2014

WV Writes and Westest 2 online writing: the impact of writing prompts on student writing proficiency

Stacey Murrell
slmurrell@k12.wv.us

WV WRITES AND WESTEST 2 ONLINE WRITING: THE IMPACT OF WRITING PROMPTS ON STUDENT WRITING PROFICIENCY

Stacey Murrell

Dissertation submitted to the faculty of the
Marshall University Graduate College in
partial fulfillment of the requirements
for the degree of

Doctor of Education
in
Curriculum and Instruction

Committee Chair: Lisa A. Heaton, Ph.D.
Terrence Stange, Ph.D.
Sam Securro, Ed.D.
Vaughn Rhudy, Ed.D.

Marshall University College of Education and Professional Development
South Charleston, West Virginia, 2014

ACKNOWLEDGMENTS

I would like to begin by acknowledging Dr. Fred Pauley, my first doctoral committee chair. His warm, easygoing, and accepting spirit made me believe in myself and regain my self-confidence as a scholar, researcher, and doctoral student. At a time when I was rebounding from personal hardships and struggling to regain my footing in the doctoral program, he accepted me and agreed to become my committee chair. I regret that he cannot be here to celebrate the end of this journey with me in person. I appreciate the time I had with him.

Dr. Lisa Heaton did not hesitate as she reached out to assist me and offered to be my committee chair when Dr. Fred Pauley passed away. I would have been lost without her kindness and willingness to aid me despite increasing her workload by taking the lead position as my committee chair. While my prospectus was nearly completed, her guidance brought this work to its conclusion. I would not have completed this work without her support and guidance. Her honesty ensured that I offered the best iteration of my work, and her belief that I could achieve my best and that this study just needed revisions in order to reach its conclusion kept me motivated.

Dr. Sam Securro became a member of my committee after the death of Dr. Fred Pauley and I am thankful that he was willing to adopt me as another doctoral student under his guidance. His patience with me as I navigated through a quantitative study and asked him question after question was boundless. He willingly offered his expertise as a statistician and his time as he worked with me during the preliminary stages of setting up the research study design, transferring the data into SPSS and analyzing the results. When I had questions or needed

clarification, he would offer his thoughts and explanations, and I always walked away with a better understanding of the processes that had confused me.

Dr. Terrence Stange provided his expertise in the areas of reading, writing, and literacy as he served on my doctoral committee. His knowledge and experience as a researcher were invaluable as he dedicated his time verifying that this work was academically and structurally correct. His thoughtful and reflective feedback helped me to consider and evaluate the purpose and intent of this work and its effectiveness.

Dr. Vaughn Rhudy fostered my academic efforts throughout this dissertation. I was able to work collaboratively with him and also have his expertise as an assessment coordinator. His extensive knowledge related to artificial intelligence scoring, online writing assessment, and the teaching and assessing of writing allowed me to have daily support and guidance. His kindness and thoughtfulness to me as we worked together and his willingness to act as a mentor academically, professionally, and personally were essential to the successful completion of this work.

Lastly, my parents, Nancy and Gary Murrell, must be acknowledged as well. They have always believed in me with an unwavering faith and love and this created in me the confidence that I needed to complete the doctoral program and this work. They made me the strong, confident, successful, and educated woman that I am today. For that, I am grateful. Not everyone can have the support that my parents gave me. I have achieved a goal that I envisioned for myself some 22 years ago as a senior in high school. This dissertation serves as evidence of the hard work, dedication, and perseverance I possessed after I embarked on this academic journey as a doctoral student in the fall of 2007.

DEDICATION

This work is dedicated to my daughter, Hannah. The life we have together is the secret place of happiness that I retreat to in order to escape the stresses of life. I hope that I will inspire her as much as she inspires me. Her existence motivates me to reach for my highest potential. I also dedicate this work to my grandmother, Lillian. Her recent death prevents her from being here to celebrate this achievement; however, her presence is in my heart.

ABSTRACT

This study's purpose was to determine the effects of students practicing writing using practice writing prompts prior to completing the summative state writing assessment. It adds data to the body of knowledge related to the use of practicing writing using practice prompts prior to students taking a high stakes state-level writing assessment. The type of research design used for this study was a quantitative, post hoc, 2 x 2 ANOVA. The data were obtained from the WESTEST 2 Online Writing composite scores and the five analytic trait scores that comprise the WESTEST 2 Online Writing scores. The study had a population of 6,459 11th grade students enrolled in West Virginia public schools. These students had all taken the WESTEST 2 Online Writing as 11th graders in the spring of 2013, and in preparation for the year-end, state level writing assessment completed either Writing Roadmap 2.0 prompts or WESTEST 2 practice prompts. Using random sampling, 190 students who wrote essays using WESTEST 2 practice prompts and 190 students who wrote essays using Writing Roadmap prompts were selected from the student population. This gave a total of 380 students in the sample size. Findings revealed that no significant effects were found when using one type of writing prompt over another on composite writing scores or on the five analytic writing scores. However, significance was demonstrated (p. 000) with the scores of females being greater than male students. Results gave stakeholders evidence that students who had a generic writing prompt versus a mirror image of the high stakes writing assessment scored no better or worse than the other. The new assessment vendors, states, counties, schools, and teachers will all benefit from these study findings as new assessment systems are adopted based on Common Core writing standards across the nation. The results are critical in supporting the discrepancy that stands between females and males and their

writing proficiencies. This study can support efforts that focus on addressing the inequalities and ensuring that the discrepancy is removed and male students become better writers.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In the fall of 2010, the West Virginia Department of Education (WVDE) released a practice writing program called WV Writes, an online program that provided formative writing assessments designed to inform teaching and improve student learning. WV Writes replaced the previous formative writing program called Writing Roadmap 2.0. Students in West Virginia public schools, grades 3-11, are required to take an online writing assessment called the West Virginia Educational Standards Test 2 or WESTEST 2 Online Writing. WV Writes is considered to be more closely aligned to the summative writing assessment than Writing Roadmap 2.0 in both prompt structure and scoring; therefore, WV Writes was additionally purchased by the WVDE and is provided free to all students as a practice writing test.

WV Writes was released in 2010 and will continue to be a state-sponsored practice writing program for West Virginia students until the end of the testing contract in 2014. The WVDE does not mandate its use in the schools; however, feedback from counties has indicated that schools with high usage of WV Writes have seen their Reading Language Arts (RLA) WESTEST 2 scores rise significantly because of high WESTEST 2 Online Writing scores. Because the WESTEST 2 Online Writing scores are calculated into the RLA WESTEST 2 scores, the higher writing scores have caused the overall increase in the total RLA scores. As students have used WV Writes, the question has emerged as to what effect on student test scores, if any, has occurred on WESTEST 2 Online Writing for those students who have used the WV Writes practice program prior to taking the summative writing test.

LITERATURE REVIEW

Prompt Structure and Construction

When reviewing the research conducted on the structure and construction of writing prompts and implications of these two characteristics, several findings were discovered. Writing prompt structures and the manner in which they were constructed have affected student perception, motivation, and writing ability. The more open a writing prompt was in its formatting, the more freedom students had to address the prompt (Furtak & Ruiz-Primo, 2008). When writing prompts were structured to provide students with a reading passage and then asked the students to complete a writing task, the students understood the requirements of the writing assignment best (Plakans, 2008; Plakans & Gebril, 2012). Students were motivated and interested in writing when the prompts were structured to engage them as readers and thinkers (Plakans & Gebril, 2012). Interestingly, while students may have believed they wrote better when they were motivated by the writing prompt, it appeared that their writing skill did not increase. Thus, even if students did not like a particular writing prompt, they still exhibited the same level of mastery in their writing ability (Olinghouse, Zheng, & Morlock, 2012).

Using Writing Prompts to Assess Writing

When using writing prompts to assess writing, several conclusions were made based on research studies. According to both Bridwell (1980) and Stoddard and MacArthur (1993), skilled writers participated in prewriting strategies more than struggling writers and spent little to no time revising their writing to ensure that it addressed the writing prompt. Struggling writers spent more time revising their essays to address the writing prompts; however, their scores did not improve. Students tapped into their metacognitive thought processes as they read writing prompts and formulated essays in response to the prompts (Nuckles, Hubner, & Renkl, 2009).

The type of prompt used to assess student writing aided and challenged students and their written attempts to address the prompt (Condon, 2004; McMaster & Campbell, 2008). Also, there were variables that may or may not have affected the proficiency scores of student writing when using writing prompts for assessment (Breland, 1983; Brown, 1986; Cherry & Meyer, 1993; Coffman, 1971; Nold & Freedman, 1977; Ruth & Murphy, 1988; White, 1985). When human scorers assessed student writing proficiency and their ability to address writing prompts, the person scoring the essays also impacted the scores (Schoonen, 2005). The type of writing prompt used to assess writing or being allowed to select a prompt were determined to not affect student proficiency on the writing task (Barry & Nielsen, 1997; Beck & Jeffery, 2007; Jennings, Fox, Graves, & Shohamy, 1999; Kobrin, Deng, & Shaw, 2011; Lee, 2008).

Gender and Writing Proficiency

When considering the variable of gender related to writing prompts and writing assessment related to affecting students' proficiency scores as their writing responses were scored, numerous studies found gender affected proficiency (Breland, 1983; Brown, 1986; Cherry & Meyer, 1993; Coffman, 1971; Nold & Freedman, 1977; Ruth & Murphy, 1988; White, 1985). Gabrielson, Gordon, and Engelhard (1995) discovered that writing prompt type had no statistical significance on writing scores; however, gender did have an effect on writing scores. Interestingly, James (2008) determined that Automated Essay Scoring (AES) programs used to assess writing skills did not show statistical significance based on gender. Both boys and girls were scored similarly regardless of gender.

Feedback and Reporting on Student Writing

When students received feedback and reporting on their writing, they were impacted in their writing proficiency. Positive feedback and encouragement on writing caused them to feel

confident and motivated about their writing and to express more than in their initial writing attempts (Lee & Laspe, 2003). Elbow (2000) found providing feedback to students on their writing must be timely and relevant to the occasion of the writing task in order to be effective. Other studies concluded with similar findings. Students and teachers must communicate regarding the findings of the feedback (Nicol & Mcfarlane-Dick, 2006). Also, effective feedback must be individualized for each student and writing occasion (Haswell, 2006). Feedback and reporting on student writing should point out strengths and weaknesses in the student writing; it should offer guidance to the students to address the weaknesses (Alter & Adkins, 2006). If students are given written feedback and reporting, then they need to be allowed revision opportunities and to see where their writing ranks on the scoring rubric (Zinn, 1998).

Analytic Trait Scoring and Holistic Scoring

Analytic trait scoring and holistic scoring were two methods of scoring used to assess student writing. When analytic trait scoring was used, the students benefited because they knew which areas of the writing needed to be improved and could make their revisions based on the analytic traits (Coe, 2000; East, 2009; Higgins, Miller, & Wegmann, 2007). Formative writing assessment benefited most when students had their writing scored using an analytic trait scoring rubric because the rubric acted as an instructional tool (Spandel, 2006). Interestingly, teachers believed that using a scoring rubric decreased subjective grading, but the reality was that using a rubric did not remove their natural subjectivity when assessing writing (Rezaei & Lovorn, 2010). Flateby (2010) found that rubrics used with analytic trait scoring improved student writing; holistic scoring from rubrics was better suited for summative writing assessments because it did not facilitate improving the writing. Attali, Lewis, and Steier (2013) discovered that automated

essay scoring and human scoring on a holistic rubric were maximized when they were used in conjunction with each other.

Automated Essay Scoring (AES) Systems

AES systems emerged as valid assessment modes in the 1990s. These programs were used summatively and formatively in educational institutions (Dikli, 2006; Warschauer & Grimes, 2008; Warschauer & Ware, 2006). Statistically, 95% of the time when automated scoring engines were compared, they were in agreement with human scorers (Cohen, Ben-Simon, & Hovav, 2003; Keith, 2003). Eliot and Mikulas (2004) found that using a formative AES system writing program improved student scores on the high stakes writing assessment. AES systems are best aligned to human scorers when they are built around prompt-specific specifications, rather than being built using a generic scoring engine (Attali, Bridgeman, & Trapani, 2010; Ramineni, 2013). When used in educational institutions for academic placement, studies proved that AES systems were as valid and reliable as their human counterparts (James, 2006; Klobucar, Elliot, Deess, Rudniy, & Joshi, 2013). While unable to comprehend essay content, several researchers argued that this ability to understand content will become a reality soon (Dean, 2013; Shermis, 2003).

Formative and Summative Writing Assessment Programs

Exposure to formative and summative writing assessment programs has grown for students as more states and school districts use these programs to determine the writing proficiency of their students. One formative AES system called Writing Roadmap 2.0 was found to directly improve student proficiency on high stakes writing assessment (Harrington, Kim, & West, 2008; Harrington &Rich, 2006; Rich, White, Hixon, D'Brot, & Perdue, 2010). Landauer, Lochbaum, and Dooley (2009) found that WritetoLearn, another AES formative writing

program, also positively impacted student results on summative writing assessments. Also, the impact of high stakes writing assessment was found to alter teacher perceptions of writing and writing instruction (Brimi, 2012). A disconnect between writing assessment and the philosophical views of writing theorists was found as well (White, 2004).

BACKGROUND

Summative Writing Assessment Prior to 2009

Table 1 indicates that West Virginia began assessing students' writing abilities in 1984. The writing assessment required students to handwrite their essays, and they were then hand scored by West Virginia Reading Language Arts teachers. Only students in grades 4, 7, and 10 were given a writing assessment until 2005. A computer-based writing assessment program was developed and introduced in 2006 called Online Writing Assessment. The Online Writing Assessment was field tested in 2007 with statistical findings indicating it was a valid and reliable assessment. In 2008, a field test was conducted for students in grades 3-11 that evaluated the summative writing program. The field test was successful, and it was determined that all West Virginia students in grades 3-11 would take the online writing assessment the next year. At the same time that the online writing assessment was scheduled to begin statewide, the release of the new state assessment WESTEST 2 was actualized. WESTEST 2 is not an online assessment program; it is a traditional paper/pencil standardized test that assesses Reading Language Arts, Math, Science, and Social Studies. The WESTEST 2 Online Writing became a component of the WESTEST 2 Reading Language Arts (RLA) scores which were then combined to generate students' overall proficiency in mastering the Reading Language Arts state content standards (A Chronicle of West Virginia's Global21 Initiative, 2009).

Table 1 History of West Virginia Writing Assessment

Year	Grades Assessed	Summative State Writing Test	Formative Writing Program
1984-2005	4, 7 and 10	West Virginia Writing Assessment	None
2006-2008	4, 7 and 10	Online Writing	None
2009-2010	3-11	WESTEST 2 Online Writing	Writing Roadmap 2.0
2010-2014	3-11	WESTEST 2 Online Writing	WV Writes

Source: A Chronicle of West Virginia's Global21 Initiative, 2009; S. Foster, personal communication, 2012

Summative Writing Assessment 2009-2014

According to Table 1, WESTEST 2 Online Writing was field tested and incorporated into the WESTEST 2 RLA students' scores in 2009. The WESTEST 2 Online Writing was a criterion-referenced exam that assessed all West Virginia public school students in grades 3-11 on their writing ability. The writing assessment assessed students using the West Virginia writing rubric which was comprised of the five analytic writing traits of organization, development, word choice/grammar usage, sentence structure, and mechanics embedded in the West Virginia 21st Century Writing Standards and Objectives (WESTEST 2 Online Writing Overview, n.d.). WESTEST 2 Online Writing results were to be used to inform instruction and improve student writing and literacy. WESTEST 2 Online Writing was given to students from 2009-2014.

What was unique about WESTEST 2 Online Writing was that it was completed by students using a secure, online browser (WESTEST 2 Online Writing Overview, n.d.). According to A Chronicle of West Virginia's Global21 Initiative (2009), using a word processing system, the students typed in their essay responses addressing a randomly assigned passage and prompt that was written and developed by the WVDE. Within the system, students in grades 4-11 saw one of the four writing genre prompts of narrative, descriptive, informative, and persuasive when they accessed the program. Third grade students were randomly assigned a passage and prompt that was narrative or descriptive. These four writing genres are embedded in

the West Virginia writing content standards and objectives for each grade level from grades 4 to 11; however, grade 3 writing content standards and objectives only addressed the two genres of narrative and descriptive writing (A Chronicle of West Virginia's Global21 Initiative, 2009).

The students' essays were scored using artificial intelligence based on a holistic scoring process using the WV writing rubric's five analytic traits of organization, development, sentence structure, word choice/grammar usage, and mechanics (WESTEST 2 Online Writing Overview, n.d.). The scores for the five analytic traits were calculated into the students' overall WESTEST 2 RLA score. Counties and schools were provided reports on their students' performance on the WESTEST 2 RLA portion of the test; in addition, they also received separate holistic scores for the five analytic writing traits (WESTEST 2 Online Writing Overview, n.d.). The data available from the WESTEST 2 results were to be used by schools and teachers to make informed decisions about the curriculum and instruction being done at the schools on the state content standards. Thus, the scoring and reporting capability of WESTEST 2 data meant that students could be further assisted when they returned to school the next year, and teachers could make informed decisions about what concepts and skills students had mastered or needed further instruction to master. Additionally, the results of the writing assessment were tied to school, district, and state accountability for the No Child Left Behind Act (A Chronicle of West Virginia's Global21 Initiative, 2009).

West Virginia Formative Writing Assessment Programs

Writing Roadmap 2008-2010. From 2008-2010, West Virginia students were provided access to practice writing on the formative writing program known as Writing Roadmap 2.0. White, Hixon, and Whisman (2011) conducted a study on a sample of 8,577 students who were given access to Writing Roadmap 2.0 during the 2009-2010 school year. The findings from this

study showed that students using Writing Roadmap 2.0 scored higher on the WESTEST 2 Online Writing taken in 2010. What was significant about this study was that despite the differences in rubrics and scoring engines in Writing Roadmap 2.0 and the 2010 WESTEST 2 Online Writing, there was statistical significance that students using Writing Roadmap 2.0 outscored their peers who did not have access to the practice writing program.

Writing Roadmap 2.0 was offered to all West Virginia students in grades 3-11 because the WVDE believed that students needed access to a practice program prior to taking the summative online writing assessment. The philosophy of the department was that by improving student writing, the students were also becoming 21st century learners with the connection to technology and literacy skills (A Chronicle of West Virginia's Global21 Initiative, 2009). Technology and literacy were two strong focuses related to the WVDE's education mission.

WV Writes 2010-2014. According to WV Writes, (n.d.), WV Writes was an online formative writing assessment program that was designed to assess students' writing progress over the course of the school year. WV Writes provided passages and prompts that assessed the writing modes of narrative, descriptive, informative, and persuasive. Teachers also created their own passages and prompts. Students in West Virginia public schools were given unlimited access to practice sessions. WV Writes gave students immediate scores and narrative feedback regarding their writing abilities (WV Writes, n.d.).

WV Writes allowed educators to target instruction in writing by providing important data on student performance relative to the WV writing rubric areas of organization, development, sentence structure, word choice/grammar usage, and mechanics (WV Writes, n.d.). Teachers were given access to the student results immediately after students completed a writing assessment, and students were given the opportunity to revise and resubmit their essays as they

progressed through the stages of the writing process. WV Writes reports were available on the district, school, classroom, and student level (WV Writes, n.d.).

Writing Roadmap 2.0 versus WV Writes

Prompt Structure. In Writing Roadmap 2.0, the prompts were called shelf prompts which meant they were created by the vendor, CTB/McGraw Hill. Writing Roadmap 2.0 offered 162 generic prompts for students in grades 3-11 in the four writing genres of narrative, descriptive, informative, and persuasive. These prompts typically consisted of three or four sentences that asked students to write about a topic after reading the sentences. For example, the descriptive prompt "Special Day" stated, "Essay Topic: Imagine that you wake up at home and suddenly remember it is a special day. Describe what you hear, what you see, and what you smell on this special morning" (Writing Roadmap 2.0, 2009, p. 2). Additionally, the prompts in Writing Roadmap 2.0 repeated for grade levels, so the prompt "Special Day" was available for teachers to assign to grades 3, 6, 8, 10, and 12. Thus, students were exposed to the same writing prompt at different grade levels. The Writing Roadmap 2.0 prompts were not formatted like the WESTEST 2 Online Writing prompts that students were assessed on at the end of the year. The WESTEST 2 Online Writing prompts were more robust in their formatting and content. They consisted of three sections which were the directions for the students, a reading passage that may be several paragraphs in length, and the assessment prompt which stated the type of essay students would write (WV Writes, n.d.).

WV Writes offered teachers and students more options when it came to prompts. It contained two essay prompt folders that teachers could select from and assign to their students (WV Writes, n.d.). It had the original 162 Writing Roadmap 2.0 prompts available, and it also had 34 additional prompts in a folder called WESTEST 2 Online Writing practice prompts.

These 34 WESTEST 2 Online Writing practice prompts were prompts that had been field tested in 2008 but were not selected for the summative assessment. The rationale behind adding these to Writing Roadmap 2.0 and calling it WV Writes was that these unused prompts could allow students an even more closely aligned testing experience when using the WV Writes program to prepare for WESTEST 2 Online Writing (A Chronicle of West Virginia's Global21 Initiative, 2009). While the prompts did not pass the validation studies needed to become WESTEST 2 Online Writing prompts, they were identical in structure to what the students would see on their writing assessment. Thus, students who practiced writing on these prompts would be even more comfortable when they took the WESTEST 2 Online Writing.

When comparing the WESTEST 2 practice prompts with the Writing Roadmap 2.0 prompts, the WESTEST 2 Practice Prompts consisted of three sections: directions, reading passage, and a prompt that directed students to compose an essay based on their reading of the passage and directions (WV Writes, n.d.). The WESTEST 2 practice prompts were written by West Virginia educators and aligned to the West Virginia 21st Century Writing Content Standards for each grade level. Each prompt was assigned to a grade level specific to that prompt, unlike the Writing Roadmap 2.0 prompts which repeated the same prompt over multiple grade levels. The writing content standards were anchors for the prompt as it aligned to the standards. For example, a grade six prompt would align to West Virginia RLA.0.6.2.03 which stated, "From a prompt, use the writing process to develop a composition that contains specific, relevant details and transitions" (Teach21 On-line CSO Level Resources, n.d., p. 1). Every grade level from 3-11 had prompts aligned to the correct writing standard and used language within the prompt that had been verified through lexile scores to be in the correct reading range for students in that grade.

Also, the prompt was distinguished as being one of the following types of writing genres: narrative, descriptive, informative, or persuasive (WV Writes, n.d.). The directions stated that students should read the passage and prompt and then write an essay. The passage might consist of one to four paragraphs that set up a scenario or situation that students would read and address in their essays' responses. The prompt itself had two or three sentences that stated the writing genre that students should use and what the purpose, audience, and tone of the essay should be. An example of a 6th grade persuasive WESTEST 2 practice prompt that appeared below the directions and the reading passage was "Prompt: Write a composition for the class website that will persuade advertisers to change one thing about how they advertise their products. Be sure to include reasons that will convince the advertisers that the change is important to you" (WV Writes, 2013, p. 1).

Analytic Writing Traits Used in Essay Scoring and Score Reports. The second difference between Writing Roadmap 2.0 and WV Writes would be how the essays reported scores in each program. Both programs used artificial intelligence scoring engines to score the essays; however, the scoring reports in Writing Roadmap 2.0 and WV Writes were different. The scoring in Writing Roadmap 2.0 consisted of a report on each of these six analytic writing traits: Ideas and Content, Organization, Voice, Word Choice, Conventions, and Fluency (see Figure 1).

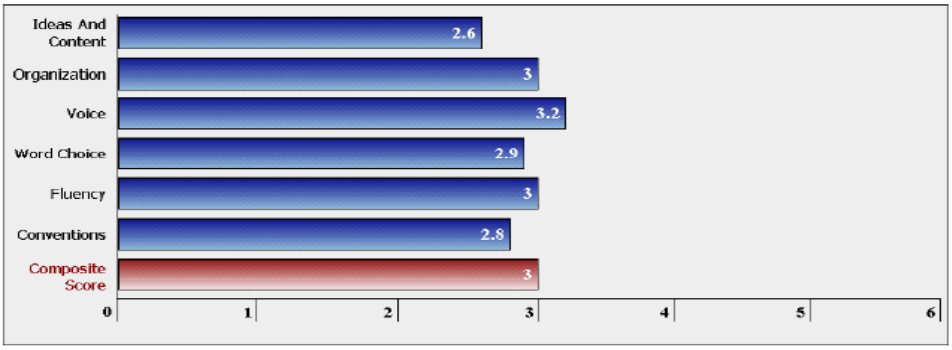

Figure 1 Writing Roadmap 2.0 Scoring Report
Source Writing Roadmap 2.0 Student Interface, 2009

As far as student feedback from the score report, Writing Roadmap offered students a score description. The score description consisted of a bulleted list of the six analytic writing traits, the score on each trait, and a brief statement of what that number correlated to in terms of proficiency (see Figure 2). For example, in Ideas and Content the student had received a score of four out of six points and the narrative feedback stated, "Your score in Ideas and Content means your writing is on its way to being complete and detailed" (Writing Roadmap 2.0, 2009, p. 1). In Writing Roadmap 2.0, students' essays were assigned a composite score which was the average of these six analytic writing traits.

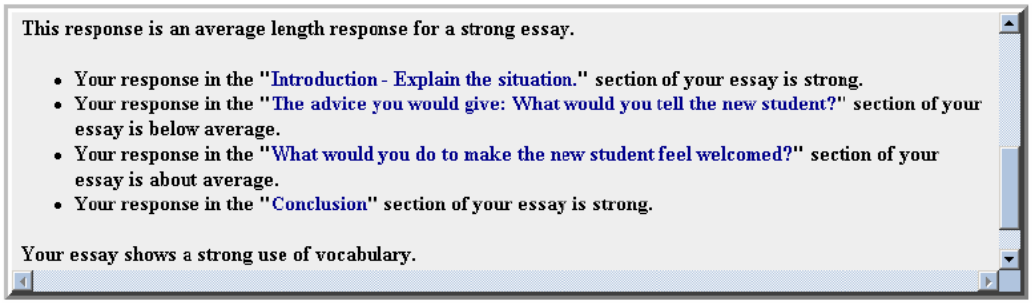

Figure 2 Writing Roadmap 2.0 Score Description
Source Writing Roadmap 2.0 Student Interface, 2009

The teachers selected whether they wanted the engine to produce a composite score from a four, five or six point scale. Both the six analytic traits and the option to select from three different point scales created confusion for teachers and students because it was not clear which scale should be selected (S. Foster, personal communication, August 2013). Writing Roadmap 2.0 was scoring the essays differently than the West Virginia writing rubric that teachers and students were to use to assess writing. Teachers were using the West Virginia writing rubrics to teach and assess writing in their classrooms. The West Virginia writing rubric had five analytic writing traits, not six, and it also used a numerical range of one to six for its proficiency levels. Whenever students used Writing Roadmap 2.0 and viewed their scoring responses, it was not the same as the West Virginia writing rubric scoring used on WESTEST 2 Online Writing.

WV Writes aligned with WESTEST 2 Online Writing in how it reported scores on student essays because it used the same five analytic writing trait rubrics that the WVDE had adopted for writing instruction and assessment (S. Foster, personal communication, 2013). The West Virginia Writing and Language 21st Century Content Standards were embedded within these writing rubrics. While the indicators within the rubrics were the same across grade levels, teachers were to keep the developmental abilities and grade level of the students they were teaching in mind as they used the rubrics as scoring tools for student writing (S. Foster, personal communication, 2013). The West Virginia rubrics consisted of five analytic writing traits: organization, development, word choice/grammar usage, sentence structure, and mechanics (see Appendix A). The West Virginia rubric proficiency ranges were given a numerical value from one to six. A score of one on the rubric meant inadequate, a score of two meant minimal, a score of three meant limited, a score of four meant adequate, a score of five meant effective, and a score of six meant exemplary. WV Writes used these writing traits and numerical values for each

proficiency level, and it also issued essays a holistic score from a six point scale. When students viewed their score reports, they would see each analytic trait with a numerical score and the overall holistic score their essay received (see Figure 3).

In addition to the score, narrative feedback was also generated for the students as soon as they scored their essays in WV Writes. The narrative feedback was listed directly below a flow chart that showed each analytic trait with a score ranging from one to six and the holistic score of the entire essay (see Figure 3). Therefore, when students received a score report from WV Writes, they were also able to read what each analytic trait score meant in terms of the strengths or deficiencies in their writing. For example, if a student essay had a score of three for Sentence Structure, the narrative feedback stated that the essay had limited sentence structure which meant there were "some errors in structure, limited evidence of sentence variety (types/length)" (see Appendix A).

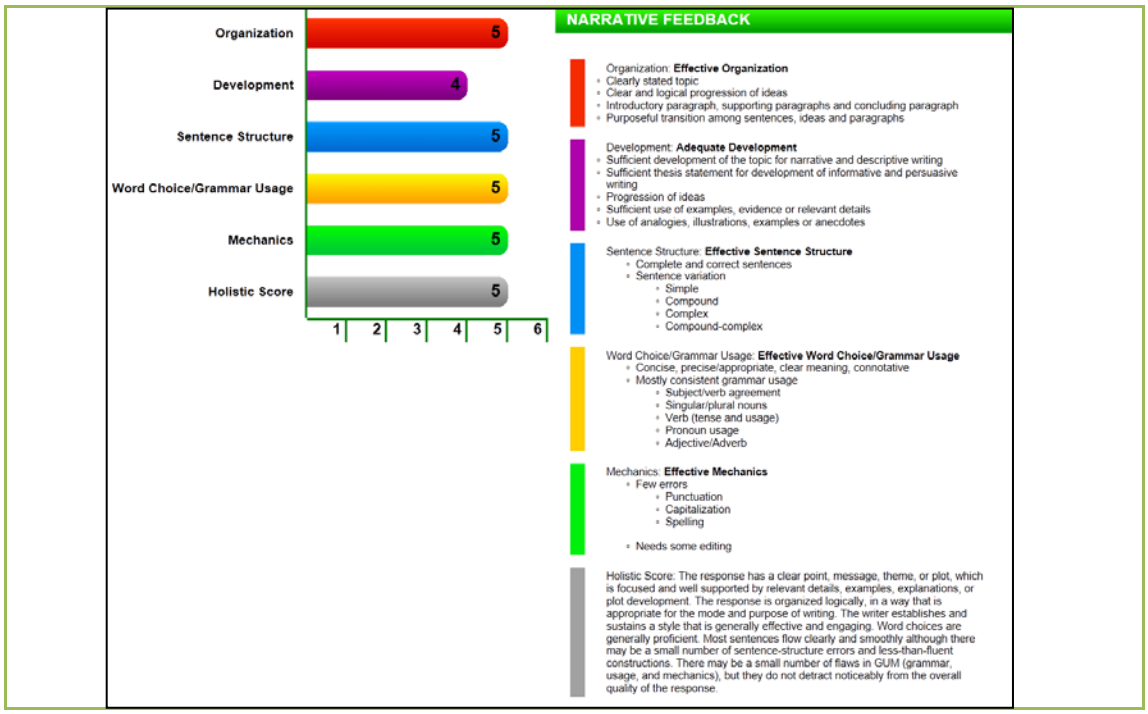

Figure 3 WV Writes Score Report and Narrative Feedback
Source WV Writes Student Interface, 2012

Scoring Engine. The scoring engines that are used in Writing Roadmap 2.0, WV Writes, and WESTEST 2 Online Writing are not the same engine. However, all three scored essays using a process that combined natural language processing, artificial intelligence, and statistical technologies which produced an automated language analysis (R. Loiacono, personal communication, 2013). This algorithm can be seen in other types of scoring engines as well (Elliot, 2001).

Writing Roadmap 2.0 was considered a generic scoring engine because it scored the essays submitted using an engine that was trained using examples of student writing based on genre and grade level bands (R. Loiacono, personal communication, 2013). For example, a student essay written to the Writing Roadmap 2.0 descriptive prompt "Special Day" for grade 3 would be scored by the engine's intelligence with the set of scoring standards set for grade 3 student writing using samples of student writing that addressed descriptive writing, but the writing samples were not necessarily about a special day. The score produced for the essay was formulated by the engine. Prior to public use, the Writing Roadmap 2.0 scoring engine was calibrated and validated using an algorithm (R. Loiacono, personal communication, 2013).

To guarantee validity in the program's ability to score student writing, Writing Roadmap 2.0's scoring engine was originally calibrated using a validation study (R. Loiacono, personal communication, 2013). First, a collection of student writing was gathered. The student writing was categorized based on the genre of writing within the student essay and the grade bands associated with the prompt. For example, a collection of student writing by third and fourth grade students composing narrative writing was compiled. The assessment system vendor, CTB/McGraw Hill, then had their employees hand score the writing samples using the six analytic traits and the scoring ranges in Writing Roadmap 2.0 (S. Foster, personal

communication, 2013). The program scoring engine algorithm was then given the scored student essays in order to associate and calculate the scores associated with each genre and grade band. To test the validity of the scoring engine with that of human scorers, the program was given another set of unscored essays that it scored (White L., Hixon, N., & D'Brot, J., 2010). The scores given by the scoring engine were compared with what the human scorers gave, and when there were high levels of agreement between the artificial intelligence scoring and its human counterpart the scoring engine was validated.

With regards to WV Writes, the same process was used; however, the WVDE chose to move beyond the previously existing scoring engine in Writing Roadmap 2.0 by creating the West Virginia writing rubric as the scoring tool, creating their own prompts, gathering student samples from West Virginia, using West Virginia educators and specialists as hand scorers, and training an engine specifically made to assess West Virginia student writing for WESTEST 2 Online Writing (S. Foster, personal communication, 2013). The WVDE progressed through a similar process when it provided CTB/McGraw Hill with over 600 student essays written to each WESTEST 2 Online Writing prompt (Rich, Harrington, Kim, & West, 2008). West Virginia educators and writing specialists were trained and worked collaboratively with CTB/McGraw Hill hand scoring employees to select writing samples that would build the foundational scoring used by the WESTEST 2 Online Writing engine (S. Foster, personal communication, 2013).

Every analytic trait for every grade level on the West Virginia writing rubric was assigned to the writing samples available for each prompt (Rich, Harrington, Kim, & West, 2008). Student paper samples had to be high enough in number to ensure that the engine would calibrate and recognize the number ranges for the five analytic traits (S. Foster, personal communication, 2013). So, a minimum of 100 student writings were needed for each proficiency

level (1-6) on the West Virginia writing rubric for each analytic trait (organization, development, sentence structure, word choice/grammar usage, and mechanics) for each prompt. Thus, it was necessary to have more than 100 papers that scored a one in organization on the grade 3 narrative prompt. This same prompt would need 100 papers that scored a two in organization on the grade 3 narrative prompt. Once all the essays had been hand scored, they were entered into the scoring engine as distinct prompt responses and then the engine was allowed to process the responses as range finding samples (S. Foster, personal communication, 2013).

Upon completion of the validation engine scoring study done for WESTEST 2 Online Writing, there were 34 prompts that were deemed unfit for the summative writing assessment. These prompts were removed from their secure settings. Because Writing Roadmap 2.0 was viewed as beneficial but somewhat misaligned to the summative writing assessment, the decision was made to create WV Writes and house those unused prompts and the scoring engine associated with those prompts in the new program (S. Foster, personal communication, 2013). The engine scoring validation process for the unused 34 prompts was the same as that used for the actual WESTEST 2 Online Writing prompts. Thus, WV Writes offered something very unique to West Virginia teachers and students - prompts and a scoring experience that fit exactly with the summative writing assessment, but lacked the validation studies necessary for use as a high stakes writing prompt (S. Foster, personal communication, 2013). The original Writing Roadmap 2.0 prompts and the engine for scoring those prompts existed alongside the WV Writes WESTEST 2 practice prompts and their scoring engine. The WVDE determined that keeping the Writing Roadmap 2.0 prompts, rather than removing them from the WV Writes program completely, alongside the WESTEST 2 practice prompts would provide more selection for the teachers as they taught students how to write (S. Foster, personal communication, 2013).

STATEMENT OF THE PROBLEM

As students are being assessed on their writing abilities in a summative manner, the notion of having a formative assessment available for them to practice in preparation for the high stakes writing assessment has been a popular trend. WV Writes was the formative writing assessment program that the WVDE purchased for students in grades 3-11 to practice their writing skills prior to taking the WESTEST 2 Online Writing. Each year thousands of essays were written and scored in WV Writes.

As so many students used WV Writes, the question emerged of whether a connection could be found between the scores students received in WV Writes and the scores they received on WESTEST 2 Online Writing. Counties, schools, and teachers speculated that their students scored higher on WESTEST 2 Online Writing because they had extensively used the practice tool of WV Writes. Within the practice program, students wrote to prompts that mirrored the WESTEST 2 Online Writing prompts and received scores in WV Writes that aligned to scoring on WESTEST 2 Online Writing.

The purpose of this investigation is to determine the effects of students writing to practice prompts on their high stakes writing assessment performance. Given there are two types of prompts available in WV Writes, the Writing Roadmap 2.0 prompts and the WESTEST 2 practice prompts, it may be expected that there will be significant changes in the scores of the writing samples among 11th graders when distinguishing between those students who wrote to the Writing Roadmap 2.0 prompts and took the WESTEST 2 Online Writing and those students who wrote to the WESTEST 2 practice prompts and took the WESTEST 2 Online Writing.

In order to prepare for the new wave of assessments being created to assess the Common Core State Writing Standards, educators, states, and policy makers must decide if they will offer

formative as well as diagnostic assessments in preparation for the summative test. It is important to determine whether or not the alignment between a formative assessment and its summative assessment will maximize student achievement and mastery on the summative test. This study will show any possible connections.

STUDY DESIGN

This study was a quantitative, post hoc, 2 x 2 ANOVA (analysis of variance) design that consisted of 380 randomly sampled 11th grade West Virginia students who used the WV Writes prompts and were assessed with the WESTEST 2 Online Writing in 2013. The experimental group was 190 11th grade students who wrote essays responding to the WESTEST 2 practice prompts in WV Writes. The control group was 190 11th grade students who wrote essays responding to the Writing Roadmap 2.0 prompts in WV Writes. The two independent variables were the types of prompt in WV Writes and the gender of the students. Factors of the writing prompt independent variable were the Writing Roadmap 2.0 and the WESTEST 2 practice prompts. The factors for gender were females and males. The dependent variable was the measured effect on the writing proficiency scores. The WV Writes and WESTEST 2 Online Writing scores of student writing on the West Virginia Writing Rubric in the five analytic traits of organization, development, sentence structure, word choice/grammar usage, and mechanics were translated into composite scores ranging from 0-30. The intervention condition for the experimental group was the WESTEST 2 practice prompts, while the intervention condition for the control group was the Writing Roadmap prompts. The posttest assessment in this study was the 2013 WESTEST 2 Online Writing scores which were calculated into composite scores.

SIGNIFICANCE OF THE STUDY

If significance was found between students scoring at mastery or above on the WESTEST 2 practice prompts and on WESTEST 2 Online Writing, then this would show that similarity between a practice writing program and a summative writing assessment increases student proficiency in writing. With the release and adoption of the Common Core State Standards, many states find themselves implementing new learning standards into the curriculum. Both assessment consortia, the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium (SBAC), have stated that the summative and formative assessments they are developing for use in the 2014-2015 school year will be online programs that students will access using computers. Both assessment consortia will offer students interim assessments for use as predictors of student performance on the summative assessments (J. D'Brot, personal communication, 2014). It has already been researched and proven that students who use formative assessments to prepare for a summative assessment will perform better than students who do not have access to the formative assessments (White, Hixon, & D'Brot, 2010).

If this study found that the students who used the WESTEST 2 practice prompts score higher on the WESTEST 2 Online Writing than those students who used the Writing Roadmap 2.0 prompts, then there were many who were affected by this study. This study held significance for students because they are expected to show their performance ability on the standards they are taught during the school year. Students will need to have practice opportunities that closely resemble the high stakes testing they are required to take at the end of the year. Teachers may also be impacted by this study because they are expected to educate and prepare students on the state content standards and objectives that the students are assessed on during high stakes

assessments. Thus, teachers will need to have access to and understand how to use the practice tests effectively. School administrators may find this study significant because they are held accountable by the state and federal government policies and regulations based on the students' performance on high stakes tests. If schools can improve their students' performance on summative assessments by using practice tests that mirror the high stakes tests, then they will want to ensure their districts and state are providing these practice tests to all students and all schools. Districts will find relevance in this study as they must offer assistance and support to schools that are unable to show mastery levels in their student achievement results on high stakes tests. If districts learn of better student performance based on closer alignment between a practice test and a high stakes test, they will want to provide these preferential practice tests to their students and schools. Lastly, states will find significance in this study because they are federally mandated to monitor and record student performance on high stakes tests for their students. If states are required to offer all students in public schools the same opportunities and fairness related to testing, then they will want to purchase and provide their students with practice tests that show statistical significance of improving student performance on a high stakes assessment.

RESEARCH QUESTIONS

The following specific research questions were posed to determine the effects of student writing responses to two different types of writing prompts in a practice writing program on the writing scores of 11th graders on their summative writing assessment.

1. What are the effects on WESTEST 2 Online Writing composite scores among 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?

2. What differences exist among the five analytic writing traits on WESTEST 2 Online Writing for 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?
3. What are the effects of gender and types of prompts on WESTEST 2 Online Writing scores among 11th graders?

OPERATIONAL DEFINITIONS

Writing Roadmap prompts refers to the generic shelf prompts available to students to practice writing in the WV Writes practice writing program.

WESTEST 2 practice prompts refers to the West Virginia writing prompts that were not used on the WESTEST 2 Online Writing and instead were made available to students to practice writing in the WV Writes practice writing program.

WESTEST 2 Online Writing composite scores are the scores that students receive on their writing responses on WESTEST 2 Online Writing. They are composite scores that range from 0-30. A composite score of 0 means no response, a composite score of 1-10 is novice, a composite score of 11-15 is partial mastery, a composite score of 16-20 is mastery, a composite score of 21-25 is above mastery, and a composite score of 26-30 is distinguished.

Five analytic writing traits are used in the scoring of both WV Writes and WESTEST 2 Online Writing. These five traits are organization, development, sentence structure, word choice/grammar usage, and mechanics. Each trait can receive a score from 0-6. A score of 0 means no response, a score of 1or 2 is novice, a score of 3 is partial mastery, a score of 4 is mastery, a score of 5 is above mastery, and a score of 6 is distinguished.

11th graders refers to West Virginia public school students who took the WESTEST 2 Online Writing their 10th grade year in high school and then used either the Writing Roadmap

prompts or the WESTEST 2 practice prompts in WV Writes and then completed the WESTEST 2 Online Writing their 11th grade year.

Gender is an independent variable that has the factors of females and males.

Type of prompt is an independent variable that has the factors of Writing Roadmap practice prompts and WESTEST 2 practice prompts.

DELIMITATIONS

1. Practice writing prompts are limited to the two types found in the WV Writes program.
2. The length of the treatment condition was one academic school year.
3. The research includes schools and counties state-wide.

LIMITATIONS

A limitation in the study is that it is dependent on the student participants using the WV Writes program. There may be students who never used the program but did take the WESTEST 2 Online Writing Assessment. These students will not be included in the sample as it is reliant on the students showing evidence of WV Writes usage in the 2012-2013 year.

CHAPTER TWO: REVIEW OF LITERATURE

INTRODUCTION

In this chapter, a relevant review of literature was conducted and organized into seven topic areas. The first examined research studies conducted on the use and implications of writing prompt components and their construction. The second topic area was about using writing prompts as assessment tools. The third section reviewed research studies that focused on narrative feedback and reporting on student writing. The fourth topic reviewed the research conducted on assessing and evaluating student writing using analytic trait scoring and holistic trait scoring. The fifth examined the variable of gender on writing proficiency. The sixth analyzed the reliability and validity of artificial intelligence scoring of student writing on formative and summative writing assessments. The seventh section explored literature on the impacts of formative and summative writing assessment programs.

PROMPT STRUCTURE AND CONSTRUCTION

Within all the studies conducted on the implications of what a writing prompt looked like and how it was constructed, a common finding was discovered (Furtak & Ruiz-Primo, 2008; Plakans, 2008; Plakans & Gebril, 2012; Ruiz-Primo & Shavelson,1996). Prompt construction affected students' perceptions and confidence in their ability to address writing prompts, and prompts that were specific in directions, yet open in interpretation allowed students to better comprehend the required writing task and perform the desired writing assessment.

Furtak and Ruiz-Primo (2008) analyzed the use of formative assessment prompts by middle school students in a science class to determine if some prompts were better than others in organizing correct information from students in their written responses and discussions on the

prompts. They discovered that students understood the concepts behind the prompts best when the prompts were structured openly and contained familiar information. When students addressed the open and more familiar prompts in writing, they showed a good range of understanding of the concepts and topics they addressed. The researchers concluded that essay prompts being used formatively were best structured in an open format that allowed students a range of response opportunities.

In a similar study, Ruiz-Primo and Shavelson (1996) concluded that assessment prompts should specifically direct students to complete a writing that illustrates evidence that the students have taken knowledge obtained through learning and applied it by addressing the prompt. The structure of the prompt should ensure that students accurately respond to the task asked of them in order to objectively judge the quality of their responses. The instructional process using written formal assessment prompts should follow a series of specifications that include a task, the details or specifics that should be embedded in the student writing, and the means of assessing the final written product. Furtak and Ruiz-Primo (2008) argued that schematic knowledge was best assessed through writing prompts. Schematic knowledge was defined as knowledge that is more organized and requires students to apply their understanding when problem solving, explaining, or predicting. When using writing prompts to assess student learning, the quality of students' understandings of the concepts is being assessed in a concrete manner.

One challenge that Furtak and Ruiz-Primo (2008) noted about using writing prompts and students' written responses to the prompts as a formative assessment tool was the time consumption that teachers experienced when attempting to read, review, and provide feedback on student writing responses. When teachers tried to give timely feedback to students' writing,

they found the task overwhelming and not an option when a deadline was imminently approaching on the concepts and knowledge being taught. Feedback on student writing will be further addressed in this section of the literature review.

According to Furtak and Ruiz-Primo (2008), successful writing prompts used formatively must show students have developed beyond the foundational understanding of the concepts and knowledge being assessed. When students expressed a lack of mastery on a concept in their writing response, then teachers needed to recognize this deficiency and provide scaffolding to close the gap between misunderstanding and clarity. Using writing prompts to assess student learning required educators and curriculum experts to develop the prompts with the purpose in mind first, then they could consider the format and type of prompt to be created.

In another study, Plakans (2008) focused on the components of writing prompts and whether or not students gained a better understanding of the prompts when they were required to read a passage or source materials and then apply the information into the writing assignment versus just reading a brief writing stimulus and then writing to it. The study examined differences between students writing to prompts that were reading-to-write tasks versus writing-only tasks for the placement of English as a Second Language (ESL) students at a university. She concluded that reading-to-write writing prompts allowed students who were experienced with writing to express their competency much easier than the writing-only writing prompts. Also, it was noted that students who were more skilled in their writing were more likely to do more prewriting or planning activities prior to writing. For writers with less skills and experience, the prewriting process was significantly less with more of their attention being placed on the drafting process. When students were given writing-only prompt tasks, they focused their efforts prior to beginning by organizing the content of their written responses. The participants would decide

what the main points of their response would be, come up with supporting details, and then determine the organization of the response. When students were given writing-only prompts, they spent more time preparing to write before they would begin their responses.

A similar research study by Plakans and Gebril (2012) was conducted on prompt construction and ESL students' writing proficiency which focused on the effect of writing assessments requiring students to read an essay prompt and then write a response addressing the prompt. Some writing assessments had students integrate source materials into their writings while more traditional writing assessments had students reading a prompt and then writing their thoughts into an essay response. The researchers wanted to see if any differences existed between student proficiency in writing when comparing the two writing prompt types: those writing prompts that had students write essays with source documents integrated into their responses and those writing prompts that had students who write more traditional responses that did not require source integration. The results of the study were broken down into three areas based on how the students comprehended the source documents, how the students implemented the source documents into their writing, and the relationship between students using source documents in their responses and their writing scores. Students who scored higher on their essays were better able to understand the content of the source materials prior to writing their essays. Students continually referred back to the source documents as they drafted their essays regardless of what range of score their final essays received. Five reasons were given by the students as to why they revisited the source documents as they wrote their essays. These included becoming informed on the topic in the sources, generating an opinion or stance about the topic, finding support from the sources, viewing the language as examples of writing, and

incorporating the organizational patterns into their own writing. The study found no significance in the scores of the students' writing and how they used the source materials.

In another study, Olinghouse, Zheng, and Morlock (2012) focused on student motivation and writing prompt construction. The study examined the effect of motivation that students felt when reading and writing to a high stakes writing assessment prompt. The authors noted that the research centering on prompts and motivation on student writing was minimal. The study also pointed out that skilled writers were able to work independently with little feedback from others when being assessed on a formal writing task. Writers must be motivated to complete a writing task in a mainly solitary setting.

Hidi and Boscolo (2006) found that when students felt motivated after reading a writing task prompt, they were positively affected to react and address the prompt. Students who were motivated possessed self-regulation and self-efficacy in their writing abilities. Bruning and Horn (2000) found that traditional academic writing prompts were focused on assessing students writing skills without allowing them to express their personal thoughts and points of view. They noted that writing prompts that encouraged student writing on personal or authentic experiences caused students to write more and be more engaged with the writing experience.

Olinghouse, Zheng and Morlock (2012) noted that little research existed in the area of high stakes writing prompt evaluation. Their study focused on the motivational ability of writing prompts on student writing. They used 222 writing prompts from 44 states with an average of 5.2 prompts per state. The study found that six variables affected the motivation that students felt when responding to a high stakes writing assessment prompt. These were time allocation, audience specification and intimacy, choice in writing task, multiple perspectives, and real world purpose. The study showed that students felt more motivated to write when the prompts had

individual or group audiences, real world application and the writing purpose allowed for exploration by the student from multiple perspectives. However, they concluded that the study did not indicate a clear picture of the best way to construct a high stakes writing prompt.

Similar to Olinghouse, Zheng, and Morlock (2012), research studies have been conducted to investigate the perceptions held by student writers as they read and respond to writing prompts. Powers and Fowles (1999) argued that the students who take writing assessments would be valuable resources for prompt developers. If students felt interested in and familiar with the contents of the writing prompt, then they were more likely to become engaged with writing a response. The researchers noted that there were no clear guidelines for developing writing prompts; however, guidelines did exist for evaluating them. According to Miller and Crocker (1990) writing assessment prompts should possess the following qualities: create thoughtfulness in the writer, encourage self-expression, relate to the majority of the writers, and be equally accessible to all groups and subgroups.

One obstacle for student writers in making use of writing prompts is that the same prompt can have different meanings to different people. Miller and Crocker (1990) concluded that students who were writing essays preferred writing prompts that were relatable, clear and concise, caused an emotional response or stance on an issue, or engaged the writer's interest. Prompts were classified as more difficult when they were considered boring, required specific knowledge to respond, were vague in purpose, or did not have real world connections. The study examined the possibility of a correlation between the student's perception of the prompt and the essay scores of the written response. The study found that there was no correlation as essay scores were similar in both high and low levels of interest, and it failed to show a significant relationship between the students' opinions and their essay scores. Interestingly, they noted that

the essays were scored holistically, and that analytic scoring of essays might provide different results.

USING WRITING PROMPTS TO ASSESS WRITING

In order to assess writing, both summative and formative writing tasks typically provided students with a writing prompt that asked them to address the prompt in their written response. The use of writing prompts to assess students' writing abilities led to researchers inquiring into the functionality and ability of writing prompts to accurately inform learning, instruction, and capture students' proficiency. Crawford, Helwig, and Tindal (2004) revealed that students with learning disabilities scored better on writing assessments when they were given extended time to compose their written responses. Writing theorists have suggested that allowing extra time for the writing process will engage students in all the stages of the writing process and, additionally, will encourage them to progress and regress through the writing stages in a non-linear fashion (Applebee, 1996; Hayes, 1996). Without adequate instruction on writing skills and the writing process, students may not maximize the extended writing time. Interestingly, previous studies of revision to writing showed that there were no significant differences between first and final drafts done by students in extended time writing assessments (Matsumura, Patthey-Chavez, Valdes, & Garnier, 2002; Scardamalia & Bereiter, 1983).

Additionally, the final drafts of student writing were of poorer quality than the first draft when students were given extended time on writing assessments (Goldberg, Roswell, & Michaels, 1996). The changes that students made to their writing were found to be cosmetic changes to spelling and mechanical errors 60% of the time that changes were even made (Graham, 1997). Bridwell (1980) and Stoddard and MacArthur (1993) concluded in their studies that skilled student writers made revisions and changes to their writing during the writing of the

draft, not after its completion. Revision occurred before and during the creation of a first draft when skilled writers were composing written responses (Faigley & Witte, 1981).

Nuckles, Hubner, and Renkl (2009) conducted a study on using learning journals to increase student learning. The students were to read prompts that activated their learning strategies and encouraged them to write. When used as knowledge activators, writing prompts facilitated self-regulated learning in the students' learning journals. Knowledge was elicited by the students when they addressed the prompts in their writing. Thus, the study illustrated that students learned metacognitively when provided writing prompts that stimulated their higher order thinking skills.

The literature review illustrated that a second area of studies focused on the genre of writing that students were asked to use when composing their responses during writing assessments. McMaster and Campbell (2008) presented the question of whether or not the type of writing task, the time spent writing to the task, and the process used to score the written response would have statistical significance. The study was related to progress monitoring of students in elementary and secondary schools and whether improvements were documented in the chronological completion of the writing tasks by the students throughout the academic year. Based on the scoring system used in the study, the researchers concluded that the written responses for narrative writing prompts would benefit from a complex scoring system beyond the one used in the study. The scoring method used in the study just looked at total word count, spelling errors, and mechanical errors. Regarding expository writing prompts, the study determined that student writing improved from a fall writing assessment to a spring writing assessment. The study concluded that using short narrative writing prompts would benefit student screening procedures and also reliably indicated growth. Also, expository prompts when

used by secondary students showed promise for indicating progress and assisting in monitoring procedures.

Condon (2004) defined a new type of writing assessment that had students write an essay based on a generative writing prompt. A generative writing prompt was a prompt that was constructed so that writers could express their experiences in their written responses. The generative prompt went beyond just being tailored to each student's personal experience because it could also become a source of data regarding the learning experiences that the student had while attending a school. Traditionally, writing was assessed by having students read a passage and then write a response to the passage. The concern with this type of writing assessment was that students who struggled with reading were already at a disadvantage. The writing assessment was not assessing their writing abilities but their reading abilities instead. In this study, the generative prompt required students to evaluate their courses taken at the school and relate them back to their personal experiences and learning. The insights and reflections made by the students in the generative writing prompt were used as a data source because the written responses had the information extracted and quantified for data analysis of the school and its programs from the students' perspectives. When tested in the study, the generative prompts scored with validity and reliability.

The writing assessment prompt does not exist in isolation; thus, Gabrielson, Gordon, and Engelhard (1995) determined what other factors, including the writing prompt, were impacting students as they expressed themselves through writing. The researchers explained that writing prompt assessments could inform instructional practices. Typically, performance assessments known as direct writing tests were used by states and national vendors. While writing standards

were not consistent across the states, there was an assumption that certain characteristics existed within writing that would deem that writing good or bad.

Schoonen (2005) speculated that writing assessment scores would be impacted by the multiple facets within the assessment process, including writing proficiency, prompt type, scoring rubric, and scoring criteria within the rubric. He concluded that the type of writing students composed their essays in affected scoring variance more than the individuals scoring the writing. This research study replicated the findings of previous studies (Barrett, 1994; Brennan, Gao & Colton, 1995; Godshalk, Swineford, & Coffman, 1966; Lee, Kantor, & Mollaun, 2002; Moon, Loyd, & Hughes, 1996; Shavelson, Baxter, & Gao, 1993; VandenBergh, DeGlopper, & Schoonen, 1988) which all indicated that the type of writing prompt being assigned created variances in the writing scores. According to Schoonen (2005), the best way to obtain reliable and valid writing scores on writing assessments was to establish the generalizability of each writing assessment prior to its full implementation.

Beck and Jeffery (2007) analyzed high stakes writing assessment prompts and student responses from California, Texas, and New York to distinguish what types of writing prompts were being used and what genres of writing were being required of the students. A concern the researchers expressed was whether the genre of writing used in the writing prompt was impacting how the essays were being scored. These researchers concluded that while the writing prompts were requiring specific writing genres to address the prompts, the essay scorers were scoring essays that wrote out of genre as generously as those that were written in the required genre. They argued that the students were being asked to write in a particular genre and then being assessed inaccurately. Both informative and argumentative essay prompts and genres were the dominant mode of writing that students were required to use in college; however, the study

pointed out that most writing prompts were asking students to write narratives. This showed a disconnect between secondary and post-secondary instruction and assessment procedures.

Kobrin, Deng, and Shaw (2011) investigated whether there were qualities of student writing and writing prompts that influenced the scoring of the student writing positively or negatively. Specifically, the study looked at prompt types, length of essays, organization of student essays, and content of student writing. One issue related to writing assessment using writing prompts that students must address in their response was that the essay prompt type would affect the quality of writing by the student. Some prompts could be more difficult than others depending on the student reading and responding (Sacchetti, 2005). Another concern was that writing prompts contained cultural bias or were not universally diverse for the students who read and responded to them, putting some students at an advantage over others (Bridgeman, Morgan, & Wang, 1997; Hinkel, 2002; Jennings, Fox, Graves, & Shohamy, 1999). Huot (1990) concluded that more research should be done on writing prompts and the scores students receive. Kobrin, Deng, and Shaw (2011) concluded their study with no significant statistical findings related to the writing prompt's relationship with the writing scores.

Lee (2008) investigated whether a relationship existed between writing prompts and students' writing responses. Prior research had not determined whether students who chose their writing prompts and wrote a response scored higher than students who were assigned a prompt without any choice (Barry & Nielsen, 1997; Jennings, Fox, Graves, & Shohamy, 1999). In Lee's (2008) study, the students who were allowed to choose their prompts did not score higher than those who were not able to choose their prompts. This finding supported prior research. One interesting finding of this study was that students' perceptions were more favorable towards their writing ability when they were allowed to choose their writing prompt. Students who were not

able to choose their writing prompts perceived their writing experience and their writing ability more negatively. Despite the students' perceptions about their writing experience or abilities, the scored essays based on prompts of choice showed no improvement in scores.

GENDER AND WRITING PROFICIENCY

Researchers (Breland, 1983; Brown, 1986; Cherry & Meyer, 1993; Coffman, 1971; Nold & Freedman, 1977; Ruth & Murphy, 1988; White, 1985) have shown that there are gender differences in the writing ability of students, with other possible influences being the composition of prompts, the demographics of students, the characteristics of essay scorers and the method used to assess the writing. Gabrielson, Gordon, and Engelhard (1995) wanted to see what effects, if any, would occur when grade 11 students were allowed to choose the writing prompt. What they discovered was that the gender variables had more of an effect on writing scores than the ability to select a prompt and compose to it. Knudson (2001) showed that gender had positive correlations with writing proficiency. Using an achievement test and a writing attitude survey, they found that the variables of attitude towards writing, grade level, and gender affected the students' success in writing. Jihun (2013) determined that the attitudes of students had a higher effect than that of learning behavior when students were categorized based on gender. The study used data from the 2007 National Assessment of Educational Progress (NAEP) and concluded that female students scored higher than their male counterparts despite the students being at a similar ranking based on their writing attitudes and behaviors.

Another study (Javed, Juan, & Nazli, 2013) focused on gender and writing proficiency concluded that gender did not have statistical significance. Four hundred forty students were given a proficiency assessment on the analytic writing traits of word choice, sentence structure,

grammar, and mechanics and after analysis, it was determined that the gender of the students did not affect their proficiency.

James (2008) challenged whether or not an Automated Essay Scoring (AES) system would show bias to particular students based on their demographic information. The researcher examined if the writing prompts used in an AES system would affect writing scores based on variables of the test takers, such as gender. The AES system used in the study was ACCUPLACER WritePlacer Plus which scored using IntelliMetric. There were no statistical differences when comparing the prompts between males and females nor when comparing native English and non-English speaking students.

FEEDBACK AND REPORTING ON STUDENT WRITING

In addition to the writing prompts used in writing assessment, the feedback and reporting that students receive directly impacted student writing proficiency. Lee and Laspe (2003) studied narrative feedback and reporting of student writing results as they refined their writing skills. They analyzed the relationship of feedback for students from teachers as students worked on their writing skills in a writing program that was meant to improve writing. They examined two different modes of increasing word production in students as the students wrote essays in response to writing prompts. The students were assigned a writing prompt that consisted of a sentence starter and were then asked to write a response to finish the beginning ideas within the writing prompt. One group was given verbal cues from the teacher to continue writing after the students had paused for over one minute; the other group was asked to write three words that required students to comply with the activity of resuming writing.

Lee and Laspe (2003) illustrated that both groups had students write more when the teacher provided verbal cues to the students. The group that had written the three words from the

teacher's prompting wrote more than the first group. One important point that the researchers noted was that an effective mode of improving student writing was to get students to write more words which led to more sentences. Providing feedback to students encouraged them to write more; thus, narrative feedback on student writing led to students motivated to write more. They also noted that just giving students writing prompts and paper to write on did not guarantee more writing. Teachers needed to use a feedback method that encouraged students to finish a task that might not be extrinsically motivating to the students on its own. Thus, teachers needed access to motivating types of narrative feedback and reporting on their students' writing skills.

Peterson and McClay (2010) also stressed the role of teachers providing constructive narrative feedback and reporting on student writing. In a study on Canadian grades 4-8 teachers' perspectives on feedback and writing assessment, it was concluded that the teachers valued feedback as critical to improving student ability and student confidence in writing. The teachers believed peer editing allowed for growth in student writing and held that using standardized rubrics and performance criteria allowed for objective assessment of student writing. Elbow (1973; 1997; 2000) noted that feedback must be timely when assessing student writing. He believed that students benefited from informal assessments of writing using feedback as an ongoing mode of correction, and discussion of student writing led to improved writing. The experience of obtaining feedback from multiple readers allowed students to not be evaluated subjectively because of the various viewpoints of the readers versus one subjective view from the teacher.

Written feedback was the main way that teachers expressed their thoughts about student writing to their students. Theorists, such as Haswell (2006), have specified that writers benefited from feedback that was specifically focused on the individual writing. Writers improved when

the feedback they received contained what errors were continually occurring in their writing. The focus of the feedback should have been a manageable number, such as one or two, of the criteria that needed to be improved by the writer. When feedback allowed the students to perceive their strengths and weaknesses and then allowed them to react or alter their writing, they were able to gain a stronger sense of how to make improvements.

Nicol and Macfarlane-Dick (2006) argued that students should be active participants when experiencing formative feedback. They recommended that when providing feedback teachers should follow these guidelines. The expectations for the writing needed to be fully and clearly expressed. Students needed to self-regulate their assessment of their own writing using the scoring rubrics or criteria. When teachers gave feedback, they needed to offer quality guidance that could be applied to the writing while it was in draft and was written in a positive tone. The feedback needed to include opportunities for teachers and fellow students to talk about the writing. Students needed to feel motivated and more confident as they received the feedback and enacted the changes in their writing. Feedback had to be timely with clear instructions for making immediate changes to the current writing responses. The most critical aspect of feedback was ensuring communication occurred continually and positively.

In the findings of the Peterson and McClay study (2010), teachers had strong feelings about the role they played when creating, providing, and guiding feedback on writing. The teachers considered the impact of what they were communicating to the students and how it affected their self-esteem. The teachers believed that the goal of feedback was to help students improve their writing. When expressing feedback, the teachers noted that it needed to include verbal feedback. Teachers struggled with assessing writing in an objective manner, so they relied on a scoring rubric when providing suggestions for improving the writing. The researchers

concluded that communication was an integral part of teaching writing as it allowed students to gain aid and also build support to improve their writing.

Stern and Solomon (2006) examined what type of feedback teachers gave students on their writing and the varying degrees of usefulness of the feedback. They identified the types of written feedback that teachers gave to students on their writing assignments to see if any categorization could be made. They determined that most teacher feedback on students' writing was focused on the grammatical and mechanical aspects of writing. Important aspects in writing such as organization, development, and sentence structure were given minimal feedback. Only 8% of the teachers used a rubric as they assessed and provided feedback to the students which might have aided them in explaining where the student writing fell in a range of proficiency.

Parr and Timperly (2010) analyzed the teachers' abilities to communicate effective feedback on student writing and the improvement of student writing. They determined that a relationship existed between improved writing and the skill level of teachers who provided the feedback. Formative assessment as assessment for learning was the philosophical stance held by the researchers, and they focused their efforts on the formative written feedback that the teachers were giving their students. One important point that they noted was that feedback could vary in its effectiveness depending on the degree of depth that the feedback attempted to correct in the students' writing. If the feedback only focused on the cosmetic aspects of improving writing, then the deeper understandings behind the student writing were not addressed; therefore, feedback needed to address the concepts behind the writing. Feedback needed to occur frequently and students needed to see examples or criteria needed to improve their writing. Effective feedback allowed students to know exactly where their proficiency levels were and

what needed to be done to the writing to reach higher levels of proficiency and skill. Feedback needed to contain information that guided the students towards improved proficiency.

Alter and Adkins (2006) reported that effective writing assessment depended upon the assessment having feedback and reporting that allowed for diagnosis of the students' writing strengths and weakness. They examined how students in a graduate level social work program were assessed on their writing skills. The study considered why and how writing was assessed which included scoring and reporting the writing. The researchers pointed out many important aspects that needed to be considered if a college or university was going to use a writing assessment as a proficiency determinant of its students. Some of the first considerations related to writing assessment programs were how the writing assessment was given to the students and what it was comprised of, who assessed the writing, how the writing was assessed, how the results were reported to the students, and how the data from the writing assessment were used to inform instruction or learning.

Alter and Adkins (2006) pointed out that if a writing assessment was to be used as a diagnostic tool, then the students needed to be provided the results of the assessment. The suggestions that students received from a diagnostic writing assessment were to be applied to their writing skills in order to improve their writing. An example of this type of diagnostic writing assessment mentioned in the study was the timed, writing test which asked students to read a prompt and then compose a response without spelling or grammatical tools for sentence corrections during a set amount of time. Another challenge that schools faced when attempting to set up a writing assessment for students was the procurement of human scorers who were able to effectively score student writing. The researchers suggested that having scorers use a selected rubric that ensured reliability and validity was the best method of scoring essays. A successful

diagnostic writing rubric identified distinct writing traits that defined the characteristics of the students writing and allowed for students to interpret and improve their writing. Lastly, the results of student writing assessments needed to result in reports that were available to the students and allowed them to review and analyze their writing score results.

Zinn (1998) conducted a study on feedback and reporting, such as the informal and formal measures that students received from their teachers related to their writing proficiency. The study pointed out that two types of writing assessment existed which were informal and formal measurements. Within these two modes of writing assessment there were best practices that teachers implemented in their classrooms to maximize the opportunities for students to improve their writing skills. One important point that Zinn noted came from the International Reading Association and the National Council of Teachers of English. There were qualities that teachers could use as foundational criteria for their assessment of writing. These were that assessments on student writing were best when they allowed students to reflect productively on their writing and teacher comments encouraged students to feel positive about their efforts at revision. Second, teacher feedback needed to be structured in language that identified what students needed to do to improve their writing without using negative language that constricted students from improving their written responses. When teachers covered a page of student writing with correction marks, this only created negativity around the writing and did not allow for constructive change and improvement. According to Zinn, teachers needed to expect and provide writing exemplars that were high quality because students needed to have their writing evaluated using high expectations of success.

Zinn (1998) argued that students felt least threatened when their writing skills were assessed using an informal method. This included not giving student writing a grade. As students

learn about writing and improving their writing, teachers needed to create a motivating and positive environment where students were able to workshop their writing. Best practices for teachers who wanted to offer students feedback on their writing pointed out that the suggestions for improvement were meaningful, specific, and positive. The students needed to feel safe in taking risks and trying new styles or approaches to writing. Writing itself was a social process, and teachers encouraged students to interact with other students and teachers as they attempted to improve their writing abilities. Students were more likely to succeed in their writing efforts when they were given goals to strive towards in their writing. Students who were taught to reflect on their writing independently found inner motivation to read, revise, and improve their writing on their own without the immediate assistance of a teacher. Students needed to be taught to be the first critical eye that reviewed their writing responses and sought out areas that needed to be improved. Along with independent critiquing, students shared and discussed their writing with other students. In addition to the processes of editing and revision, students were given a variety of writing prompts and writing experiences. The more authentic a writing task, the more benefit students experienced as they transitioned into adult writers. Also, when students wrote under diverse conditions, it allowed teachers to evaluate their effectiveness in multiple writing experiences.

When evaluating writing informally, teachers did not need to formally mark and grade every aspect of the student writing. According to Zinn (1998), it was better to have minimal markings and grading on student writing as teachers could not possibly grade every writing assignment in a formal manner. Students improved as writers just by experiencing writing and writing frequently and often. When evaluating student writing, teachers needed to have clear and specific criteria that they used to determine the level of mastery within the responses. Students

needed also to be given the evaluation criteria as soon as they were exposed to the assignment in order to know what was expected of their writing and what was considered proficient.

ANALYTIC TRAIT SCORING AND HOLISTIC SCORING

Higgins, Miller, and Wegmann (2007) found that analytic trait scoring allowed students to benefit as they applied their understanding of the trait scores into their writing responses. They presented their beliefs that teachers effectively taught writing and assessed writing without succumbing to the pressures of teaching to the high stakes writing assessments their students took. They examined research studies that supported their argument that using the 6 + 1 Traits analytic scoring system for writing developed in the early 1980s increased student achievement and proficiency on state summative writing assessments. The six analytic traits that students were taught to understand and assess their writing in the 6 + 1 system were ideas, organization, voice, word choice, sentence fluency, and conventions. Arter, Spandel, Culham, and Pollard (1994) employed the 6 + 1 analytic scoring system as the focus of their research study. Students who were instructed on the 6 + 1 traits showed improved writing proficiency in the areas they were taught to understand. Thus, if students were taught the writing traits then they applied this knowledge to their own writing and improved these qualities.

Spandel (2006) supported the argument that writing rubrics were valuable tools for teachers to use for writing instruction. With a well-constructed rubric, teachers evaluated students' writing and offered guidance for improving the writing and communicated with the students about what strengths and weaknesses existed in the writing. Revision was a key component of the writing process, and writing rubrics were essential tools that guided students as they increased their proficiency as skilled writers. Teachers assessed writing using rubrics that extended beyond fixing spelling errors or writing longer essays; rubrics allowed students to

enhance their critical thinking behind the writing. Rubrics instructed by providing explanations of the criteria being assessed and also illustrated the ranges of possible proficiency of the writing responses. Students read their results on a rubric and recognized their strengths as writers which built confidence in them.

Coe (2000) concluded that when students were taught how to understand and self-assess their own writing using the six writing traits in the 6 + 1 scoring system, their summative writing assessment scores were greatly improved. For example, students who used the 6 + 1 analytic scoring system during writing instruction predicted future success in the state writing assessment. The study created a model that used the sum of the six traits' scores and this predicted success for 79% of the students. Thus, students who understood and used analytic writing scoring systems on their writing improved their writing in the areas of focus.

East (2009) investigated the reliability of scoring done by individuals who were using an analytic scoring rubric on a foreign language writing assessment. Holistic scores were considered more practical for high stakes testing typically because they offered a single score as the final assessment of students' writing. According to several researchers, holistic scores were often assigned to writing assessments not for the criteria within the rubric, but rather for the more cosmetic writing traits such as good penmanship (Charney, 1984; Grobe, 1981; Stewart & Grobe, 1979). Even scorers who were well-trained ended up scoring holistically based on only one or two criteria, not the entire set (Sakyi, 2001). In analytic scoring, each criteria of the rubric was assigned a score but this did not calculate into a holistic score. The analytic scores allowed for more specific attention to each criteria area and student writing weaknesses were targeted for focus and improvement (Weigle, 2002).

Rezaei and Lovorn (2010) argued for teachers to be trained to ensure that when using a scoring rubric, analytic traits or holistic, their subjectivity did not blur their assessment scoring process. They speculated that although teachers used rubrics to assess writing, they were not necessarily scoring student writing consistently or correctly. Their study investigated how reliable and valid rubrics were when used to assess students' writings. While rubrics were viewed by educators and researchers as valuable tools that created equality in scoring student writing, the researchers questioned the reality of this. They noted that rubrics can be holistic, analytical, or a combination of both. When teachers used rubrics to assess writing, they sometimes believed that they were removing subjectivity from the scoring. Objective scoring occurred when using writing rubrics for scoring, but it depended on how the rubric was used, not just that it was the assessment tool. The study concluded that teachers using the scoring rubric did not increase their reliability of grading. The teachers actually had more variance in their assigned scores when using a rubric versus when they did not use a rubric. The teachers felt compelled to assign low scores to students who had writing that contained spelling errors and grammatical mistakes, despite the content of the writing being conceptually superior to student writing samples that had excellent spelling and grammar but wrote off topic or had poor organization and development. Using rubrics did not guarantee objectivity and accuracy in the scoring of writing.

DelleBovi (2012) conducted an action research study on preservice education teachers to determine how best to prepare future teachers to assess students' writing abilities. While teachers recognized the importance of writing and teaching writing skills to students, they felt uncertain about how to implement the assessment practices in their classrooms. The research study used a six point rubric that scored holistically, but the teachers were also taught to assess students'

writing on six analytic writing traits individually and provide feedback on the six traits along with the overall holistic score. When rubrics were adopted for use in the classroom, it was critical that students be given the rubrics and taught how to understand their criteria. The most beneficial feedback extended beyond correcting grammatical errors and asking rhetorical questions such as “What do you mean here?” to clear statements that addressed the content of the writing. The study concluded with the following recommendations. When preservice teachers were given instruction on teaching and assessing writing, they were better informed and knowledgeable using writing in their content areas. The use of holistic rubrics was effective when the criteria within the rubric were used for instruction and feedback. Rubrics were best used to fit the purpose and content of a writing assignment.

Flateby (2010) examined whether one writing assessment rubric would be more reliable or valid than another. The study compared the Cognitive Level and Quality of Writing Assessment (CLAQWA) rubric with the Collegiate Assessment of Academic Proficiency (CAAP) rubric to see if students’ writing scores would be similar when scored on both rubrics. The CLAQWA rubric was created at the University of South Florida (USF) over 13 years ago and contained a cognitive and a writing scoring section within its analytic scoring traits. The CAAP only provided teachers with a holistic score. Determined to obtain a more formative writing assessment, the researcher revised the CLAQWA rubric and then trained teachers on its use in the classroom. The CLAQWA rubric was then taken online as the university provided student exemplars and comments on the exemplars that teachers and students accessed and reviewed to better understand the scoring rubric. When teachers used the CLAQWA rubric, they scored the essays and also gave written comments on the strengths and weaknesses of the writing which allowed the students to alter their writing skills. The researcher concluded that a

statistically significant relationship existed between students' writing analytic trait scores on the CLAQWA rubric and the CAAP holistic score which meant that the USF faculty should consider adopting the CLAQWA rubric. Obtaining a single holistic score was useful for data and accountability purposes; however, the analytic scores from a rubric impacted students' writing abilities and transformed them into skilled writers when used correctly.

Related to the reliability and validity of scoring rubrics, Knoch (2007) studied the accuracy of the scoring made by the individuals who used the assessment tool. This study compared two rating scales used to assess student writing for rater reliability. The study attempted to create a rating scale based on coherence in order to maintain an empirical mode of assessment. Using topical structure analysis (TSA) as the objective measure for coherence the individuals then rated the essays. The goal of the study was to determine if TSA was operational. This was proven when the scores of the raters were analyzed because the scores from the TSA measure were more accurate. In conclusion, the research proved that using empirical foundations when developing writing performance descriptors on writing assessment rubrics made for more accurate scoring. Empirically based descriptors were determined to be the best measure of reliability, not educator intuition, when rubrics were constructed.

Attali, Lewis, and Steier (2013) looked at another consideration related to rater reliability which was the use of automated essay scoring (AES) and human essay scoring. This study examined alternate ways of evaluating the reliability of human essay scorers using holistic rubrics. Often, assessing students' writing skills on high stakes assessments was subjective as the people scoring the essays were human and assessed subjectively, not objectively. A new mode of essay scoring known as automated essay scoring was becoming an acceptable and statistically supported mode of assessing students' writing. Prior research studies (Attali, 2007; Attali &

Burnstein, 2006) had shown that AES scoring closely resembled human scoring of writing. The limitation of AES scoring was that it did not assess the higher order thinking skills within the writing, such as the content accuracy and organization of ideas. The researchers hoped to find a method that increased human scorers' reliability and increased the reliability of combined AES and human scoring. They concluded that these results could be attained by creating distinctions between what the AES scoring assessed and what the human scorers assessed. An important finding from the Attali study (2007) was that scoring argumentative writing samples required the scorers to assess higher order thinking skills as human scorers held more agreement and the AES scoring showed lower performance. The informational writing samples scored by AES and human scorers were better aligned which suggested that the scoring of this genre was based on language control, not higher order thinking skills.

The perceived limitations of writing assessment rubrics being used in the United Kingdom were analyzed in another study. Fox (2000) argued that the current writing assessment rubric being used by teachers in the United Kingdom failed to accurately and concisely assess students' writing abilities. He described the current rubric, offered his criticisms, and then issued up his own version of the rubric which he argued aligned to the writing standards and allowed for a more authentic assessment of students' writing. Using the United Kingdom National Curriculum writing standards rubric, teachers were to score an essay with one of the best fitting descriptors which can be a W (working towards level 1), level 1, levels 2c, 2b, and 2a and level 3, which is the highest. Writing was evaluated using the criteria of three areas which were quality/style of writing, spelling and punctuation (together) and handwriting.

Fox (2000) criticized that the current assessment tool had unclear criteria because they were not indiscriminate of each other. He further stated that the progressions of development

were unclear as well and that no validity or reliability studies were conducted on the rubric. He pointed out that the lowest level of W may be too broad as students' writing may never progress outside of this descriptor. He also argued that there were other writing areas that could be included in the rubric such as development, use of the writing process, creativity, use of multiple genres, and independence of writing. Fox (2000) offered up his solution for a new writing rubric that still aligned to the writing standards. This rubric assessed seven dimensions of writing. These were handwriting, spelling, punctuation, vocabulary, syntax, genre, and communication of meaning. He also wanted to subdivide the W category into three parts which would allow for better assessment in this lower category. The dimensions were given numerical analytic trait scores and then averaged to find the students' proficiency levels in their writing.

Similar to the critique of the United Kingdom's writing assessment rubrics, another study Knoch (2011) provided guidance and recommendations for scoring rubric construction and use when assessing writing. This study pointed out the concerns around the creation, implementation, and scoring on a writing assessment rubric that was used diagnostically. The researcher noted that little research had been conducted on the diagnostic assessment of writing and its related rating scales that determined proficiency. With the goal of diagnostic writing assessment being to improve students' writing abilities, the following recommendations were made. First, diagnostic writing assessments were to use an analytic trait scale rather than a holistic one because this allowed writers to pinpoint areas of strengths and weakness. The rubric was to be constructed for use by both the writing scorers and the writing students because that best aligned to diagnostic assessment. The writing scoring rubric criteria was founded on a theoretical understanding of language and writing. The levels within the rubric were to be based on the goal of the writing assessment and the context for its use. Lastly, the reports that were

generated for the students to view and diagnose their writing skills were to be detailed and not an averaged number because the students benefited more from the descriptive explanations than a final score.

Dryer (2013) argued that creating and constructing writing rubrics were not easy tasks, and he pointed to the 83 different writing rubrics that states have provided for teachers as evidence of the difficulty. Two main features of writing rubrics were the analytic traits and levels of performance in the rubrics. When essays were scored using a rubric with specific descriptions and qualities, then the individuals scoring the essays tended to perceive and evaluate the writing better (Mills & Jaeger, 1998). The findings of the study suggested that the analytic traits, style, organization, and thesis, encouraged the essay scorers to evaluate the ideas behind the writing. Second, the definition and choice of terminology within the performance levels were recognized by the essay scorers as being unique to each writer and essay genre which was improved from the notion that one set of performance levels fit all writers and purposes. Lastly, essay scorers realized that writing is not error free, even the best writing may contain errors and this is acceptable (Dryer, 2013).

AUTOMATED ESSAY SCORING (AES) SYSTEMS

Emerging in the 1960s, automated essay scoring (AES) systems began to offer researchers the ability to score student writing in mass quantities. As the push for high stakes writing assessments emerged and with advances in technology in the 1990s, AES systems became a researched and valid mode of assessing student writing performance (Warschauer & Grimes, 2008). Warschauer and Grimes explained the history of AES software as an assessment tool for writing. Supporters of AES believed that it relieved the burden of grading student writing from teachers and allowed them more time to teach writing. Opponents of AES believed that it

was a gateway leading to controlled teacher behaviors and an eventual replacement of teachers by machines. AES first appeared as an assessment program in the 1960s. Using multiple regression analysis of textual features, the scoring mechanism scored writing based on hand graded papers (Shermis, Mzumara, Olson, & Harrington, 2001). In the 1990s, AES reappeared with improved functionality as several companies offered competing programs for schools to use after the push of high stakes testing began to take shape. The validity of AES scoring was assured by using the same type of calculations used for human scorers. Thus, the writing was assessed by two humans, and the AES replaced one of the human scorers. Studies (Chodorow & Burnstein, 2004; Elliot & Mikulas, 2004) conducted on the reliability of AES systems have shown that AES scoring agreed or agreed within one point from a human scorer more than 95% of the time.

There were many different types of AES systems used in the United States. Dikli (2006) Warschauer and Ware (2006) summarized the main AES programs available which were Project Essay Grader (PEG), Intelligent Essay Assessor (IEA), E-rater and Criterion, IntelliMetric and MYAccess!, and Bayesian Essay Test Scoring System (BETSY). PEG, which emerged in 1966, was shown to have similar scores to humans. It had the capability to note errors in the writing it scored. It focused more on the mechanical and grammatical errors, and it received criticism that it was unable to accurately score organization and development. IEA scored written responses using the process of latent semantic analysis (LSA) which meant that the scoring engine compared the semantics within a training set of papers with the semantics of student essays. This meant that the responses were read for semantic meaning and then compared to masses of writing (Landauer, Laham, & Foltz, 2000). IEA scored the organization and development of written responses. It was created by the University of Colorado and belongs to Pearson

Education. IEA was unique because it could be transplanted into any online program and be leased to it. IEA used multiple sources as comparable documents for scoring student writing and required fewer hand scoring anchor papers for training. Students received either a holistic score or feedback comments (Foltz, Gilliam, & Kendall, 2000).

Dikli (2006) noted that E-rater and Criterion both scored writing using the linguistic features of the writing to guide its scoring which was called artificial intelligence (AI) and it employed natural-language processing (NLP). NLP consisted of speech recognition which continuously diagnosed speech into known words, syntactic analysis which recognized word clusters such as noun and verb phrases, discourse analysis which analyzed the context of sentence structures and meanings, information extraction which located and extracted text, and machine translation which translated one language to another. E-rater scoring was formed by papers that were scored by two humans on a holistic scale of six points. Criterion relied on the E-rater scoring tools and it also analyzed writing and gave feedback beyond the holistic score that E-rater gave. Criterion was maintained by Educational Testing Services (ETS). Criterion was originally used primarily in high stakes writing assessments as a supplemental scorer with human scorers (Burstein, 2003; Kukich, 2000). It used natural language processing (NLP) and an assistive tool known as Critique to score essays which gave individualized feedback that was unique to each student's needs.

Warschauer and Ware (2006) noted that MY Access! created by Vantage Learning was used in public schools. Its scoring engine had a trademarked name, IntelliMetric, and it scored student writing by analyzing features in the writing and comparing them with hand scored essay samples (Elliot, 2003). IntelliMetric was the first AI based scoring program. MY Access! used the IntelliMetric system and allowed for instructional uses beyond just a score report as it

provided diagnostic feedback on the students' writing. MY Access! gave written responses a holistic score on either a range of 1-6 or a range of 1-4. Additionally, it scored the analytic traits of focus and meaning, organization, content and development, language use and style, and mechanics and conventions. Students read feedback about their writing, and they used an editing tool within the program that allowed them to correct spelling and grammar. BETSY was considered a research tool, not a commercial product provided by assessment vendors. It classified text after being trained on its contents.

Warschauer and Ware (2006) also looked at the types of research that had been conducted on AES systems. One area of research focused on the psychometric issues related to AES. AES systems had scored 95% or more of the time in agreement with human scorers which was statistically valid (Cohen, Ben-Simon, & Hovav, 2003; Keith, 2003). Another study found that AES scoring was best used in standardized, high stakes testing; however, it also worked effectively in classroom settings as well (Powers, Burstein, Chodorow, Fowles, & Kukich, 2002). One big question asked about AES scoring was how effective it was at improving student writing. A study conducted by Vantage Learning found that students using MY Access! had improved writing scores from 2.00 to 2.84 in a six week period (Elliot & Mikulas, 2004). Another Vantage Learning study found that 81% of students who used MY Access! prior to taking the California High School Exit Examination passed it (Elliot & Mikulas, 2004). A concern expressed by Warschauer and Ware (2006) was that the current studies were all conducted by testing vendors.

Warschauer and Grimes (2008) wanted to look at how effectively AES was when used in the classroom formatively to improve student writing. Although small in number, one study of AES determined that students only submitted their essays one time (71% of the time) which was

in contrast with the limitless opportunities that students had to revise and rescore their essays in AES systems (Attali, 2004). Warschauer and Grimes (2008) wished to add to the AES systems research. They found that teachers and students perceived the value of using a formative AES writing program; however, they did not use the program frequently. Additionally, the AES systems offered the main benefit of allowing students to revise and redo their writing to improve it; yet, the students who revised between drafts of writing did only superficial revisions.

Warschauer and Grimes (2008) agreed in their findings with Attali (2004) as 72% of students only submitted one version of their writing to be scored. The opportunity to revise and resubmit was there, but the majority of students did not complete this task. Interestingly, the complaint that AES writing programs had the negative effect of replacing the teachers' abilities to teach was not found. The teachers in the study had more time to teach writing; however, other programs and curriculum requirements took time away from any writing instruction that might have been done.

Warschauer and Grimes (2008) also examined another concern expressed which was AES systems took control from the teachers because they were not able to teach writing in its shadow. This was also unfounded because the teachers who taught the writing process still taught it, and teachers who did not enjoy teaching writing did not teach writing in much depth to their students. Lastly, some argued that AES systems caused students to write in static and formulaic modes; however, this study showed that AES scored creative writing effectively, but the high stakes assessments of writing encouraged more formulaic writing. They concluded that AES systems did not offer all solutions nor did they offer only disappointment. It was a tool that can be used positively or negatively depending on the contexts of its use (Warschauer & Grimes, 2008).

Rather than dismiss AES systems, another theorist (Dean, 2013) argued that modifying the current use as a high stakes testing system to better align with the philosophical beliefs held by writing theorists and teachers allowed the systems to better serve students. Dean determined that AES systems were being used as high stakes testing tools; however, they needed to be adapted to measure more than writing skills in a timed, testing environment. AES systems needed to be altered to better fit with the philosophical beliefs held by writing theorists. These beliefs were that writing was a social and humanistic process. AES systems could be extended beyond their current use as a summative assessment, and they needed to be formative writing programs that better aligned to the teachings and philosophical views of writing teachers. AES systems functioned within the contexts of their purpose, and this purpose historically was to assess writing in a high stakes assessment method. Dean (2013) argued that the current operating purpose needed to be adapted in AES systems by altering the algorithms and programming to accommodate more formative practices.

Dean (2013) noted that one area of misunderstanding related to using the current AES systems outside of summative writing assessment was related to the terminology of the analytic trait scoring categories. For example, the scoring engine scored organization and development based on the units of text in the student writing; however, a human scorer defined these terms as being related to the specific content of the writing. AES systems used features within the programming that recognized errors using algorithms and textual analysis that looked for patterns within the texts. Dean also noted that although content was not recognized by AES systems as of the publication of the article, research was being conducted to accomplish this using NLP. In addition to the AES systems analyzing texts based on their features, the systems also used statistical model training which allowed the engine scoring to align with human

scoring. A set of training papers which were scored by humans were used as the foundation of AES scoring systems. When the training set was specific to the prompt being written to and scored, the better the agreement between the AES system scores and human scores was. Studies (Bejar, 2011; Ben-Simon & Bennett, 2007; Habermand & Sinharay, 2010) found that statistical significance existed when creating an AES scoring engine if the statistical methods and procedures were analyzed.

According to Attali, Bridgeman, and Trapani (2010), the two dominant methods used to develop AES scoring systems were prompt-specific and generic. Both were determined by the training done and the set of papers used in the training. While prompt-specific models were built by using student writing to a specific prompt which became the training set, generic models used student writing that did not address a specific prompt. Generic models were trained using collected paper samples that were the same genre of writing, such as descriptive, narrative, informative, and persuasive essay types. Studies (Klobucar, Elliot, Deess, Rudniy, & Joshi, 2013; Ramineni, 2013) concluded that the prompt-specific models scored more accurately than generic models. Another intricate element within the model designs was the importance of quality student samples to act as foundational texts for the training sets. Thus, when AES systems were found to be limited or inaccurate, it typically stemmed from the methods used to create the scoring engine.

Ramineni (2013) concluded that an AES system created to be unique to the particular needs of a college's writing placement program produced results that correlated with human scorers better than generic prompts being used. The study used Criterion Online Writing Evaluation Service to build its AES scoring system and used four unique prompts and training sets of papers on those prompts as the basis for the AES scoring. The unique prompts that were

customized by the university faculty scored comparably between the scoring engine and the human scorers while the generic prompts offered no agreement in the scores by the engine and the human scorers.

Klobucar et al. (2013) analyzed the effectiveness of an AES system used by universities to assess writing abilities of at-risk students for classroom placement. In addition to researching the AES scoring system, the study also looked at student and teacher perceptions of the AES system. They concluded that the AES system accurately assessed the student writing samples, aligned to the university's goal for its use as an assessment tool, and was accepted by the students and teachers. No statistical differences were found in the scores based on variables, such as ethnicity or race, and it accurately identified at-risk students who needed remedial placement.

James (2006) noted that automated essay scoring (AES) writing programs were being used for placement in colleges and universities. The researcher hoped to determine whether there was more accuracy in the AES scoring versus the scoring of college writing professors. While historically studies have shown that AES scoring and human scoring were comparable, the question raised was if the human scorers in the studies were calibrated to the scoring rubric and mimicked the AES scoring because of this. James (2006) decided to use non-calibrated writing professors to determine if any changes would occur in the scoring agreement between the ACCUPLACER OnLine WritePlacer Plus program that used the IntelliMetric scoring engine and the writing professors. She also wanted to see if predictive findings could be made based on the essay scoring of both the AES system and the college professors. The study concluded that scores corresponded well between the AES system and the professors. Since the study used untrained and non-calibrated professors to score the student writing, it provided even greater validity to the question of reliability of AES scoring in the program. Additionally, when

combining the AES scores and the human scores to predict future success as students, the combined scores were 77% correct which was a better indicator than the 75% produced by the professors' essay scoring and the 70% made by the AES scoring system. Thus, when used in combination, the AES scores and the human scores greatly predicted future success in the students (James, 2006).

A unique study by McCurry (2010) in Australia wanted to see if an AES system could adapt itself and accurately score essays that were written to open-ended writing prompts which were prompts that asked students to select any topic and write an informative or argumentative essay. McCurry (2010) conducted a study on the use of AES scoring of open-ended writing prompts to determine if the scoring engines would have rating similarity with human scorers. The open-ended writing prompts which were called the AST Writing Test created by the Australian Council for Educational Research required students to select a social issue and then construct an essay that represented their opinions on the topic. Two different AES systems were given training papers from another AST Writing Test that were used to calibrate the scoring engines which were then used to score the essays from the study. Neither AES systems agreed with human scoring which indicated that the reliability and validity of using AES scoring on these open-ended writing prompts was ineffective.

Scharber and Dexter (2004) conducted a study that gathered and analyzed student perceptions of a formative writing assessment program that provided automatic scores on student writing. They wanted to see what reactions students had when they used an automated essay scoring program formatively. Prior studies (Clariana, 1993; Peat & Franklin, 2002) had shown that computer-based formative assessments had improved learning. Another important aspect of computer-based formative assessments that studies found was the valuable feedback they offered

students (Charman, 1999). Scharber and Dexter (2004) noted that literature on student learning and performance affected by formative online programs was scarce. Two prior studies (Boyle, Bryon, & Paul, 1997; Charman & Elmes, 1998) on formative online assessment illustrated positive effects on student learning. One major finding that emerged from Scharber and Dexter's study (2004) was the emotional reaction that students had as they used an automated formative essay scoring program. Because of this finding, the researchers suggested that future use of formative essay scoring programs needed to address the emotions that students had towards a computer scoring their writing. They also concluded that feedback for students using automated formative programs needed to be more extensive than a simple number score. It needed to offer students specific guidance and details of how to improve their writing.

While not against AES systems specifically, one researcher (Condon, 2013) argued that the entire structure and philosophy of high stakes online writing assessment were erroneous and need replaced. He analyzed the different types of writings and the modes in which they are assessed. He argued that AES systems were unfairly pointed to as unacceptable assessment systems. Instead, the high stakes writing assessment system as a whole needed to be stopped because it was what was faulty. Rather than use statistical data in a writing assessment score, students needed to be evaluated in richer modes, such that provided data on the complexity in which writing was grounded. High stakes writing assessments were too restrictive in manner to offer useful insights and assistance to improve student writing ability. AES systems were not the problem, but the type of writing that the AES systems scored was the problem. High stakes writing assessments needed to be abandoned and replaced with authentic writing assessments.

Shermis (2003) addressed the two criticisms towards AES first that the scoring engines were unable to follow logic in an argument. Second, critics argued that computer programs

stifled the human interaction needed for writing. Shermis agreed that AES scoring was not able to assess with definite accuracy the content of the writing; however, he pointed out that students writing to cheat the engine were good writers in the first place and knew the information in order to alter it or cheat with it. He also discussed that there was new technology in AES scoring called discourse analysis which allowed the program to summarize the text it was scoring and determine what was the main idea or thesis sentence which was shared with the student to determine if this was in fact the argument or point the student was attempting to make in the essay. New linguistic evaluations were being developed that enabled the AES scoring to distinguish spelling errors in numerous variations, such as 67 misspellings for a word, and still recognize the misspelled word for the word that was intended by the student.

Foltz, Gilliam, and Kendall (2000) tested the scoring ability of an AES system that used LSA to see if the program assessed the content of the essays as well as the writing ability of the students. They determined that when the engine scored the essays higher the essays were also factually correct as well. Thus, students who understood the content they were communicating were able to express themselves in written expression expertly as well. The study concluded that the quality of writing reflected the quality of the content of the writing.

According to Dikli (2006), AES systems provided instant essay scores and feedback. In order to improve writing, feedback needed to be made by teachers and then students needed to be able to apply their understandings based on the feedback to their writing. AES systems were successful when they gave correct feedback and allowed teachers and students to communicate about the writing. This mode of essay scoring was also beneficial for high stakes writing assessments because of the cost of scoring large samples of student writing. One shortcoming of AES systems was that they were unable to score the writing process; only the final product of

writing was assessed. Another shortcoming with AES programs was that they were typically in English, not other languages, and this created challenges for non-English users. AES scoring often received criticism that writers tricked the scoring engine, and vendors were actively creating algorithms that flagged cheating attempts.

FORMATIVE AND SUMMATIVE WRITING ASSESSMENT PROGRAMS

Since the 1990s, both formative and summative writing assessment programs have been given to students as methods of evaluating their writing skills. While the functions of these programs differed depending on the purpose of the institution, the reliability and validity of the programs' abilities to effectively and accurately score student writing and thus determine student proficiency in writing was considered.

The AES program, Writing Roadmap 2.0, which was the precursor to WV Writes in West Virginia, was the focus of several studies which were done to determine its effect on the state standardized test WESTEST 2 Online Writing. In one study, Harrington and Rich (2006) argued that student performance on the WESTEST 2 Online Writing would be affected by their usage of the practice writing program, Writing Roadmap 2.0. The study concluded that across the counties included in the sample, there was a significant improvement on the high stakes writing scores when the students wrote essays in Writing Roadmap 2.0. The size of a county also impacted the student score improvements as larger counties showed better score increases than small counties. Also, counties who had low test scores from 2005 used Writing Roadmap 2.0 more than those that scored at normal levels.

Rich, Harrington, Kim, and West (2008) reviewed two aspects related to West Virginia adopting an AES writing assessment system including both formative (Writing Roadmap 2.0) and summative assessments (WESTEST 2 Online Writing). First, they considered how West

Virginia set up the validation of its high stakes online writing assessment. Second, the study considered the effects of a practice writing program, called Writing Roadmap 2.0, on the writing scores of the high stakes assessment.

According to Rich et al. (2008), from 2005 to 2007, the state of West Virginia used an AES program to score the writing of its 7th and 10th graders to determine their writing proficiency. Prior to 2009, the writing scores students received were not calculated into their accountability scores for No Child Left Behind (NCLB); however, the scores were to be used from 2009 to 2014 as part of the Reading Language Arts (RLA) scores. Along with using the high stakes online writing assessment, West Virginia began using a practice writing program called Writing Roadmap 2.0 from 2005-2007 for the 7th and 10th grade students being assessed in writing.

Rich et al. (2008) further explained that the first part of the process for West Virginia was to validate the AES system for accurate scores. The WESTEST 2 Online Writing used a prompt-specific scoring engine and relied on a scoring algorithm that used artificial intelligence, natural language processing, and statistical model processing. There were two parts of validating the AES system. The first was to train the scoring algorithm using scored sets of training papers that had been scored by experts in writing and hand scoring. The AES system was then provided student samples that have no set score and allowed to score the training set on its own. The scores given to the papers by the scoring engine were then compared with human scorers to determine that the AES was scoring in agreement with human scorers. Calculations were made between the scoring engine and the hand scorers to ensure that acceptable agreement was occurring. Typically, 40-70% agreement was considered acceptable in the validation studies.

West Virginia used approximately 600 training papers for the validation study. Three hundred were used in the first stage to train the scoring engine, and the remaining were used in the blind scoring in the second stage. For each high stakes prompt, if the inter-rater reliability fell below the accepted range, the scoring engine was retrained using more student samples and calibration with human scorers (Rich et al., 2008). Any prompts that were unable to reach the acceptable agreement ratings between the engine and humans were dropped as assessment prompts. These prompts would later be used in the WV Writes program as the WESTEST 2 practice prompts. One concern that emerged during the validation study was that some essays were flagged as unable to be scored by the engine. West Virginia, with the assistance of CTB/McGraw Hill, determined that a process of flagging before scoring would be put in place in the high stakes assessment (S. Foster, personal communication, August 2013). The flagged and unscored essays were directed to human scorers who hand scored the essays rather than having the essays receive inaccurate flagging or low scores based on their unusual characteristics.

The second consideration of the study (Rich et al., 2008) was to determine if student usage of Writing Roadmap indicated higher scores on WESTEST 2 Online Writing. The data used for this study came from two years of Writing Roadmap usage in 2006 and 2007. Additionally, the study looked at performance levels of students on the Online Writing five analytic traits and their scores on the Writing Roadmap five analytic traits. Gender and ethnicity were also considerations in the study. It was concluded that students who practiced writing four essays or more in Writing Roadmap had the largest score gains on WESTEST 2 Online Writing.

White, Hixon, D'Brot, and Perdue (2010) measured the impact, if any, of Writing Roadmap 2.0 on the writing scores of students taking WESTEST 2 Online Writing. The study determined that statistical significance existed in the difference between the use of Writing

Roadmap 2.0 five or more times by students and their WESTEST 2 Online Writing scores. Thus, students who wrote five or more essays in Writing Roadmap 2.0 scored higher on WESTEST 2 Online Writing than those students who did not use Writing Roadmap 2.0. The strongest relationship between Writing Roadmap usage and better WESTEST 2 Online Writing scores was at the elementary grades with middle and high school students showing a decline in the relationship.

Beyond AES programs, several studies have been done on writing assessment overall. Llosa, Beck, and Zhao (2011) conducted a study to determine the most common types of writing genres required of students in high stakes writing assessments in New York City schools. They also looked at what differences if any existed between English Language Learners (ELLs) and non-English Language Learners (non-ELLs) as they experienced these writing assessments. They had two major findings. The first was that argumentative writing was the main genre used to assess students in their study, and the second was that differences existed between the (ELLs) and (non-ELLs) with how they expressed their ideas in written work.

The researchers (Llosa, Beck, & Zhao, 2011) pointed out that there was not adequate information about secondary school students and their unique difficulties with learning to write because most writing assessment programs did not offer feedback beyond the analytic trait scores the students received on their scored essay. There was not any instructional feedback or narrative about what students needed to improve in the revisions of their essays. The score by itself was valuable because it showed whether or not students had mastered the writing task asked of them, but it did not offer specific information. They believed that students needed to be granted access to diagnostic tools to improve their writing. Students needed to have their writing weaknesses diagnosed and then be allowed to strengthen their writing skills by making the

suggested improvements. Another limitation the researchers pointed out in the existing high school writing programs were the types of essays that the writing prompts asked students to respond to were typically narrative essays. The narrative genre did not require the higher order skills and critical thinking that persuasive or argumentative essay writing genres required of students, so the students were not being exposed to more challenging types of writing genres.

In the findings of their study, the researchers (Llosa, Beck, & Zhao, 2011) learned that the state standards addressed several genres of writing and the classroom teachers taught many diverse genres of writing. These included personal narratives, informational reports, explanation essays, exposition (argumentative or persuasive) essays, and other such as journal entries. While the state standards and the teachers were exposing students to these multiple types of writing genres, the high stakes writing assessments for high school ELL students asked students to write only one type of essay which was the explanation essay. The ELL students had little to no exposure to explanation writing genres because multiple writing genres throughout the school year were addressed and explanation essays were not usually covered in instruction during high school. The ELL students were being assessed on their writing with a genre that they were unfamiliar with and had little knowledge of its format and the expectations associated with the genre. The high school non-ELL students were assessed in the high stakes writing assessment mainly on exposition writing which they had been exposed to from the standards and the instruction of their teachers as the researchers were able to document in their studies and findings.

The second study the researchers (Llosa, Beck, & Zhao, 2011) conducted looked at what the students found challenging while writing their essays. The most prevalent issue that students faced was the struggle to find a way to express their thoughts into writing. The researchers used

the term, translation, to describe this process that both ELL and non-ELL students progressed through as they composed. Within translation, the students had trouble with grammatical errors and misspellings; in addition, they also struggled to select what words best expressed their thoughts. The second difficulty for students was the ability to state their opinion and generate enough reasons, details, or examples to support their stance on a topic.

The researchers (Llosa, Beck, & Zhao, 2011) concluded with the belief that states needed to offer high school students better diagnostic writing assessments prior to the high stakes writing assessments. It was noted that in order to improve student writing, there needed to be opportunities for students and teachers to use diagnostic writing assessments that offered students more than an analytic score. The feedback from diagnostic writing tools needed to allow students to improve their writing by giving them detailed information about their writing strengths and weaknesses which were not available in the New York City schools' high stakes writing assessments. One concern was that teachers were only teaching the writing genres that students were assessed on in the high stakes writing, but the study found otherwise because teachers were teaching students multiple genres of writing in the classroom. The main shortcoming was for ELL students because they did not receive any explanation genre writing instruction, but their high stakes writing assessment used only that genre of writing prompt. The recommended change from this study was that schools needed to offer students a diagnostic writing program that would provide helpful and specific feedback for improving writing within the genres of writing modes that the high stakes writing assessment would require students to use. The diagnostic program needed to allow students to breakdown and individually address the writing traits within isolation, rather than a broad recommendation being offered by the diagnostic assessment.

Olinghouse, Zheng, and Morlock (2012) noted that when reviewing the current state high stakes writing assessments, several similarities were found in them. The writing assessments were usually referred to as on-demand or direct writing assessments. The students had a certain amount of time in which to read a prompt, organize their thoughts, and write a response to the prompt. Similarities existed among state high stakes writing assessments mainly in the purposes that the assessments were given. First, the results of the student writing assessments acted as indicators of student achievements. These snapshots often had implications for students, teachers, schools, and districts as they related to state and federal accountability. The high stakes writing assessments had a similar format which was a picture or text prompt that required students to respond in writing. Typically, states did not connect the writing assessment with other content areas such as reading language arts or social studies. Additionally, states used writing assessments to monitor the student achievement levels related to NCLB and to standardize writing assessment across students.

Brimi (2012) explored the effect of a summative writing assessment in Tennessee on the teaching of writing to its students. His study wanted to determine what impact the summative writing test had on how the teachers taught writing in the areas of preparing for the assessment, using the writing process, and practicing multiple genres of writing. Also to be considered was the background of the teachers, including how they learned to teach writing, and if the summative writing assessment altered in any way their teaching methods. When analyzing the teacher interviews, Brimi concluded that the Tennessee Comprehensive Assessment Program (TCAP) had affected how the teachers approached their instruction of writing to students. The study found that the teachers had little formal training in writing instruction, and they did not feel comfortable teaching the writing process or the different writing modes associated with the

different writing genres. Thus, this study showed that teachers were more comfortable teaching writing in preparation of a summative test than to teach the process of writing.

White (2004) considered the argument that writing assessments needed to be better connected with writing theory. Writing assessments were too focused on the mechanical aspects of scoring writing, and this needed to be revised to incorporate the theories and research of writing into the assessment process (Huot, 2002). White (2004) noted many of the concerns related to writing assessment, such as too much attention placed on the technology of writing assessment which led to validity in writing assessment becoming too simplistic. Writing assessment and writing practice needed to be united in order to accurately assess students' writing skills. When considering the faulty assumptions currently being held about writing assessment and writing practice, these represented the current mindset. When leading writing programs, the importance of the leader being an expert in assessment was not a consideration. A writing program was created and implemented without assessment being incorporated into the program. Non-educators may have pressured states to administer writing assessments. Determinations about writing and writing programs were made by individuals who were not knowledgeable in the field of writing or in implementing writing programs in schools.

CHAPTER THREE: METHODS

INTRODUCTION

In order to best understand the data collection used in this research, this chapter describes the methods and data analysis. In the following pages, the research design, population, instruments, procedures, and data analysis are explained.

RESEARCH DESIGN

This study analyzed the effect of writing prompts on students' writing proficiency among 11th graders. The type of research design used for this study was a quantitative, post hoc, 2 x 2 ANOVA. In this design type, the experimental group received an intervention treatment while the control group did not receive the treatment or intervention. The students' composite writing scores and the five analytic traits scores of organization, development, sentence structure, word choice/grammar usage, and mechanics on the 2013 WESTEST 2 Online Writing were the posttest data. The composite scores ranged from 0-30 and the five analytic trait scores ranged from 0-6. The posttest data were used to measure the students' writing proficiency after the intervention was completed.

The two independent variables included the type of prompt that students wrote to in the WV Writes program and the gender of the students. The factors of the writing prompt independent variables were the Writing Roadmap prompts and the WESTEST 2 practice prompts. The factors for gender were males and females. The dependent variables were the measured effects on writing proficiency scores. The experimental group consisted of 190 11th grade students who practiced writing using WESTEST 2 Practice prompts in WV Writes. The

control group included 190 11th grade students who practiced writing using the Writing Roadmap 2.0 prompts in WV Writes.

POPULATION AND SAMPLE

After collecting the student data, it was found that 39,464 11th grade students from across the state of WV took WESTEST 2 Online Writing in 2013. In order to get an accurate count of students who met the conditions of using either WESTEST 2 practice prompts or Writing Roadmap practice prompts, the researcher filtered the 39,464 students in order to identify the population of 6,459 students who had used WV Writes as 11th graders prior to taking the WESTEST 2 Online Writing in 2013. Of that population, 3,521 students used WESTEST 2 practice prompts (the treatment group) and 2,938 students used Writing Roadmap practice prompts (the control group). Once these groups were established, the researcher conducted random sampling for each category: those who used WESTEST 2 practice prompts and those who used Writing Roadmap practice prompts. In order to achieve the desired number given by the sample size calculator (380), the WESTEST 2 practice prompt student group was uploaded into the SPSS system and was used to randomly select 190 students from the 3,521. Similarly, the sample for the control group who used Writing Roadmap practice prompts was determined using SPSS to randomly select another 190 students from the 2,938 who had practiced using this type of prompt.

PROCEDURES

This study required approval from the Institutional Review Board (IRB) (see Appendix B). Additionally, the West Virginia Department of Education (WVDE) approved this study. The researcher submitted a WVDE Research Proposal Application on October 31, 2013 and was granted permission to access student scores and data pending approval from the IRB and the

researcher's doctoral committee. The researcher requested a data file that was downloaded into Excel for review and input into the Statistical Package for the Social Sciences (SPSS) Version 21. The comma, delineated data file from the West Virginia Education Information System (WVEIS) contained student data from the 11th grade state population including a unique student identification number, grade, county, school, gender, and writing scores on WESTEST 2 Online Writing for 2013 (11th grade year) for each student. No student names were included in the file because they were not necessary to this study.

The writing scores identified for each student number were broken into five analytic traits including organization, development, sentence structure, word choice/grammar usage, and mechanics. These five trait scores were associated with each student's number for the 11th grade WESTEST 2 Online Writing Assessment (posttest). The five analytic trait scores were calculated into a composite score that ranged from 0-30.

The researcher was also given a WV Writes data file in comma delimited form which was downloaded into Excel for review and input into the SPSS system. This file included the student number, grade, county, school, gender, and writing scores on the practice prompts in WV Writes for the 2012-2013 academic year. The students were categorized into either the experimental group or the control group based on whether they wrote essays for WESTEST 2 practice prompts or Writing Roadmap prompts.

INSTRUMENT

Data were collected from the WV Writes practice writing program database that captured and saved student writing and writing scores for one academic year. From September 2012 to May 2013, the experimental group wrote essay responses in WV Writes to the WESTEST 2 practice prompts, and the control group wrote essay responses to the Writing Roadmap prompts.

The students received scores in the WV Writes program for the five analytic writing traits of organization, development, sentence structure, word choice/grammar usage, and mechanics on the essay responses they composed. These five analytic traits were reported in the WV Writes program and also made available in a downloadable format. The posttest instruments were the 11th grade (taken in the spring of 2013) WESTEST 2 Online Writing scores. The five analytic traits were also reported in WESTEST 2 data results and were available in a downloadable format. The analytic trait scores were computed into composite scores ranging from 0-30.

DATA ANALYSIS

WV Writes composite scores of both groups were averaged into the composite scores that ranged from 0-30. The composite scores of the students on the 2013 WESTEST 2 Online Writing in the experimental and control groups were analyzed using the SPSS system. The specific descriptive procedures and inferential statistical techniques needed to analyze the data are described in Table 2 in regard to each of the research questions posed for this investigation.

Table 2 Descriptive and Inferential Analyses for Research Questions

Research Question	Data Analysis	Statistics
1. What are the effects on WESTEST 2 Online Writing composite scores among 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?	Descriptive and Inferential	Means, Standard Deviations, Variances, and t test
2. What differences exist among the five analytic writing traits on WESTEST 2 Online Writing for 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?	Descriptive and Inferential	Means, Standard Deviations, Variances, and t test
3. What are the effects of gender and types of prompts on WESTEST 2 Online Writing scores among 11th graders?	Descriptive and Inferential	Means, Standard Deviations, Variances, and 2 x 2 ANOVA

CHAPTER FOUR: FINDINGS

The purpose of this study was to determine the effects of writing prompts in a practice writing program on students' writing proficiency among 11th graders. Presented in this chapter are the descriptive data and statistical analysis for the three research questions:

1. What are the effects on WESTEST 2 Online Writing composite scores among 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?
2. What differences exist among the five analytic writing traits on WESTEST 2 Online Writing for 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?
3. What are the effects of gender and types of prompts on WESTEST 2 Online Writing scores among 11th graders?

POPULATION AND SAMPLE

This study had a population of 6,459 11th grade students enrolled in West Virginia public schools. These students had all taken the WESTEST 2 Online Writing as 11th graders in the spring of 2013, and in preparation for the year-end, state level writing assessment completed either Writing Roadmap 2.0 prompts or WESTEST 2 practice prompts. The student population of 6,459 was categorized into two groups: those who practiced writing using WESTEST 2 practice prompts (3,521) and those who practiced writing using Writing Roadmap prompts (2,938). Using a random sampling technique in SPSS, the researcher selected 190 students from the WESTEST 2 practice prompt group and 190 students from the Writing Roadmap practice prompt group. This made the student sample for the study 380 students.

MAJOR FINDINGS

The data were obtained from the WESTEST 2 Online Writing composite scores for research questions one and three. For research question two, the data were obtained from the five analytic trait scores that comprised the WESTEST 2 Online Writing scores. Using SPSS, the researcher analyzed the data using descriptive and inferential statistics. The results are reported for the research questions below.

Research Question One

What are the effects on WESTEST 2 Online Writing composite scores among the samples of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts? This research question was answered by analyzing the data from the 380 sampled students and also the entire population of 6,459 students using descriptive statistics and an independent samples t-test.

Descriptive statistics for sampled students. Table 3 shows the mean posttest scores for the two types of writing prompts. As can be seen, there is very little difference in mean scores for the prompt conditions, although scores for Writing Roadmap practice prompts resulted in greater variability (standard deviation) compared to the WESTEST 2 practice prompts. The Writing Roadmap prompt students' mean score was 19.86 (*SD* 4.49) with the standard error mean of .326. For the WESTEST 2 practice prompt students, the mean score was 19.37 (*SD* 3.85) with the standard error mean of .279. In both cases, the mean scores were similar to each other. There was greater variance for the Writing Roadmap prompt students when compared to WESTEST 2 practice prompt students.

Table 3 RQ1 – Mean WESTEST 2 Posttest Scores for WESTEST 2 and Writing Roadmap Prompts

	Group Statistics				
	SAMPLE	n	Mean	Std. Deviation	Std. Error Mean
WESTEST 2 2013 POSTTESTSCORES	WRITING ROADMAP	190	19.8684	4.49674	.32623
	WESTEST 2 PRACTICE	190	19.3789	3.85361	.27957

T-test for Sampled Students. Descriptive data were further analyzed with a t-test for independent samples to determine if there were significant effects for WESTEST 2 Online Writing composite scores for the types of prompt groups (see Table 4). The hypotheses were:

1. Null hypothesis: There are no differences in the WESTEST 2 Online Writing composite scores among the samples of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.
2. Alternate hypothesis: There are differences in the WESTEST 2 Online Writing composite scores among the samples of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.

Based on the data in Table 4, no statistical significance was found on the WESTEST 2 Online Writing scores between the prompt groups in the sampled population. A probability of .255 ($p > .05$) does not indicate a difference in using one type of prompt over another. Thus, the results failed to reject the null hypothesis.

Table 4 RQ1 – WESTEST 2 Posttest Scores for WESTEST 2 and Writing Roadmap Prompts

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
WESTEST 2 2013 POSTTESTSCORES	Equal variances assumed	1.139	378	.255	.48947	.42963	-.35530	1.33424

Descriptive Statistics for Total Population of Students. Table 5 shows the mean posttest scores for the two types of writing prompts among the population. As shown here, there is very little difference in mean scores for the two prompt conditions as noted by the data in Table 5. Additionally, these groupings varied fractionally in all cases regarding standard deviations and standard errors or its variability.

Table 5 RQ1 – Mean WESTEST 2 Posttest Scores for WESTEST 2 and Writing Roadmap Prompts

	TYPE OF WRITING PROMPTS USED FOR PRACTICE	N	Mean	Std. Deviation	Std. Error Mean
WESTEST 2 2013 POSTTESTSCORES	WRITING ROADMAP	2938	19.8213	4.17840	.07709
	WESTEST 2 PRACTICE	3521	19.6260	4.35317	.07336

T-test for Total Population of Students. To confirm the descriptive results noted in Table 5, the data were inferentially analyzed with a t-test for independent samples for the total population among the prompt groupings. These results are seen in Table 6.

The hypotheses were:

1. Null hypothesis: There are no differences in the WESTEST 2 Online Writing composite scores among the population of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.

2. Alternate hypothesis: There are differences in the WESTEST 2 Online Writing composite scores among the population of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.

Based on the t-test data in Table 6, no statistical significance was found for the WESTEST 2 Online Writing scores for both groups among the entire population. The significance of .067 ($p > .05$) did not indicate a difference in using one type of prompt over another with equal variances assumed. Thus, the results failed to reject the null hypothesis.

Table 6 RQ1 – WESTEST 2 Posttest Scores for WESTEST 2 and Writing Roadmap Prompts

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
WESTEST 2 2013 POSTTESTSCORES	Equal variances assumed	1.829	6457	.067	.19535	.10681	-.01404	.40473

Research Question Two

What differences exist among the five analytic writing traits on WESTEST 2 Online Writing for 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts? The total sample size of students was 380. The researcher used a random sampling technique in SPSS that selected 190 students who used the WESTEST 2 practice prompts and 190 students who used the Writing Roadmap prompts. To analyze multiple outcomes (five analytic traits), a multivariate analysis and a between subjects analysis were obtained, which included various descriptive data and inferential testing.

Descriptive Statistics. Table 7 shows the data for the sampled students who practiced writing using Writing Roadmap prompts and students who practiced writing using WESTEST 2 practice prompts. To determine the effect of multiple outcomes (five analytic traits), a multivariate analysis was obtained across gender and types of prompts. These initial results are shown in Table 7. A significant effect (p .000) is noted for gender which favored the female students. Although there was no overall significance found for the practice prompt variable, its p-level (.092) was below the .10 threshold and warranted further analysis.

Table 7 RQ2 – Multivariate Tests: Gender and Types of Prompts

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.957	28935.909[b]	5.000	6451.000	.000	.957
	Wilks' Lambda	.043	28935.909[b]	5.000	6451.000	.000	.957
	Hotelling's Trace	22.427	28935.909[b]	5.000	6451.000	.000	.957
	Roy's Largest Root	22.427	28935.909[b]	5.000	6451.000	.000	.957
Gender	Pillai's Trace	.053	71.692[b]	5.000	6451.000	.000	.053
	Wilks' Lambda	.947	71.692[b]	5.000	6451.000	.000	.053
	Hotelling's Trace	.056	71.692[b]	5.000	6451.000	.000	.053
	Roy's Largest Root	.056	71.692[b]	5.000	6451.000	.000	.053
Type of Prompt	Pillai's Trace	.001	1.895[b]	5.000	6451.000	.092	.001
	Wilks' Lambda	.999	1.895[b]	5.000	6451.000	.092	.001
	Hotelling's Trace	.001	1.895[b]	5.000	6451.000	.092	.001
	Roy's Largest Root	.001	1.895[b]	5.000	6451.000	.092	.001
Gender * Type of Prompt	Pillai's Trace	.002	2.185[b]	5.000	6451.000	.053	.002
	Wilks' Lambda	.998	2.185[b]	5.000	6451.000	.053	.002
	Hotelling's Trace	.002	2.185[b]	5.000	6451.000	.053	.002
	Roy's Largest Root	.002	2.185[b]	5.000	6451.000	.053	.002

a. Design: Intercept + Gender + Type of Prompt + Gender * Type of Prompt

b. Exact statistic

Between Subjects Analysis for Sampled Students. A between subjects analysis was obtained to further test for a potential effect on writing posttest scores for analytic traits and types of prompts to determine if there were effects on the WESTEST 2 Online Writing five analytic scores for the prompt groups. The hypotheses were:

1. Null hypothesis: There are no differences in the WESTEST 2 Online Writing five analytic trait scores among the samples of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.
2. Alternate hypothesis: There are differences in the WESTEST 2 Online Writing five analytic trait scores among the samples of 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.

Although in Table 7, there was no overall significance found for the practice prompt variable, its p-level (.092) was below the .10 threshold and warranted further analysis. This was made by examining the expanded output for between subjects and types of prompting for each of the five analytic traits. As can be seen in Table 8, significant differences occurred for posttest writing scores for *development* (p .018), *sentence structure* (p .048) and *organization* (p .008). In each case, the p-level indicated a less than 5% chance of error.

Table 8 RQ2 – Between Subjects Analysis for Analytic Traits and Types of Prompts*

Dependent Variable	Type III Sum of Squares	df	Mean Square	F-Ratio	Sig.
POSTTEST WRITING DEVELOPMENT SCORES	4.664	1	4.664	5.592	.018
POSTTEST WRITING SENTENCESTRUCTURE SCORES	3.140	1	3.140	3.928	.048
POSTTEST WRITING WORDCHOICE SCORES	1.819	1	1.819	2.271	.132
POSTTEST WRITING MECHANICS SCORES	2.866	1	2.866	3.546	.060
POSTTEST WRITING ORGANIZATION SCORES	5.335	1	5.335	6.959	.008

* Table was abridged to include only the between subjects output for the five analytic traits and types of prompts.

Research Question Three

What are the effects of gender and types of prompts on WESTEST 2 Online Writing scores among the population of 11th graders? The population consisted of female students who used Writing Roadmap (*n* = 1,431) and WESTEST 2 practice prompts (*n* = 1,784) and male students who used Writing Roadmap (*n* = 1,737) and WESTEST 2 practice prompts (*n* = 1,505). To determine such effects, descriptive statistics and a 2 x 2 Analysis of Variance were obtained.

Descriptive Statistics. Table 9 shows the mean scores and standard deviations of WESTTEST posttest scores for female and male students who used Writing Roadmap prompts and WESTEST 2 practice prompts. The data varied somewhat for these groupings. Females averaged about two points greater than males, while the males differed about .70 standard deviation points in variability. While gender showed some subtle differences, the types of prompts among them averaged a difference of .02 and .05 points.

Table 9 RQ3 – WESTEST 2 Mean Scores for Females and Males by Types of Prompt

WESTEST 2 2013 POSTTESTSCORES			
GENDER	TYPE OF WRITING PROMPTS USED FOR PRACTICE	Mean	Std. Deviation
FEMALES	WRITING ROADMAP	20.7346	3.70932
	WESTEST 2 PRACTICE	20.6766	3.65341
	Total	20.7024	3.67795
MALES	WRITING ROADMAP	18.9529	4.40912
	WESTEST 2 PRACTICE	18.5469	4.73457
	Total	18.7354	4.59007
Total	WRITING ROADMAP	19.8213	4.17840
	WESTEST 2 PRACTICE	19.6260	4.35317
	Total	19.7148	4.27534

2 x 2 ANOVA for Total Population of Students. To further analyze the descriptive results, a 2 X 2 Analysis of Variance was obtained on the total population to determine if significance could be found between the types of prompts and gender groups (Table 10). The hypotheses were:

1. Null hypothesis: There are no differences in the WESTEST 2 Online Writing composite scores between of the population of female and male students who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.
2. Alternate hypothesis: There are differences in the WESTEST 2 Online Writing composite scores between of the population of female and male students who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts.

The data in Table 10 shows significance for the gender variable, favoring the females, based on multiple comparisons testing. However, no significance was found for types of

prompts. As noted in Table 10 (p. 096), there was no interaction effect of the type of prompt (Writing Roadmap practice and WESTEST 2 practice) by gender. Thus, the alternate hypothesis was correct, and the null hypothesis was rejected.

Table 10 RQ3 – Tests of Between-Subjects Effects for Gender and Types of Prompt

Dependent Variable: WESTEST 2 2013 POSTTESTSCORES						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5722.135[a]	3	1907.378	111.036	.000	.054
Intercept	2296850.385	1	2296850.385	133708.381	.000	.958
Type of Prompt	49.258	1	49.258	2.867	.090	.000
Gender	5638.416	1	5638.416	328.234	.000	.053
Type of Prompt * Gender	47.570	1	47.570	2.769	.096	.000
Error	101092.874	5885	17.178			
Total	2403192.000	5889				
Corrected Total	106815.009	5888				

a. R Squared = .054 (Adjusted R Squared = .053)

SUMMARY OF FINDINGS

In conclusion, the findings reported in this chapter are based on the descriptive and inferential statistical analyses of the data collected from the writing score results for students using either Writing Roadmap practice prompts or WESTEST 2 practice prompts prior to completing the 2013 WESTEST 2 Online Writing. Research question one findings show no differences in the WESTEST 2 Online Writing composite scores among the 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts. Research question two findings show no significant differences in the WESTEST 2 Online Writing five analytic trait scores among the 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts. Research question

three findings show differences in the WESTEST 2 Online Writing composite scores between female and male students who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts. There is statistical significance when looking at the effect of gender on writing scores with females scoring greater than male students.

CHAPTER FIVE: SUMMARY, DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

SUMMARY INTRODUCTION

This study's purpose was to determine the effects of students practicing writing using practice prompts on the summative state writing assessment. Within the practice writing program WV Writes, there were two types of prompts that teachers could assign to their students, either Writing Roadmap 2.0 prompts or WESTEST 2 practice prompts. It was expected that there could be significant differences on the high stakes writing assessment scores of the students who wrote essays using Writing Roadmap 2.0 prompts compared to the students who wrote essays using the WESTEST 2 practice prompts.

In order to comply with W. Va. Code §§18-2E-1, 18-2E-1a, 18-2E-2, 18-2-5, 18-2E-8(c)(1), 18A-3-6 and Section 1111 of Public Law 107 – 110, No Child Left Behind Act of 2001, and the Elementary and Secondary Education Act, the state of West Virginia is required to develop a comprehensive assessment system to assess students based upon the adopted state standards. The content areas of English Language Arts/Literacy and Mathematics have come under close scrutiny with expectations to significantly raise achievement scores in these content areas. As students are being assessed on their writing abilities in a summative manner, the notion of having a formative assessment available for them to practice in preparation for the high stakes writing assessment has been a popular trend. WV Writes was the formative writing assessment program that the WVDE purchased for students in grades 3-11 to practice their writing skills prior to taking the WESTEST 2 Online Writing. Each year thousands of essays were written and scored in WV Writes. As so many students used WV Writes, the question emerged of whether or not a connection could be found between the scores students received in WV Writes and the

scores they received on WESTEST 2 Online Writing. Counties, schools, and teachers speculated that their students scored higher on WESTEST 2 Online Writing because they had extensively used the practice tool of WV Writes. In order to prepare for the new wave of assessments being created to assess the Common Core State Writing Standards, educators, states, and policy makers must determine if they will offer formative as well as diagnostic writing assessments in preparation for the summative writing test. It is important to determine whether or not the alignment between a formative assessment and its summative assessment will maximize student achievement and mastery on the summative test.

SUMMARY OF POPULATION AND SAMPLES

This study represented a population of 6,459 11th grade students enrolled in West Virginia public schools who used the WV Writes practice program prior to completing the WESTEST 2 Online Writing. The population was generated from an official database from the West Virginia Department of Education. From the study population of 6,459, there were 3,521 students who used WESTEST 2 practice prompts and 2,938 students who used Writing Roadmap prompts prior to taking the WESTEST 2 Online Writing. The researcher employed a random sampling technique from the SPSS system to obtain representative samples for each group with 190 subjects being sampled from the WESTEST 2 practice prompt condition and 190 subjects being sampled from the Writing Roadmap practice prompt condition, which resulted in a sample size of 380.

SUMMARY OF DESIGN AND INSTRUMENTS

The type of research employed for this study was a quantitative, two group, post hoc design, with random selection and assignment. There were two independent variables: practice prompts and gender. The factors of the prompt variable were WESTEST 2 practice prompts and

Writing Roadmap prompts. The dependent variable was the year-end, online, state writing assessment (WESTEST 2 Online Writing). In addition to the total composite writing proficiency scores, its five analytic trait scores of organization, development, sentence structure, word choice/grammar usage, and mechanics were analyzed for the population as a whole and for the samples, across gender and types of prompts. There were three major instruments employed. Two of these were the practice prompts (WESTEST 2 and Writing Roadmap) and the posttest assessment (WESTEST 2 Online Writing). Every year the WESTEST 2 Online Writing has been taken by West Virginia students, the reliability of its engine scoring has been validated through comparability studies conducted by the West Virginia Department of Education's Office of Assessment.

SUMMARY OF RESEARCH QUESTIONS, FINDINGS, AND DISCUSSION

This study was conducted to determine the effects of writing prompts in a practice writing program on students' writing proficiency among 11th graders. The effects of writing prompts on WESTEST 2 Online Writing composite scores and on the five analytic writing trait scores were examined using three research questions. A combination of descriptive and inferential techniques was employed to assess the statistical significance of the practice prompts and posttest writing assessment scores. These included a t-test for independent samples, a two-way analysis of variance (ANOVA), and a multivariable analysis. Descriptive data included mean scores, standard deviations, variance, and frequencies.

Research Question One

What are the effects on WESTEST 2 Online Writing composite scores among 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?

Results for research question one found no statistical significance when comparing writing proficiency scores on WESTEST 2 Online Writing composite scores for those who used either WESTEST 2 practice writing or Writing Roadmap prompts.

Research Question Two

What differences exist among the five analytic writing traits on WESTEST 2 Online Writing for 11th graders who used Writing Roadmap prompts compared to their peers who used WESTEST 2 practice prompts?

When analyzing all five analytic traits, these mean scores were similar and no major differences in variance were found. Likewise, based on the t-test results, no statistical significance was found on the WESTEST 2 Online Writing for either group in the sampled population. There was a range of probability outcomes which did indicate a difference in the five analytic traits when using one type of prompt over another.

The findings for research questions one and two did not align with existing studies which concluded that the types of prompts used to assess writing impacted students as they addressed the prompt (Condon, 2004; Nuckles, Huber, & Renkl, 2009). Condon (2004) concluded that practice writing prompts that required students to read a passage and then answer a prompt could place students with reading disabilities at a disadvantage. The WESTEST 2 practice prompts were structured using a passage and prompt; however, the similarity in the year-end writing composite scores would indicate that passage and prompt length were not causing significant differences. Nuckles, Huber, and Renkl (2009) found that student proficiency was increased when writing prompts activated critical thinking skills. The fact that the more robust WESTEST 2 practice prompts did not cause student proficiency to raise on the WESTEST 2 Online Writing any greater than the simplistically formatted Writing Roadmap prompts contradicts their study.

However, this study did find similar results to previous studies (Barry & Nielsen, 1997; Beck & Jeffery, 2007; Kobrin, Deng, & Shaw, 2011; Lee, 2008) which all reported that the types of writing prompts used to assess writing showed no effect on student proficiency. Barry and Nielsen (1997) determined that choosing a writing prompt versus being assigned a prompt did not impact writing proficiency scores of students which confirms the finding of this study that practice prompt type did not affect achievement scores. Beck and Jeffery (2007) concluded that high stakes writing assessment scores were impacted by the scoring process conducted by human scorers, not the types of writing prompts being used in the assessment. This study aligns with Beck and Jeffery (2007) in that the findings are the same related to writing prompt types not impacting writing scores. Kobrin, Deng, and Shaw (2011) determined that the writing prompt type did not affect the quality of student writing and this study aligns to their determination in that there was no significance in scores between the two types of practice writing prompts. Lee (2008) found that students who perceived a preference for one prompt over another believed they scored better when in reality the type of writing prompt did not affect the scoring whether a student liked the writing prompt or not. This study also accords with Lee (2008) because the perception was that one type of prompt, WESTEST 2, was preferential over the other; however, no statistical evidence supported this belief.

Research Question Three

What are the effects of gender and types of prompts on WESTEST 2 Online Writing scores among 11th graders?

The entire student population was the basis for research question three. It consisted of female (n=3,216) and male students (n=3,243) who used Writing Roadmap and WESTEST 2 practice prompts. To determine such effects, descriptive statistics and a 2 x 2 analysis of variance

(ANOVA) were obtained. Based on the 2 x 2 ANOVA, statistical significance was found for gender, with females scoring greater than males. However, there was no interaction effect between gender and the types of prompts (WESTEST 2 practice and Writing Roadmap) on the WESTEST 2 Online Writing posttest scores.

A finding of gender causing an effect on writing proficiency was also noted in studies conducted by other researchers. Gabrielson, Gordon, and Engelhard (1995) wanted to see what effects, if any, would occur when grade 11 students were allowed to choose the writing prompt. What they discovered was that the gender variable had more of an effect on writing scores than the ability to select a prompt and compose to it. This study aligns with Gabrielson, Gordon, and Engelhard (1995) because regardless of which prompt type, female students still scored greater than did males. Knudson (2001) showed that gender had positive correlations with writing proficiency. Female students scored higher than males on an achievement writing test which agrees with this study's findings.

IMPLICATIONS

When providing practice writing opportunities for students to increase their proficiency scores on high stakes writing assessments, this study showed that the types of prompts students accessed did not impact their writing proficiency scores. Counties, schools, and teachers may have believed that their students who used WESTEST 2 practice prompts were better prepared and more proficient on WESTEST 2 Online Writing; however, in reality, providing students with practice opportunities using either Writing Roadmap prompts or WESTEST 2 practice prompts created no significant difference. As the Common Core writing standards are implemented by states and high stakes assessments are created to determine student writing proficiency, this study provided evidence that students who had a generic writing prompt versus a mirror image of

the high stakes writing assessment were no better or worse than the other. The new assessment vendors, states, counties, schools, and teachers will all benefit from these study findings as the new assessment systems are adopted based on Common Core writing standards across the nation.

Additionally, when considering the gender of students and their writing abilities, this study supported the findings reported in other studies which all pointed out that gender is an influential variable on writing proficiency scores. Such results can provide important information and create awareness among education stakeholders, such as states, counties, schools, teachers, and parents about the need to support efforts to consistently monitor and enhance writing abilities. This is particularly true for teachers who can provide instructional resources and support to ensure that male students are able to increase and improve their writing achievement and related proficiencies.

Related implications are that teachers and students can use any type of practice writing prompt to prepare for a summative writing assessment. These practice writing prompts need not be clones of the writing prompts that will appear on high stakes writing assessments. Additionally, when the composite writing scores were broken into the five analytic trait scores, the scores on these were similar for all students regardless of the types of practice prompts used.

RECOMMENDATIONS FOR FUTURE STUDIES

A number of opportunities for future studies emerged from this study and its conclusions about practice writing prompts and writing proficiency.

1. While the type of practice writing prompt did not impact the writing proficiency, it remains unknown whether more frequent use of practice prompts increases writing proficiency.

2. The issue of frequency of usage of the practice writing prompts prior to the high stakes writing assessment could be considered. Did students practice consistently throughout the year or did they engage the practice program in a period just before the WESTEST 2 Online Writing?
3. Because WV Writes has been available since 2009, a longitudinal, trend study could be conducted for the academic years from 2009-2013.
4. Teachers used the WV Writes practice program prior to students taking WESTEST 2 Online Writing; however, the methods used by teachers to teach writing skills were not examined as a variable.
5. This study did not examine differences that may have existed for students who did not use a practice program. Thus, the lack of program use could be studied or compared to determine its effect.
6. There have been no surveys or collections of teacher and student perceptions about the WV Writes program. There may be an interest in perceptions held by the teachers and students who engaged in practice writing programs, e.g., what they thought were the most valuable and least effective aspects, in regard to the year-end writing assessment.
7. While the current study examined a population of 11^{th} graders, it may be informative to compare practice prompts at lower grade levels, e.g., at upper elementary and middle school. Language Arts instruction is more specifically taught at these levels and the effects of practice programs may be more pronounced.

REFERENCES

A Chronicle of West Virginia's Global21 Initiative. (2009). Retrieved from http://wvde.state.wv.us/oaa/pdf/WV%2021st%20Century%20Learning%20Chronicle_FINAL.pdf

Applebee, A. (1996). Writing and reasoning. *Review of Educational Research, 54*(4), 577-596. doi:10.3102/00346543054004577

Attali, Y. (2004). Exploring the feedback and revision features of Criterion. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), San Diego, CA.

Attali, Y. (2007). *Construct validity of e-rater in scoring TOEFL essays* (Report No. RR-07-21). Princeton, NJ: Educational Testing Service. Retrieved from www.ets.org/Media/Research/pdf/RR-07-21.pdf

Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a generic approach in automated essay scoring. *Journal of Technology, Learning & Assessment, 10*. Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1603/1455

Attali, Y., Lewis, W., & Steier, M. (2013). Scoring with the computer: Alternative procedures for improving the reliability of holistic essay scoring. *Language Testing, 30*(1), 125-141. doi:10.1177/0265532212452396

Barrett, T. (1994). Generalizability of writing tasks at fourth grade in the Riverside Unified School District. Paper presented at the annual meeting of California Educational Research Association, San Diego, CA. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED377498

Bejar, I. I. (2011). A validity-based approach to quality control and assurance of automated scoring. *Assessment in Education: Principles, Policy & Practice, 18*(3), 319–341. Retrieved from http://dx.doi.org/10.1080/0969594X.2011.555329

Breland, H. (1983). The direct assessment of writing skills: A measurement review (Report No. 83-6). New York, NY: College Entrance Examination Board.

Bridgeman, B., Trapani, C., & Bivens-Tatum, J. (2011). Comparability of essay question variants. *Assessing Writing, 16*(4), 237-255. Retrieved from http://dx.doi.org/10.1016/j.asw.2011.06.002

Bridgeman, B., Morgan, R., & Wang, M. (1997). Choice among essay topic: Impact on performance and validity. *Journal of Educational Measurement, 34*(3), 273-286. doi:10.1111/j.1745-3984.1997.tb00519.x

Bridwell, L. (1980). Revising strategies in twelfth-grade students' transactional writing. *Research in the Teaching of English, 14*, 197-222. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ236505

Brimi, H. (2012). Teaching writing in the shadow of standardized writing assessment: An exploratory study. *American Secondary Education*, *41*(1), 52-77. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=f5h&AN=84030212&site=ehost-live

Brown, R. (1986). Testing black student writers. In K.L. Greenberg, H.S. Wiener, & R.A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 98-108). New York: Longman.

Burstein, J. (2003). The e-rater scoring engine: Automated essay scoring with natural language processing. In M.D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp.113-121). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Charman, D. (1999). Issues and impacts of using computer-based assessments CBAs for formative assessment. In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 85-93). London: Kogan Page Limited.

Charman, D., & Elmes, A. (1998). A computer-based formative assessment strategy for a basic statistics module in geography. *Journal of Geography in Higher Education, 22*(3), 381-385. Retrieved from http://dx.doi.org/10.1080/03098269885787

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18*, 65-81. Retrieved from http://www.jstor.org/stable/40170979

Clariana, R. B. (1993). A review of multiple-try feedback in traditional and computer based instruction. *Journal of Computer Based Instruction, 20*(3), 74-76.

Coe, M. (2000). *Direct writing assessment in action: Correspondence of six-trait writing assessment scores and performance on an analog to the Washington Assessment of Student Learning writing test*. Portland, OR: Northwest Regional Educational Laboratory.

Coffman, W. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 271-302). Washington, DC: American Council on Education.

Condon, W. (2004). Looking beyond judging and ranking: Writing assessment as a generative practice. *Assessing Writing, 14*(3), 141-156. Retrieved from http://dx.doi.org/10.1016/j.asw.2009.09.004

Condon, W. (2013). Large-scale assessment, locally-developed measures, and automated scoring of essays: Fishing for red herrings? *Assessing Writing, 18*(1), 100-108. Retrieved from http://dx.doi.org/10.1016/j.asw.2012.11.001

DelleBovi, B. M. (2012). Literacy instruction: From assignment to assessment. *Assessing Writing, 17*(4), 271-292. Retrieved from http://dx.doi.org/10.1016/j.asw.2012.07.001

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment, 5*(1) 1-35.

Dryer, D. (2013). Scaling writing ability: A corpus-driven inquiry. *Written Communication, 30*(1), 3-35. Retrieved from http://dx.doi.org/10.1177/0741088312466992

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing, 14*(2), 88-115. Retrieved from http://dx.doi.org/10.1016/j.asw.2009.04.001

Elliot, S. (2001). Intellimetric: From here to validity. In M. Shermis & J. Burstein (Eds.), *Automated essay scoring; A cross disciplinary perspective* (pp. 67-81). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Elliot, S. (2003). IntelliMetric: From here to validity. In J.C. Burstein (Ed.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Elliot, S., & Mikulas, C. (2004). The impact of MY Access! use on student writing performance: A technology overview and four studies. Paper presented at the annual meeting of the American educational Research Association, San Diego, CA.

Gabrielson, S., Gordon, B., & Engelhard Jr., G. (1995). The effects of task choice on the quality of writing obtained in a statewide assessment. *Applied Measurement in Education, 8*(4), 273-291. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ517180

Graham, S. (1997). Executive control in the revising of students with learning and writing difficulties. *Journal of Educational Psychology, 89*(2), 223-234. Retrieved from http://dx.doi.org/10.1037//0022-0663.89.2.223

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 15*(1), 75-85. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ242214

Haberman, S. J., & Sinharay, S. (2010). The application of the cumulative logistic regression model to automated essay scoring. *Journal of Educational & Behavioral Statistics, 35*, 586–602. Retrieved from http://dx.doi.org/10.3102/1076998610375839

Harrington, H., & Rich, C. (2006). *2006 West Virginia writing assessment comparison study*. West Virginia Department of Education website. Retrieved from http://wvde.state.wv.us/oaa/pdf/2006%20WV%20Writing%20Asssessment%20Comparision%20Study.pdf

Haswell, R. (2006). The complexities of responding to student writing; or, looking for shortcuts via the road of excess. *Across the Disciplines, 3*. Retrieved from http://wac.colostate.edu/atd/articles/haswell2006.cfm

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Hidi, S., & Boscolo, P. (2006). Motivation and writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144-157). New York: Guilford.

Higgins, B., Miller, M., & Wegmann, S. (2007). Teaching to the test…not! Balancing best practice and testing requirements in writing. *Reading Teacher, 60*(4), 310-319. doi:10.1598/RT.60.4.1

Hinkel, E. (2002). *Second language writers' test: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60*(2), 237-263. Retrieved from http://dx.doi.org/10.3102/00346543060002237

James, C. (2006). Validating a computerized scoring system for assessing writing and placing students in composition courses. *Assessing Writing*, *11*(3), 167-178. Retrieved from http://dx.doi.org/10.1016/j.asw.2007.01.002

James, C. (2008). Electronic scoring of essays: Does topic matter? *Assessing Writing*, *13*(2), 80-92. Retrieved from http://dx.doi.org/10.1016/j.asw.2008.05.001

Javed, M., Juan, W., & Nazli, S. (2013). A study of students' assessment in writing skills of the English language. *International Journal of Instruction*, *6*(2), 129-144. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=90564975&site=ehost-live

Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-taker's choice: An investigation of the effects of topic on language test performance. *Language Testing*, *16*(4), 226-256. doi: 10.1177/026553229901600402

Keith, T. Z. (2003). Validity and automated essay scoring systems. In M.D. Shermis & J.C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 147-167). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Klobucar, A., Elliot, N., Deess, P., Rudniy, O., & Joshi, K. (2013). Automated scoring in context: Rapid assessment for placed students. *Assessing Writing*, *18*(1), 62-84. Retrieved from http://dx.doi.org/10.1016/j.asw.2012.10.001

Knoch, U. (2007). Little coherence, considerable strain for reader: A comparison between two rating scales for the assessment of coherence. *Assessing Writing*, *12*(2), 108-128. Retrieved from http://dx.doi.org/10.1016/j.asw.2007.07.002

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, *16*(2), 81-96. Retrieved from http://dx.doi.org/10.1016/j.asw.2011.02.003

Knudson, R. E. (2001). Writing experiences, attitudes, and achievement of first to sixth graders. *The Journal of Educational Research*, *89*(2). Retrieved from http://dx.doi.org/10.1080/00220671.1995.9941199

Kobrin, J., Deng, H., & Shaw, E. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, *16*(3), 154-169. doi:10.1016/j.asw.2011.01.001

Kukich, K. (2000). Beyond automated essay scoring. *IEEE Intelligent Systems*, *15*(5), 22-27.

Landauer, T. K., Laham, D., & Foltz, P. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems*, *15*(5), 27-31.

Landauer, T. K., Lochbaum, K. E., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice*, *48*(1), 44-52.

Lee, H. (2008). The relationship between writers' perceptions and their performance on a field-specific writing test. *Assessing Writing*, *13*(2), 93-110. doi:10.1016/j.asw.2008.08.002

Lee, Y., Kantor, R., & Mollaun, P. (2002). Score reliability as an essential prerequisite for validating new writing and speaking tasks for TOEFL. Paper presented at the annual meeting of Teachers of English to the Speakers of Other Languages (TESOL). Salt Lake City, UT. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED464964

Llosa, L., Beck, S. W., & Zhao, C.G. (2011). An investigation of academic writing in secondary schools to inform the development of diagnostic classroom assessments. *Assessing Writing*, *16*(4), 256-273. Retrieved from http://dx.doi.org/10.1016/j.asw.2011.07.001

Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal*, *103*(1), 3-25. Retrieved from http://dx.doi.org/10.1086/499713

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, *15*(2), 118-129. Retrieved from http://dx.doi.org/10.1016/j.asw.2010.04.002

Mills, C. N., & Jaeger, R. J. (1998). Creating descriptions of desired student achievement when setting performance standards. In L. Hansche (Ed.), *Handbook for the development of performance standards* (pp. 73-85). Washington, DC: US Department of Education and the Council of Chief State School Officers.

Moon, T. R., Loyd, B. H., & Hughes, K. R. (1996). Generalizability analyses of a large-scale writing assessment. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. Retrieved from http://dx.doi.org/10.1080/03075070600572090

Nuckles, M., Hubner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning & Instruction*, *19*(3), 259-271. Retrieved from http://dx.doi.org/10.1016/j.learninstruc.2008.05.002

Olinghouse, N. G., Zheng, J., & Morlock, L. (2012). State writing assessment: Inclusion of motivational factors in writing tasks. R*eading & Writing Quarterly*, *28*(1), 97-119. doi:10.1080/10573569.2012.632736

Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology*, *33*(5), 515-523. Retrieved from http://dx.doi.org/10.1111/1467-8535.00288

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111-129. Retrieved from http://dx.doi.org/10.1016/j.asw.2008.07.001

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, *17*(1), 18-34. Retrieved from http://dx.doi.org/10.1016/j.asw.2011.09.002

Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002). Comparing the validity of automated and human scoring of essays. *Journal of*

Educational Computing Research, *26*(4), 407-425. Retrieved from http://dx.doi.org/10.1092/UP3H-M3TE-Q290-QJ2T

Ramineni, C. (2013). Validating automated essay scoring for online writing placement. *Assessing Writing*, *18*(1), 40-61. Retrieved from http://dx.doi.org/10.1016/j.asw.2012.10.005

Sacchetti, M. (2005, March 9). For SAT takers a new test; Essay section may be struggle for students. *Boston Globe*. Retrieved from http://www.boston.com/news/education/k_12/articles/2005/03/09/for_sat_takers_a_new_test/

Sakyi, A. A. (2001). Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. In A.J. Kunnan (Ed.), *19th language testing research colloquium. Fairness and validation in language assessment* (pp. 131-152). Cambridge, UK: Cambridge University Press.

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing*, *22*(1), 1-30. doi:10.1191/0265532205lt295oa

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*(3), 215-232. Retrieved from http://dx.doi.org/10.1111/j.1745-3984.1993.tb00424.x

Shermis, M. (2003). Facing off on automated scoring. *Assessment Update*, *15*(2), 4-6.

Shermis, M. D., Mzumara, H. R., Olson, J., & Harrington, S. (2001). On-line grading of student essays: PEG goes on the World Wide Web. *Assessment & Evaluation in Higher Education*, *26*(3), 247-259. Retrieved from http://dx.doi.org/10.1080/02602930120052404

Spandel, V. (2006). Speaking my mind: In defense of rubrics. *English Journal*, *96*(1), 19-22. Retrieved from http://dx.doi.org/10.2307/30046656

Stewart, M., & Grobe, C. (1979) Syntactic maturity, mechanics of writing and teachers' quality ratings. *Research in the Teaching of English*, *13*(3), 207-215. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ210748

Teach21 On-line CSO Level Resources. (n.d.). Retrieved from http://wveis.k12.wv.us/Teach21/public/cso/cso.cfm?tsele1=1&tsele2=999

VandenBergh, H., DeGlopper, K., & Schoonen, R. (1988). Direct measures of writing ability: Validity and task effects. In F.H. VanEemeren & R. Grootendorst (Eds.), *Language Performance* (pp. 370-378). Dordrecht, Germany: Foris Publications.

WESTEST 2 Online Writing Overview. (n.d.). Retrieved from http://wvde.state.wv.us/oaa/writing/wa_index.html

Weigle, S. (2002). *Assessing writing*. New York: Cambridge University Press.

White L., Hixon, N., & D'Brot, J. (2010). Research Brief: Impact of Writing Roadmap 2.0 on WESTEST 2 Online Writing assessment scores. West Virginia Department of Education website. Retrieved from http://wvde.state.wv.us/oaa/pdf/research/Research%20Brief%20-%20WRM2.0%20Impact%20FINAL%2001.27.10.pdf

Writing Roadmap 2.0 Student Interface. (2009). Retrieved from http://writingroadmap2.com

WV Writes Student Interface. (2012). Retrieved from http://student.wvwrites.com

WV Writes. (n.d.). Retrieved from http://wvde.state.wv.us/oaa/wvwrites/wvwrites.html

Zinn, A. (1998). Ideas in practice: Assessing writing in the developmental classroom. *Journal of Developmental Education, 2*(2), 28-34. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ584348

APPENDIX A: WEST VIRGINIA GRADE 11 WRITING RUBRIC

West Virginia Writing Rubric - Grade 11

	ORGANIZATION	DEVELOPMENT	SENTENCE STRUCTURE	WORD CHOICE/ GRAMMAR USAGE	MECHANICS
Score Of 6	**Exemplary Organization** • Clear and logical progression of ideas • Strong beginning, middle and end • Strong introductory and concluding paragraph; related and cohesive supporting paragraphs • Strong use of transition	**Exemplary Development** • Topic sentence • Strong use of examples, evidence or relevant details • Clear focus maintained for intended audience ➢ Descriptive: exemplary examples and sensory details ➢ Narrative: exemplary ideas convey the experience ➢ Informative: exemplary explanation ➢ Persuasive: convincing argument	**Exemplary Sentence Structure** • Complete and correct • Strong evidence of sentence variety: (types/length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple and compound	**Exemplary Word Choice/Grammar Usage** • Vivid, specific, precise • Consistent grammar usage ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Exemplary Mechanics** • May have minor errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Needs little or no editing
Score of 5	**Effective Organization** • Logical progression of ideas • Clear beginning, middle and end • Introductory paragraph, supporting paragraphs and concluding paragraph • Sufficient use of transition	**Effective Development** • Topic sentence • Clear use of examples, evidence or relevant details • Clear focus ➢ Descriptive: effective examples and sensory details ➢ Narrative: effective ideas convey the experience ➢ Informative: effective explanation ➢ Persuasive: convincing argument	**Effective Sentence Structure** • Complete and correct • Clear evidence of sentence variety: (types/length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple and compound	**Effective Word Choice/Grammar Usage** • Specific, precise • Mostly consistent grammar usage ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Effective Mechanics** • Few errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Needs some editing
Score of 4	**Adequate Organization** • Some evidence of a logical progression of ideas • Beginning, middle and end • Introductory paragraph, supporting paragraphs and concluding paragraph • Some use of transition	**Adequate Development** • Topic sentence • Sufficient use of examples, evidence or relevant details • Some evidence of a continued focus ➢ Descriptive: adequate examples and sensory details ➢ Narrative: adequate ideas convey the experience ➢ Informative: adequate explanation ➢ Persuasive: convincing argument	**Adequate Sentence Structure** • Minor errors in structure • Some evidence of sentence variety: (types/length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple and compound	**Adequate Word Choice/Grammar Usage** • Appropriate, somewhat simplistic • Somewhat consistent ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Adequate Mechanics** • Some errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Needs editing but doesn't Impede readability
Score of 3	**Limited Organization** • Limited evidence of a logical progression of ideas • Beginning, middle and/or end • Introductory paragraph, concluding paragraph and limited supporting paragraphs • Limited use of transition	**Limited Development** • Attempted topic sentence • Some use of examples, evidence or supporting details • Some evidence of focus ➢ Descriptive: limited examples and sensory details ➢ Narrative: limited ideas convey the experience ➢ Informative: limited explanation ➢ Persuasive: limited evidence to support argument	**Limited Sentence Structure** • Some errors in sentence structure • Limited evidence of sentence variety: (types/ length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple sentences; limited use of compound	**Limited Word Choice/Grammar Usage** • Inadequate, repetitive, simplistic • Several inconsistencies in grammar usage ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Limited Mechanics** • Frequent errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Begins to impede readability
Score of 2	**Minimal Organization/Minimal Response** • Lacks evidence of a logical progression of ideas • Lacks a beginning, middle and/or end • Lacks introductory paragraph, supporting paragraphs and/or concluding paragraph • Lacks clear transition	**Minimal Development/Minimal Response** • Poorly stated topic sentence • Lacks sufficient examples, evidence or supporting details • Unclear focus ➢ Descriptive: minimal examples and sensory details ➢ Narrative: minimal ideas convey the experience ➢ Informative: minimal explanation ➢ Persuasive: lacks evidence to support argument	**Minimal Sentence Structure/Minimal Response** • Contains fragments and/or run-ons • Minimal evidence of sentence variety: (types/ length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple sentences; minimal use of compound	**Minimal Word Choice/Grammar Usage/Minimal Response** • Inadequate, incorrect • Frequent inconsistencies in grammar usage ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Minimal Mechanics/Minimal Response** • Consistent errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Impedes readability
Score of 1	**Inadequate Organization** • Little or no progression of ideas • No evidence of beginning, middle and/or end • Inadequate paragraphing • No transition	**Inadequate Development** • Lacks topic sentence • Lacks examples, evidence or supporting details • No focus ➢ Descriptive: inadequate examples and sensory details ➢ Narrative: inadequate ideas convey the experience ➢ Informative: inadequate explanation ➢ Persuasive: no argument	**Inadequate Sentence Structure** • Contains numerous fragments and/or run-ons • Little or no evidence of sentence variety: (types/length) ➢ Declarative, interrogative, exclamatory, imperative ➢ Simple sentences; minimal or no use of compound	**Inadequate Word Choice/Grammar Usage** • Rambling, inappropriate, incorrect • Distracting inconsistencies in grammar usage ➢ Subject/verb agreement ➢ Singular/plural nouns ➢ Verb (tense and usage) ➢ Pronoun usage ➢ Adjective/Adverb	**Inadequate Mechanics** • Serious and consistent errors ➢ Punctuation ➢ Capitalization ➢ Spelling • Impedes communication/understanding

APPENDIX B: INSTITUTIONAL REVIEW BOARD LETTER

MARSHALL UNIVERSITY
www.marshall.edu

Office of Research Integrity
Institutional Review Board
401 11th St., Suite 1300
Huntington, WV 25701

FWA 00002704

IRB1 #00002205
IRB2 #00003206

June 3, 2014

Lisa Heaton, PhD
Curriculum and Instruction, COEPD

RE: IRBNet ID# 605194-1
At: Marshall University Institutional Review Board #2 (Social/Behavioral)

Dear Dr. Heaton:

Protocol Title:	[605194-1] WV Writes and WESTEST2 Online Writing	
Expiration Date:	June 3, 2015	
Site Location:	MUGC	
Submission Type:	New Project	APPROVED
Review Type:	Exempt Review	

In accordance with 45CFR46.101(b)(4), the above study was granted Exempted approval today by the Marshall University Institutional Review Board #2 (Social/Behavioral) Designee for the period of 12 months. The approval will expire June 3, 2015. A continuing review request for this study must be submitted no later than 30 days prior to the expiration date.

This study is for student Stacey Murrell.

If you have any questions, please contact the Marshall University Institutional Review Board #2 (Social/Behavioral) Coordinator Bruce Day, ThD, CIP at 304-696-4303 or day50@marshall.edu. Please include your study title and reference number in all correspondence with this office.

AUTHOR'S VITA

Stacey Murrell PO Box 551, Milton, WV 25541
slmurrell@k12.wv.us

ACADEMIC DEGREES

Marshall University, Huntington, WV

English Literature	B.A.	1997
English Literature	M.A.	2000
English Education, 5-12	M.A.T.	2002

PROFESSIONAL EXPERIENCE

English Language Arts (ELA) Acuity and WV Writes Coordinator Aug. 2010- present
Office of Assessment and Accountability
West Virginia Department of Education, Charleston, WV

- Coordinate the ELA Acuity and WV Writes formative assessment programs for WV public schools, teachers, and students.
- Plan, organize, write, manage, and implement passage and item development that aligns to the WV NxG CSOs/CCSS CSOs.
- Hire and train teacher item writers and performance task writers to create items for WV NxG CSOs assessments and WV Writes writing prompts.
- Provided customer support for Acuity and WV Writes users, such as county superintendents, county test coordinators, principals, teachers, and parents.
- Travel extensively to RESAs/counties/schools to provide professional development for educators and administrators on assessment and accountability, such as using technology, program adoption and implementation, and data analysis.
- Present updates, data, and information at Office of Assessment County Test Coordinator meetings, WESTEST 2 Planning and Review meetings, Acuity and WV Writes Planning and Review meetings, Office of Assessment meetings, and vendor meetings.
- Prepare state-level memos, data, and documentation for dissemination to counties, schools, and teachers.
- Communicate state-wide assessment information using listservs.
- Assist with customer support during WESTEST 2 Online Writing.
- Monitor WESTEST 2 test administration and assist with WESTEST 2 administration.
- Represent WV as an assessment coordinator at national assessment meetings, such as NAEP item reviews and as a member of CCSSO's FAST SCASS.

High School English Teacher June 2007- Aug. 2010
Huntington High School, Huntington, WV

- Taught AP Language, Honors English 10, and 10th grade courses as a full-time teacher.

- Developed lesson plans, curriculum, and activities in accordance with the West Virginia Department of Education's English Language Arts Curriculum Standards and Objectives.
- Upheld Cabell County School District's learning, discipline, and grading

High School English Teacher Aug. 2005- June 2007

Gahanna Lincoln High School, Gahanna, OH

- Taught 10th, 11th, and 12th grade courses as a full-time teacher.
- Developed lesson plans, curriculum, and activities in accordance with the Ohio Department of Education's English Standards of Learning.
- Upheld Gahanna Jefferson School District's learning, discipline, and grading standards.

High School English Teacher Aug. 2004- June 2005

Horizon Science Academy High School, Columbus, OH

- Taught 8th grade, 10th grade, 11th grade, and Creative Writing courses as a full-time teacher.
- Tutored academically at-risk students at "Saturday School" for the 10th grade OGT Writing and Reading tests.
- Sponsored and co-sponsored student clubs and activities.

High School English Teacher Aug. 2003- June 2004

Woodbridge Senior High School, Woodbridge, VA

- Taught 10th grade English courses and Center for the Fine and Performing Arts Advanced English 10 Humanities courses as a full-time teacher.
- Developed lesson plans, curriculum, and activities in accordance with the Virginia Department of Education's 10th grade English Standards of Learning.
- Upheld the Prince William County learning, discipline, and grading standards.

Part Time English Instructor Aug. 2001- Dec. 2002

Marshall University, Huntington, WV

- Taught English composition courses.
- Created coursework, syllabus, and schedule in a computer-based classroom.
- Responsible for upholding University teaching and grading policies.

CERTIFICATION/LICENSE

- WV permanent teaching certification, grades 5-12 English Language Arts

PUBLICATIONS AND PAPERS

O'Byrne, B., & Murrell, S. L. (2012). Adolescent blogging practices and the new literacies. Manuscript submitted for publication, *Journal of Computer Assisted Learning*, Charleston.

O'Byrne, B., Murrell, S., & Bailey, D. (2011). Literacy in multimedia environments: Preliminary findings. In *Proceedings of Society for Information Technology & Teacher Education International Conference 2011* (pp. 1600-1606). Chesapeake, VA: AACE.

O'Byrne, B., Murrell, S.L., & Bailey, D. (2010). Impact of web-authoring tools on literacy and learning in the K-8 classroom. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2010* (pp. 2483-2488). Chesapeake, VA: AACE.