady scores for the language indicator on the TNReady writing assessment: \(F(1, 1594) = 1.84, p = .174\). Female students’ language mean scores on the common assessment \((M = 3.13, SD = 1.17)\) were higher than male students’ mean scores on the common team-created assessment \((M = 2.88, SD = 1.26)\).

**Results for Null Hypothesis 8**

Null hypothesis 8 was accepted for the focus and organization achievement on the 2016–2017 TNReady writing assessment. A statistically significant difference did not exist at the \((p = .091)\) level in TNReady scores for the focus and organization indicator on the TNReady writing assessment: \(F(1, 1594) = 2.86, p = .091\). Female students’ focus and organization mean
scores on the common assessment \((M = 3.09, SD = 1.26)\) were higher than male students’ mean scores on the common team-created assessment \((M = 2.80, SD = 1.4)\).

**Summary**

This study had two purposes. The first purpose was to determine if a difference in achievement levels existed on the 2016–2017 TNReady writing assessment between eleventh-grade students who took common assessments (team-created) and students who were administered individual assessments (teacher-created). The study results demonstrated a statistically significant difference in two categories of the writing rubric: development and conventions. The null hypotheses were accepted for the other two reporting categories on the TNReady writing assessment: language and focus and organization.

The second purpose of this study was to determine if a difference existed in the category scores between male and female students who took common (team-created) assessments as determined by TNReady writing achievement levels. This null hypothesis was rejected for development. Female students in this sample demonstrated a significantly higher achievement score in the development category on TNReady. These results indicated that female students scored significantly higher than male students in one out of four categories in the TNReady writing test. Neither gender scored significantly higher than the other in the other three writing domains. The null hypotheses were accepted for conventions, language, and focus and organization.
CHAPTER FIVE: CONCLUSIONS

Overview

Chapter 5 presents a discussion of the study findings presented in Chapter 4. Each of the eight hypotheses are discussed. Chapter 5 also presents implications of the study results, the limitations of the study, and recommendations for future research.

Discussion

The purpose of this study was to examine the differences in TNReady writing scores among eleventh-grade students who took team-created common formative writing assessments at least twice a semester and students who took teacher-created formative assessments at least twice a semester. School officials who understand which type of formative assessment better supports student achievement on end-of-course state testing can make more informed classroom and formative assessment decisions. Teachers, school, and district leaders can use the results of this study to plan and prepare both classroom instruction and develop assessment resources. All stakeholders can benefit from understanding how to best prepare students to achieve higher scores on state-mandated end-of-course assessments.

The study used end-of-year TNReady writing scores for the classes from two groups of eleventh-grade students during the 2016–2017 school year. Gender and types of assessment, either team-created common formative or teacher-created formative assessments were considered in this study. The dependent variables were development, focus and organization, language, conventions, and gender. The dependent variables were identified from the state writing rubric scoring categories and were reported on the students’ score report provided by the state of Tennessee.
Null Hypothesis 1

Null hypothesis 1 was rejected. The category of development on the state writing rubric evaluates students’ abilities to choose relevant evidence from a text to support their response and then elaborate on that response. The development category requires students to pull information from a text to support their response to the prompt or complete a written task for the writing assessment. The goals of the development category on the rubric directly relate to the goals of the recently implemented CCSS. These standards place an emphasis on choosing appropriate evidence from the text to support propositions made by the student. TNReady is a high-stakes assessment; therefore, teachers may engage in instruction that directly relates to the tested standard (Popham et al., 2014; Stobart & Eggen, 2012). Teachers generally have an inclination to teach to the assessment and narrow the focus of instruction to highlight the purpose of the assessment (Olinghouse et al., 2012). The development category’s standard of supporting evidence is a relatively new writing trend in the state of Tennessee. This was a major shift in how writing was taught in that creative responses were no longer celebrated and writing had a purpose other than to express one’s view. Writers had to explain how they arrived at their view of a topic and support their responses. This shift in writing has been embraced more by secondary teachers than earlier grade ELA teachers, which may impact essay development at the high school level.

Null Hypothesis 2

Null hypothesis 2 was rejected. The rubric category of conventions addressed the need to use correct grammar and engage in as few grammatical mistakes as possible. The conventions of grammar are generally relatable to rules of grammar and usage. Some teachers may have a lack of professional knowledge concerning writing pedagogy; this lack of content knowledge may
influence the amount of time teachers devote to teaching grammar (conventions; Herman et al., 2015). The conventions category accounts for 25% of the students’ TNReady writing achievement score. Teacher teams may consider addressing the conventions category through direct instruction because this category counts as one fourth of the students’ score and is dependent on rules of grammar. Teachers could feel the need to teach to the test in this category (Brimi, 2012; Tuttle, 2013). Conventions are most often associated with instruction in the earlier grades. High school teachers may focus more time on the development and organization of essays rather than on conventions. Therefore, it is probable that teacher teams creating common assessments put a greater emphasis on conventions because these assessments are created and vetted by a team of teachers (Gillespie et al., 2014, Harris et al., 2013; Mulatash, 2011).

**Null Hypothesis 3**

Null hypothesis 3 was accepted. The language indicator evaluates students’ ability to use vocabulary, sentence variety, and tone when writing an essay. A vertical alignment exists in both the standards and the writing rubric in the language category at the elementary, middle, and high school grade levels. The expectations and requirements on the standards and writing rubrics are consistently increased as students progress through the grade levels. Summative end-of-course assessments are considered high stake assessments. Summative scores influence district decisions and are reported to stakeholders; thus, it is possible that teachers focus their instruction on preparing students to score well on the assessment rather than demonstrating mastery of the content (Beaver & Weinbaum, 2012; Shepard et al., 2016).
Null Hypothesis 4

Null hypothesis 4 was accepted for focus and organization as reported by the TNReady writing scores of this sample of eleventh-grade students. In the focus and organization category, students are scored on their ability to draft a coherent essay with effective beginning and concluding statements combined with an organizational structure that orients the reader to the argument, explanation, or narrative structure of the essay.

Null Hypothesis 5

Null hypothesis 5 was rejected. Female students scored higher than males in the development category of the TNReady writing assessment. This finding was consistent with the literature. Female students have been found to consistently score higher than males in writing in general. The study findings were consistent with the literature concerning female students outperforming male students on summative achievement tests (Reynolds et al., 2015; Scheiber et al., 2015). Female students consistently outperformed males in the areas of sentence composition and essay composition (Pargulski & Reynolds, 2017. The development category of the TNReady writing assessment required students to read a passage and write an essay in response to a prompt. Students were also asked to use evidence from the passage to support their response. The skills required of the development category relate to both fluency and reading ability. Female students have also had consistently higher NAEP reading scores than male students and female students develop verbal fluency at a faster and more proficient rate than male students (Reilly, Neumann, & Andrews, 2019).
Null Hypothesis 6

Null hypothesis 6 was accepted. The conventions category remains identical on the TNReady writing rubric regardless of the grade level and the standards remain consistent from the middle school level through high school; conversely, the other three categories have varying expectations and align vertically as students move through the grade levels (CCSS initiative, 2010; Tennessee Department of Education, 2018). The vertical alignment of the standards and increasing expectations for student writing on the TNReady writing rubric may explain the results of null hypothesis 6. Students who were administered team-created common assessments scored higher on the one category that remained unchanged throughout their high school career. Vertical alignment of the standards can influence students’ knowledge of the writing process along with students’ ability to produce written expression (Graham & Herbert, 2011; Troia & Olinghouse, 2015).

Null Hypothesis 7

Null hypothesis 7 was accepted. The language category of the state writing rubric scores students on their ability to produce writing that conveys appropriate style and tone and sentence variety along with writing for meaning and reader interest and using sensory language. Female students’ mean scores were higher than male mean scores on the 2017 TNReady writing assessment; however, the differences were not statistically significant. Females typically develop reading fluency and the ability to express themselves in writing at an earlier age than males. The early development in reading fluency translates to writing, as females have more experience expressing their thoughts through writing (Petersen, 2018). Females are also more likely to express their feelings in writing and generally use words that more broadly express ideas,
concepts, and theories. Males generally rely on word choice that is directly related to the topic or task presented (Ishikawa, 2015).

Null Hypothesis 8

Null hypothesis 8 was accepted. The focus and organization category of the state writing rubric evaluates students on their ability to create a relevant, well-organized, and coherent essay based on the text provided on the TNReady writing assessment. This category of the rubric requires students to express their thoughts while still drawing on information from the text; thus, students must comprehend the text, task, and then transfer those thoughts to an organized essay. Males are generally less willing to provide information in their writing. The transference of limited information to the essay hinders male students’ ability to write a coherent essay. Female students frequently provide a larger context for the topic in their writing, which leads to more coherence and a more meaningful production of writing (Ishikawa, 2015). Adams et al. (2019) found that males struggle more often with transferring their thoughts to writing. Adams et al. also noted that males have a negative attitude toward writing in general. More than any other category, the focus and organization category of the state writing assessment directly reflects students’ ability to formulate an essay.

Implications

This study contributed to the larger body of formative assessment literature by addressing a gap regarding measuring writing achievement in quantitative analysis. Writing is generally viewed from a qualitative method of research (Tuttle, 2013). Few studies have analyzed writing data quantitatively; most studies have focused on the product rather than the instructional process of teaching writing (Graham et al., 2011). The purpose of this study was to determine if the type
of formative assessment administered to students—whether common or individual teacher-created—influenced the summative assessment results on the TNReady state writing assessment. The study results demonstrated that a relationship does exist between the type of formative assessment administered (teacher team-generated common formative assessments or individual teacher-generated formative assessment) and achievement levels in some categories on the state writing assessment.

The general consensus in the literature is that formative assessment increases student achievement scores (Black & Wiliam, 2009; Wiley & Lyon, 2015). A majority of teachers report using some type of formative assessment in their classrooms (Abrams et al., 2015). This study addressed the difference between two of the most frequently used types of formative assessment: team-created and individual teacher created formative assessments. The study findings demonstrated that a relationship does exist between common assessments created by teacher-teams and higher summative achievement levels by students. The study results indicated that students who were administered common assessments achieved a higher score on the conventions category and female students scores higher on the development category of the state writing assessment. These results imply that the type of assessment given over the course of a semester can influence the results of state summative testing. This finding is consistent with previous literature and may help district leaders to develop professional development opportunities and plan for creating formative assessments for teachers in their districts. Perhaps districts could train English teachers on creating team-created writing assessments. The policy implication for these findings would suggest incorporating team-created common assessments into the instruction.
**Limitations**

This study has limitations regarding internal and external validity. The common assessments were developed at the district level either by curriculum supervisors and the ELA instructional coach or teams of teachers. The assessments were aligned to both Tennessee state standards and the district curricula (Tennessee Department of Education, 2015). The items included in the formative assessments were selected from the prompts and tasks provided by the state summative assessment testing company. The individual teacher-created formative writing assessments were created by individual teachers and were administered for instructional purposes. These teacher-created assessments may not have been vetted as thoroughly for validity as team-created assessments.

None of the schools examined in this study administered both common and individual teacher-created formative assessments. The expectations regarding the assessment, planning, and administration of formative assessments would be more consistent if both types of formative assessments were administered at each of the eight schools in the sample. Class periods were not the same length despite all classes running on a block schedule. Standards varied for teacher-made tests and common assessments that limit question types, number of questions, or content. Every school in this study did not administer both types of formative assessments at multiple points during the semester; thus, it is possible that bias or error may have occurred.

Another limitation to this study was the unequal population of students included in the study. While an equal number of schools participated in either common or teacher-created formative writing assessments, more students participated in teacher-created assessments due to the population of the schools. In this study, 780 students were administered common assessments and 818 students were administered teacher-created assessments. Also, the data analyzed in this
study were from the 2016–2017 school year. This archival data was relevant and consistent with prior literature; however, it was not recent data.

**Recommendations for Future Research**

The first recommendation for future studies includes the need to conduct a true experiment of administering both common and individual teacher-created formative assessments in a classroom setting. The sample used for this study was large \( n = 1,598 \) and included eight schools. An experiment in which both types of formative assessments are administered in a school or district would provide more insight and a deeper understanding of how formative assessment prepares students to perform on summative achievement assessments. An experimental study would also provide teachers the opportunity to gather longitudinal data as students participate in both types of formative assessment across grade levels.

Furthermore, additional studies should examine how common formative assessments are created and how common formative assessments are used to inform instructional practice in the classroom. The purpose of this study focused solely on the administration of the assessment and its possible influence on student achievement levels on the state summative writing assessment. It would be interesting to know how teachers who are part of teacher-teams describe their experiences in writing state standards and creating the state writing rubric. It would also be interesting to explore the process by which common assessments are created, and identify if a relationship exists between how the assessments are created and achievement levels on state summative assessments. It would be helpful to explore how teachers provide feedback to students concerning their common formative assessment scores and identify how often this feedback occurs during the course of a semester. Future studies could focus on the factors surrounding the creation of common formative assessments and how the data are used to inform
instruction. Common formative assessments are created in teacher teams; thus, additional research should focus on those teachers’ knowledge and application of the state writing standards.

**Summary**

Chapter 5 addressed the two research questions and provided a discussion of the study results. The first null hypothesis was rejected for the reporting category of development. The second null hypothesis was rejected for the reporting category of conventions. Students who were administered common assessments had higher achievement scores on the summative end-of-course assessment in two out of four writing rubric categories. The fifth null hypothesis was rejected for development category in terms of gender. Female students who were administered common formative assessments demonstrated higher achievement scores in the development category of the summative end-of-course assessment than male students. Chapter 5 also included implications of the findings in addition to recommendations for future research.
REFERENCES


Jones, S. M. (2012). Mapping the landscape: Gender and the writing classroom. *Journal of Writing Research, 3*, 161–179. doi:10.17239/jowr-2012.03.03.2


Konstantopoulos, S., Li, W., Miller, S. R., & van der Ploeg, A. (2017). *Do interim assessments reduce the race and SES achievement gaps?* Retrieved from ERIC database. EJ1139993


October 18, 2019

Justin Calhoun
IRB Application 3939: A Comparative Study of Teacher-Team Generated Formative Assessments on Student Writing Achievement Test Levels

Dear Justin Calhoun,

The Liberty University Institutional Review Board has reviewed your application in accordance with the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and finds your study does not classify as human subjects research. This means you may begin your research with the data safeguarding methods mentioned in your IRB application.

Your study does not classify as human subjects research because it will not involve the collection of identifiable, private information.

Please note that this decision only applies to your current research application, and any changes to your protocol must be reported to the Liberty IRB for verification of continued non-human subjects research status. You may report these changes by submitting a new application to the IRB and referencing the above IRB Application number.

If you have any questions about this determination or need assistance in identifying whether possible changes to your protocol would change your application’s status, please email us at irb@liberty.edu.

Sincerely,

[Signature]
Administrative Chair of Institutional Research
Research Ethics Office

Liberty University | Training Champions for Christ since 1971
### APPENDIX B: EXPLANATORY WRITING RUBRIC

#### TNReady Grades 9-12 Informational/Explanatory Rubric

Revised: May 2017

<table>
<thead>
<tr>
<th>Score</th>
<th>Focus &amp; Organization</th>
<th>Development</th>
<th>Language</th>
<th>Conventions</th>
</tr>
</thead>
</table>
| 4     | In response to the task and the stimuli, the writing:  
utilizes effective organizational strategies to create a unified whole and to aid in comprehension,  
effectively clarifies relationships among ideas and concepts to create cohesion,  
contains an effective and relevant concluding statement or section. | In response to the task and the stimuli, the writing:  
utilizes well-chosen, relevant, and sufficient evidence from the stimuli to thoroughly and insightfully develop the topic.  
comprehensively and accurately explains and elaborates on the evidence provided, demonstrating a clear, insightful understanding of the topic, task, and stimulus. | The writing:  
illustrates consistent and sophisticated command of precise language, domain-specific vocabulary, and literary techniques appropriate to the task.  
illustrates sophisticated command of syntactic variety for meaning and reader interest.  
utilizes sophisticated and varied transitional words and phrases.  
effectively establishes and maintains a formal style and an objective tone. | The writing:  
illustrates consistent and sophisticated command of grade-level conventions of standard written English.  
may contain a few minor errors that do not interfere with meaning. |
| 3     | In response to the task and the stimuli, the writing:  
contains a relevant introduction.  
utilizes effective organizational strategies to create a mostly unified whole and to aid in comprehension,  
clarifies most relationships among ideas and concepts, but there may be some gaps in cohesion,  
contains a relevant concluding statement or section. | In response to the task and the stimuli, the writing:  
utilizes relevant and sufficient evidence from the stimuli to adequately develop the topic.  
adapts and accurately explains and elaborates on the evidence provided, demonstrating a sufficient understanding of the topic, task, and stimulus. | The writing:  
illustrates consistent command of precise language, domain-specific vocabulary, and literary techniques appropriate to the task.  
illustrates consistent command of syntactic variety for meaning and reader interest.  
utilizes appropriate and varied transitional words and phrases.  
establishes and maintains a formal style and an objective tone. | The writing:  
illustrates consistent command of grade-level conventions of standard written English.  
contains occasional minor and/or major errors, but the errors do not significantly interfere with meaning. |
| 2     | In response to the task and the stimuli, the writing:  
contains a limited introduction.  
demonstrates an attempt to use organizational strategies to create some unification, but ideas may be hard to follow at times,  
clarifies some relationships among ideas and concepts, but there are lapses in focus,  
contains a limited concluding statement or section. | In response to the task and the stimuli, the writing:  
utilizes mostly relevant, but insufficient evidence from the stimuli to partially develop the topic. Some evidence may be inaccurate or repetitive.  
explains some of the evidence provided, demonstrating only a partial understanding of the topic, task, and stimuli. There may be some level of inaccuracy in the explanation. | The writing:  
illustrates inconsistent command of precise language, domain-specific vocabulary, and literary techniques.  
illustrates inconsistent command of syntactic variety.  
uses basic or repetitive transitional words and phrases.  
establishes but inconsistently maintains a formal style and an objective tone. | The writing:  
illustrates inconsistent command of grade-level conventions of standard written English.  
contains frequent errors that may significantly interfere with meaning. |
| 1     | In response to the task and the stimuli, the writing:  
contains no or an irrelevant introduction,  
demonstrates an unclear organizational structure, ideas are hard to follow most of the time,  
fails to clarify relationships among ideas and concepts; concepts are unclear and/or there is a lack of focus,  
contains no or an irrelevant concluding statement or section. | In response to the task and the stimuli, the writing:  
utilizes mostly irrelevant or no evidence from the stimuli, or mostly only personal knowledge, to inadequately develop the topic. Evidence is inaccurate or repetitive.  
inadequately or inaccurately explains the evidence provided, demonstrating little understanding of the topic, task, and stimulus. | The writing:  
illustrates little to no use of precise language, domain-specific vocabulary, and literary techniques.  
illustrates little to no syntactic variety.  
utilizes no or few transitional words and phrases.  
does not establish or maintain a formal style and an objective tone. | The writing:  
illustrates limited command of grade-level conventions of standard written English.  
contains numerous and repeated errors that seriously impede meaning. |

1. Evidence includes facts, extended definitions, concrete details, quotations, or other information and examples as appropriate to the task and the stimuli.  
2. Domain-specific vocabulary refers to the terminology used in the stimuli and/or associated with the topic.  
3. Literary techniques, such as metaphor, simile, and analogy, help to manage the complexity of the topic and are expected at grades 11-12.  
4. Conventions of standard written English include sentence structure, grammar, usage, spelling, capitalization, and punctuation.
## APPENDIX C: ARGUMENTATIVE WRITING RUBRIC

### TNReady Grades 9-12 Argument Rubric

Revised: May 2017

<table>
<thead>
<tr>
<th>Score</th>
<th>Focus &amp; Organization</th>
<th>Development</th>
<th>Language</th>
<th>Conventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>The writing:</td>
<td>The writing:</td>
</tr>
<tr>
<td></td>
<td>• contains an effective and relevant introduction.</td>
<td>• utilizes well-chosen, relevant, and sufficient evidence from the stimuli to thoroughly and insightfully support logical claim(s) and counterclaim(s), while acknowledging strengths and limitations of both.</td>
<td>• illustrates consistent and sophisticated command of precise language and domain-specific vocabulary appropriate to the task.</td>
<td>• demonstrates consistent and sophisticated command of grade-level conventions of standard written English.</td>
</tr>
<tr>
<td></td>
<td>• states a precise claim and maintains a sophisticated argument.</td>
<td>• thoroughly and accurately explains and elaborates on the evidence provided, connecting the evidence to claim(s) and counterclaim(s) and demonstrating a clear, insightful understanding of the topic, task, and stimuli.</td>
<td>• utilizes sophisticated and varied transitional words and phrases.</td>
<td>• may contain a few minor errors that do not interfere with meaning.</td>
</tr>
<tr>
<td></td>
<td>• utilizes effective organizational strategies to logically sequence claim(s), counterclaim(s), reasons, and evidence to create a unified whole.</td>
<td>• effectively clarifies relationships among claim(s), reasons, evidence, and counterclaim(s) to create cohesion.</td>
<td>• effectively establishes and maintains a formal style and an objective tone.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• effectively clarifies relationships among claim(s), reasons, evidence, and counterclaim(s) to create cohesion.</td>
<td>• contains an effective and relevant concluding statement or section.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>The writing:</td>
<td>The writing:</td>
</tr>
<tr>
<td></td>
<td>• contains a relevant introduction.</td>
<td>• utilizes relevant and sufficient evidence from the stimuli to adequately support claim(s) and counterclaim(s), while acknowledging strengths and limitations of both.</td>
<td>• illustrates consistent command of syntactic variety for meaning and reader interest.</td>
<td>• demonstrates consistent command of grade-level conventions of standard written English.</td>
</tr>
<tr>
<td></td>
<td>• states a precise claim and maintains a clear argument.</td>
<td>• adequately and accurately explains and elaborates on the evidence provided, connecting the evidence to claim(s) and counterclaim(s) and demonstrating a sufficient understanding of the topic, task, and stimuli.</td>
<td>• utilizes appropriate and varied transitional words and phrases.</td>
<td>• contains occasional minor and/or major errors, but the errors do not significantly interfere with meaning.</td>
</tr>
<tr>
<td></td>
<td>• utilizes adequate organizational strategies to logically sequence claim(s), counterclaim(s), reasons, and evidence to create a mostly unified whole.</td>
<td>• clarifies most relationships among claim(s), reasons, evidence, and counterclaim(s), but there may be some gaps in cohesion.</td>
<td>• establishes and maintains a formal style and an objective tone.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• clarifies most relationships among claim(s), reasons, evidence, and counterclaim(s), but there may be some gaps in cohesion.</td>
<td>• contains a relevant concluding statement or section.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>The writing:</td>
<td>The writing:</td>
</tr>
<tr>
<td></td>
<td>• contains a limited introduction.</td>
<td>• utilizes mostly relevant but insufficient evidence from the stimuli to partially support claim(s) and counterclaim(s). Some evidence may be inaccurate or repetitive.</td>
<td>• illustrates inconsistent command of syntactic variety.</td>
<td>• demonstrates inconsistent command of grade-level conventions of standard written English.</td>
</tr>
<tr>
<td></td>
<td>• states a weak argument.</td>
<td>• explains some of the evidence provided, connecting some of the evidence to claim(s) and counterclaim(s) and demonstrating only a partial understanding of the topic, task, and stimuli. There may be some level of inaccuracy in the explanation.</td>
<td>• utilizes basic or repetitive transitional words and phrases.</td>
<td>• contains frequent errors that may significantly interfere with meaning.</td>
</tr>
<tr>
<td></td>
<td>• demonstrates an attempt to use organizational strategies to sequence claim(s), counterclaim(s), reasons, and evidence, but ideas may be hard to follow at times.</td>
<td>• clarifies some relationships among claim(s), reasons, evidence, and counterclaim(s), but there are lapses in focus.</td>
<td>• establishes but inconsistently maintains a formal style and an objective tone.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• clarifies some relationships among claim(s), reasons, evidence, and counterclaim(s), but there are lapses in focus.</td>
<td>• contains a limited concluding statement or section.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>In response to the task and the stimuli, the writing:</td>
<td>The writing:</td>
<td>The writing:</td>
</tr>
<tr>
<td></td>
<td>• contains no or an irrelevant introduction.</td>
<td>• utilizes mostly irrelevant or no evidence from the stimuli, or mostly/personal knowledge to inadequately support claim(s) and counterclaim(s). Evidence is inaccurate or repetitive.</td>
<td>• illustrates little to no use of precise language and domain-specific vocabulary.</td>
<td>• demonstrates limited command of grade-level conventions of standard written English.</td>
</tr>
<tr>
<td></td>
<td>• states an unclear argument.</td>
<td>• inadequately or inaccurately explains the evidence provided; evidence, claim(s), and counterclaim(s) appear disconnected, demonstrating little understanding of the topic, task, and stimuli.</td>
<td>• illustrates little to no syntactic variety.</td>
<td>• contains numerous and repeated errors that seriously impede meaning.</td>
</tr>
<tr>
<td></td>
<td>• demonstrates an unclear organizational structure; ideas are hard to follow most of the time.</td>
<td>• fails to clarify relationships among claim(s), reasons, evidence, and counterclaim(s). concepts are unclear and/or there is a lack of focus.</td>
<td>• utilizes no or few transitional words and phrases.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• fails to clarify relationships among claim(s), reasons, evidence, and counterclaim(s).</td>
<td>• contains no or an irrelevant concluding statement or section.</td>
<td>• does not establish or maintain a formal style and an objective tone.</td>
<td></td>
</tr>
</tbody>
</table>

1 Evidence includes facts, definitions, concrete details, quotations, or other information appropriate to the task and stimuli.

2 Domain-specific vocabulary refers to the terminology used in the stimuli and/or associated with the topic.

3 Conventions of standard written English include sentence structure, grammar, usage, spelling, capitalization, and punctuation.
APPENDIX D: NARRATIVE WRITING RUBRIC

TNReady Grades 9-12 Narrative Rubric
Revised: May 2017

<table>
<thead>
<tr>
<th>Score</th>
<th>Focus &amp; Organization</th>
<th>Development</th>
<th>Language</th>
<th>Conventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>In response to the task and the stimulus, the writing: • effectively utilizes relevant narrative techniques, such as dialogue, pacing, description, reflection, and/or multiple plot lines, in order to thoroughly and insightfully develop experiences, events, and/or characters. • effectively incorporates relevant, well-chosen details from the stimulus. • effectively demonstrates a clear, insightful understanding of the task and stimulus by using relevant, well-chosen, descriptive details in order to convey a vivid picture of the experiences, events, setting, and/or characters.</td>
<td>The writing: • illustrates consistent and sophisticated command of precise language, including sensory language. • illustrates sophisticated command of syntactic variety for meaning and reader interest. • utilizes sophisticated and varied transitional words and phrases. • effectively establishes and maintains an appropriate style and tone.</td>
<td>The writing: • demonstrates consistent and sophisticated command of grade-level conventions of standard written English.1 • may contain a few minor errors that do not interfere with meaning.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>In response to the task and the stimulus, the writing: • adequately engages and orientates the reader by setting out a relevant problem, situation, or observation, establishing a point(s) of view, and introducing a narrator and/or character. • utilizes adequate organizational strategies to establish a sequence of events and/or experiences that build on one another to create a coherent whole. • contains an effective conclusion that follows from and reflects on what is experienced, observed, or resolved over the course of the narrative.</td>
<td>The writing: • illustrates consistent command of precise language, including sensory language. • illustrates command of syntactic variety for meaning and reader interest. • utilizes appropriate and varied transitional words and phrases. • establishes and maintains an appropriate style and tone.</td>
<td>The writing: • demonstrates consistent command of grade-level conventions of standard written English.1 • contains occasional minor and/or major errors, but the errors do not significantly interfere with meaning.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>In response to the task and the stimulus, the writing: • conveys a limited, possibly confusing, situation, observation, or problem that may include a point(s) of view, a narrator, and/or character. • contains a limited sequence of events and/or experiences that may be confusing or contain gaps that interfere with the progression of events and/or experiences. • contains a weak conclusion that may be only loosely related to the narrated events or experiences.</td>
<td>The writing: • illustrates inconsistent command of precise and sensory language. • illustrates inconsistent command of syntactic variety. • utilizes basic or repetitive transitional words and phrases. • establishes but inconsistently maintains an appropriate style and tone.</td>
<td>The writing: • demonstrates inconsistent command of grade-level conventions of standard written English.1 • contains frequent errors that may significantly interfere with meaning.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>In response to the task and the stimulus, the writing: • contains an unclear, irrelevant, or no situation, observation, problem, or point of view. • contains no or an ineffective sequence of events and/or experiences that may be brief, confusing, or very hard to follow. • contains no or an irrelevant conclusion.</td>
<td>The writing: • illustrates little to no use of precise language and sensory language. • illustrates little to no syntactic variety. • utilizes no or few transitional words and phrases. • does not establish or maintain an appropriate style and tone.</td>
<td>The writing: • demonstrates limited command of grade-level conventions of standard written English.1 • contains numerous and repeated errors that seriously impede meaning.</td>
<td></td>
</tr>
</tbody>
</table>

1 Conventions of standard written English include sentence structure, grammar, usage, spelling, capitalization, and punctuation.