

EXPLORING TASK AND GENRE DEMANDS IN THE PROMPTS AND RUBRICS OF
STATE WRITING ASSESSMENTS AND THE NATIONAL ASSESSMENT OF
EDUCATIONAL PROGRESS (NAEP)

By

Ya Mo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Curriculum, Teaching and Educational Policy—Doctor of Philosophy
Measurement and Quantitative Methods—Doctor of Philosophy

2014

ABSTRACT

EXPLORING TASK AND GENRE DEMANDS IN THE PROMPTS AND RUBRICS OF STATE WRITING ASSESSMENTS AND THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

By

Ya Mo

My dissertation research examines constructs of writing proficiencies in state and national assessments through content analysis of writing prompts and rubrics; predicts students' writing performance on the National Assessment of Educational Progress (NAEP) from assessment variations using multi-level modeling; and explores genre demands in state writing assessments through syntactic analysis of writing prompts to identify the ambiguity and implicit expectations and content analysis of rubrics and state standards to identify the genres specified.

Through content analysis of 78 prompts and 35 rubrics from 27 states' writing assessments, and three representative prompts and rubrics from the NAEP, the research presented in Chapter 1 finds that state writing assessments and the NAEP seem to align in their adoption of the writing process approach, their attention to audience and students' topical knowledge, their accommodations through procedure facilitators, and their inclusion of organization, structure, content, details, sentence fluency, and semantic aspects as well as general conventions, such as punctuation, spelling, and grammar in their assessment criteria. However, the NAEP's writing assessment differs from many states' by having explicit directions for students to review their writing, giving students two timed writing tasks, making informative composition—which was rarely included in state assessments—one of the three genres assessed, and including genre-specific components in their writing rubrics. The fact that all of the NAEP's

writing rubrics are genre-mastery rubrics with genre-specific components can be considered one of its biggest differences from most state writing assessments.

To examine the impact of the variations between state and national writing assessments through Hierarchical Linear Modeling, the research presented in Chapter 2 examines the relationship between students' NAEP performances and the amount of difference between state and NAEP direct writing assessments using content analysis of the state and NAEP prompts and rubrics detailed above. This study finds that students' preparedness for the tasks, namely the similarity between the assessments of their home states and the NAEP, plays a role in students' performance on the NAEP. Students from those states with writing assessments similar to the NAEP performed significantly better than students from states with writing assessments that differed markedly from the NAEP.

Through syntactic analysis of the same set of state prompts and content analysis of rubrics and standards, the research presented in Chapter 3 explores genre demands in state writing assessments. In total, this study found that 23% of prompts possessed one of two problematic features: 14% of prompts were ambiguous, and 9% of prompts had implicit genre expectations. Almost one third of those prompts that possessed problematic features were used with genre-mastery rubrics. The content analysis of state writing standards also suggests that 22% of them do not cover all the genres assessed in their corresponding writing assessments. The ambiguity and implicit genre expectations in writing prompts and the limited congruence of state writing assessments with learning expectations pose potential threats to the valid interpretation and use of these writing assessments.

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Professor Gary Troia. He inspired my interests in writing assessments, guided me through writing research, made the IES-funded K-12 Writing Alignment Project data available for my dissertation research, always gave me prompt feedback, and offered me his support along every step of my doctoral study. I look up to and learn from his productivity, diligence, and vision as a scholar.

I am also indebted to my co-advisor, Professor Mark Reckase. Being an outstanding teacher, he introduced me to measurement theories and sparked my interest in assessments. His devotion and passion towards the field of measurement is always inspirational to me.

I am very grateful to my other dissertation committee members—Professor Susan Florio-Ruane and Professor Peter Youngs. They have always been extremely helpful, giving me all the support that I need and sharing with me insights that helped develop my dissertation.

Finally, I extend my heartfelt thanks to my family and dear friends. They gave me their unconditional love and support, which motivated me through every step of my academic pursuits.

This dissertation study uses a portion of data collected and coded in the K-12 Writing Alignment Project, funded by grant number R305A100040 from the U.S. Department of Education, Institute of Education Sciences, to Michigan State University. Statements do not necessarily reflect the positions or policies of this agency, and no official endorsement by it should be inferred.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
INTRODUCTION	1
CHAPTER 1: Examining Writing Constructs in U.S. State and National Assessments	5
1. Introduction	5
2. Review of Literature	7
Genre Theories in Composition	10
3. Research Questions	17
4. Mode of Inquiry	17
4.1 State and NAEP Direct Writing Assessments	17
4.2 Coding Taxonomy	19
4.3 Procedure	20
5. Results	21
5.1 How do the features of writing tasks and rubrics vary across a sample of states and NAEP?	21
Writing Process	22
Writing Context	23
Writing Components	24
Writing Mechanics	24
Writing Knowledge	25
5.2 What are the connections between these prompts and rubrics, especially in terms of their genre demands?	25
Prompts	25
Rubrics	27
Connections between Prompts and Rubrics	28
5.3 What are the similarities and differences between NAEP and state writing assessments?	29
5.4 Insights from a combined use of the two approaches	30
Prompts	31
Rubrics	31
Prompts and Rubrics Associations	33
6. Discussion	34
6.1 Prevalent Writing Assessment Practices	34
6.2 Genre Demands in Direct Writing Assessments	38
6.3 State and National Alignment	40
7. Implications	41
8. Limitations	42
CHAPTER 2: Predicting Students' Writing Performance on NAEP from Assessment Variations	

	44
1. Introduction	44
2. Research Questions	49
3. Method	49
3.1 State and NAEP Direct Writing Assessments	49
3.2 Coding taxonomy	51
3.3 Coding Procedure	53
3.4 Distance between State Assessments and the NAEP	54
3.5 NAEP Sample	55
3.6 Students' NAEP Composition Performance	55
3.7 Students' Characteristics in NAEP	57
3.8 Structure of the Data Set and Statistical Analyses	59
3.9 Statistical Models	60
Unconditional model (Model 1)	61
Main effect model (Model 2)	61
Main effect model (Model 3)	62
Main effect model (Model 4)	62
4. Results	63
5. Discussion	69
6. Implications	71
7. Limitations	72
CHAPTER 3: Genre Demands in State Writing Assessments	74
1. Introduction	74
2. Research Questions	80
3. Method	81
3.1 State Direct Writing Assessments and Standards	81
3.2 Data Coding	82
Genre demands in prompts	83
Genres of prompts	83
Genre expectations in rubrics	84
Genre expectations in state standards	84
3.3 Data Analyses	85
4. Results	86
4.1a. How many state writing prompts possessed the problematic features of ambiguity or implicit genre expectations?	86
4.1b. Which key words in prompts were associated with ambiguity and implicit genre expectations, and how frequently do they appear?	89
4.2. What is the relationship between prompts' genre specification and rubrics' genre-mastery expectations?	95
4.3. What is the relationship between genre expectations in state standards and writing prompts?	99
5. Discussion	100
5.1 Ambiguity in prompts	100
5.2 Genre Expectation in Standards, Rubrics, and Prompts	102
5.3 Validity of State Writing Assessments	103

6. Implications	103
7. Limitations	105
CHAPTER 4: Summary and Moving Forward	106
1. Major Findings	106
1.1 Prevalent Writing Practices	106
1.2 Genre Demands in Direct Writing Assessments	107
1.3 State and National Alignment	107
1.4 The Relationship between the Variability between State and National Assessments and Students' NAEP Performance	108
1.5 The Relationship between Students' Characteristics and their NAEP Performance	108
1.6 Ambiguity in Prompts and Genre-mastery Rubrics	110
1.7 Genre Expectation in Standards and Genres Assessed	110
2. Implication for Writing Assessment Practices	111
2.1 For State Writing Assessment and NAEP	111
2.2 Writing Prompt Design	112
3. Implication for Writing Instruction	112
4. Next Steps for Research	114
APPENDICES	116
Appendix A Tables	117
Appendix B Coding Taxonomies	132
Appendix C State Direct Writing Assessments	146
BIBLIOGRAPHY	152

LIST OF TABLES

Table 1 Prompt-Rubric Contingencies for 81 Prompts	28
Table 2 States with Genre-Mastery Rubrics and/or State with Rubrics Containing Genre-Specific Components	32
Table 3 Genre Assessed in States with both Genre-Mastery Rubrics and Rubrics Containing Genre-Specific Components	33
Table 4 Sample Sizes, Achievement, and Student Demographics, 27 State Grade 8 HLM Sample	56
Table 5 HLM Model Results	66
Table 6 Frequency (F) and Percentage (P) of Key Words Usage in Genres	91
Table 7 Prompts with Problematic Features and Used with Genre-Mastery Rubrics	96
Table 8 NAEP Coding & Frequency Counts and Percentage of States	117
Table 9 Sample Sizes, Achievement, and Student Demographics, 27 State Grade 8 NAEP Reporting Sample	120
Table 10 Comparison of Sample Sizes and Student Demographics for 27 State Grade 8 NAEP Reporting Sample and HLM Sample	121
Table 11 Raw Unweighted Descriptive Statistics of Variables in HLM Models	123
Table 12 Genre Expectations in Standards and Genre Assessed	125
Table 13 Prompt Coding—Troia & Olinghouse’s (2010) Coding Taxonomy	132
Table 14 Rubric Coding—Troia and Olinghouse’s (2010) Coding Taxonomy	136
Table 15 Rubric Coding—Jeffery’s (2009) Coding Taxonomy	141
Table 16 Seven-Genre Coding Scheme for Prompts—Adapted from Jeffery (2009) and Troia & Olinghouse (2010)	142
Table 17 Standards Genre Coding—Troia and Olinghouse’s (2010) Coding Taxonomy Modified to Accommodate Jeffery’s (2009) Genre Coding Taxonomy	144
Table 18 State Direct Writing Assessments	146

LIST OF FIGURES

Figure 1 Genre Categories for 81 Prompts	26
Figure 2 Criteria Categories for 38 Rubrics	27

INTRODUCTION

There are persistent discrepancies between state and national writing assessment results (Lee, Grigg, & Donahue, 2007; Salahu-Din, Persky, & Miller, 2008). High proficiency levels are often reported for state-mandated assessments, while low proficiency levels are reported for the National Assessment of Educational Progress (NAEP). A possible explanation for this gap is that state and national assessments vary in the ways they define the writing construct and measure proficiency (Jeffery, 2009). The No Child Left Behind Act of 2001 (NCLB) gave states the freedom to adopt vastly different standards for English language arts, and allowed states to define content area proficiency levels and flexibly design their accountability systems (U.S. Department of Education, 2004). As a result, “states’ content standards, the rigor of their assessments, and the stringency of their performance standards vary greatly” (Linn, Baker, & Betebenner, 2002, p.3). However, little is known about how these tests vary.

When the content and format of state-mandated assessments are comparable to the national assessment, students are indirectly prepared for the NAEP. However, whether students actually achieve higher scores on the NAEP when their state assessments are more similar to it and lower scores when their state assessments are less similar is unknown. In other words, whether this variation between state and national writing assessments predicts students’ performance on the NAEP remains unexamined.

Currently, the Common Core State Standards (CCSS) have been formally adopted by 45 states and the District of Columbia. Developed by two multistate consortia, the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment for Readiness for College and Careers (PARCC), K-12 assessments aligned with CCSS will be in place starting with the 2014-2015 academic year. While this multistate effort may address the persistent discrepancies

between state and national writing assessments, it cannot explain the existing gap. A study of the state and national writing assessments will not only contribute to explaining the existing gap, but also inform policymakers and test designers by identifying the central characteristics of writing constructs valued in the past and advise them in the further development of new writing assessments.

The research presented in Chapter 1 examines what constitutes the writing construct in state writing assessments and the NAEP, and explores the similarities and differences between them through content analysis of state and NAEP writing prompts and rubrics. My adoption of Troia & Olinghouse's (2010) comprehensive coding taxonomy and Jeffery's (2009) genre-based coding schemes for content analysis ensures a broad presentation of recent thinking about writing development, instruction, and assessment, and allows an in-depth look into the variability of conceptions of writing constructs across states.

The research presented in Chapter 2 builds on the research presented in Chapter 1 by examining whether the differences between state and national writing assessments can explain some of the discrepancies found in the results of these assessments. This study quantifies these differences as the Euclidean distance between state and NAEP writing constructs as defined by the 90 indicators in Troia & Olinghouse's (2010) and Jeffery's (2009) coding taxonomies. The study explores the relationship between these differences and students' NAEP performance through Hierarchical Linear Modeling (HLM). The findings suggest that students' performances on the NAEP reflect both their writing abilities and how well they are prepared for the type of assessments the NAEP conducts. However, the large amount of unexplained variance between students' performances on NAEP from state to state suggests that there are more state-level variables to be explored. This result does not suggest that state and NAEP assessments should be

made more similar to each other; rather, components of these assessments such as prompts and rubrics should be examined to see whether they reflect evidence-based practices and whether they ensure the valid interpretation and use of the results of those assessments.

Following the recommendations of the research presented in Chapter 2, the research presented in Chapter 3 investigates the prompts in state writing assessments in depth and identifies ambiguities and implicit genre expectations in the design of these prompts. Ambiguity is defined as the presence of two or more genre demands in a prompt, while implicit genre expectations in prompts means a lack of verbs (e.g., argue, convince) or nouns (e.g., stories) that explicitly signal the desired genre. This is especially problematic when a prompt that is ambiguous or has implicit genre expectations is used with a rubric that emphasizes genre mastery. Therefore, the study also examines the use of genre-mastery rubrics with prompts that possess problematic features. When state writing assessment prompts are ambiguous or contain implicit expectations, a question is raised about whether the assessment is effectively and accurately evaluating the students' mastery of the genre in question. State standards provide an answer by specifying what students are expected to learn. Therefore, this study also examines state standards to identify the range of genres expected of middle school students. This study highlights the connection between genre demands in writing prompts and genre-mastery expectations in rubrics and state standards.

Together, this research investigates the writing constructs underlying state and national writing assessments, explore(s) the relationship between the variability in state and national assessments and students' NAEP performance, and examine(s) an important component of writing assessments—prompts—in depth. The findings should raise people's awareness that

students' performances on the NAEP do not only measure their writing abilities but also reflect how well they are prepared for the type of assessments the NAEP uses.

Poorly developed assessments will provide inaccurate evaluations of students' abilities, impact curriculum in unwarranted ways, and lead to wrong decisions regarding students' promotion and retention, as well as imprecise ratings of teacher effectiveness. These findings can advise test designers about what central characteristics of the writing construct have been valued in the past, and can be used in the development of new writing assessments. Furthermore, it is hoped that these findings will direct the assessment and writing research communities' attention to validity-related issues in large-scale writing assessments and encourage more research to study components of these large-scale writing assessments.

CHAPTER 1: Examining Writing Constructs in U.S. State and National Assessments

1. Introduction

In the U.S., persistent discrepancies exist between state and national writing assessment results (Lee, Grigg, & Donahue, 2007; Salahu-Din, Persky, & Miller, 2008). The results of the National Assessment of Educational Progress (NAEP) show low proficiency levels, yet state-mandated assessments often report high proficiency levels. This inconsistency suggests that, in order to ensure that the results of state and national assessments are comparable, more uniform academic and assessment standards may be necessary.

One solution to this gap, the Common Core State Standards (CCSS), has already been formally adopted in 45 states and Washington, D.C. Two multistate consortia, the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment for Readiness for College and Careers (PARCC) worked together to develop K-12 assessments aligned with the CCSS. These assessments will be implemented for the 2014-2015 school year.

Although these multistate efforts have attempted to address the persistent discrepancy between the results of state writing assessments and the NAEP, they do not explain the existing gap. One possible explanation of this gap is the varying ways in which state and national assessments define the writing construct, and the differences in the measures they use to determine proficiency levels (Jeffery, 2009). It is difficult to state with certainty whether these variations fully account for the inconsistent results, though, because little is known about how these assessments actually vary.

The No Child Left Behind Act of 2001 (NCLB) required states to implement statewide accountability systems that consisted of challenging state standards and annual testing for all grade 3-8 students. At the same time, these NCLB requirements were flexible enough that states

were able to adopt dramatically different standards for English language arts instruction and assessment, some of which placed little emphasis on writing (Jeffery, 2009); this flexibility also let each state define their own content area proficiency levels and design appropriate accountability systems to assess those proficiency levels (US Department of Education, 2004). As a result, “states’ content standards, the rigor of their assessments, and the stringency of their performance standards vary greatly” (Linn, Baker, & Betebenner, 2002, p.3).

Variation in states’ standards, assessments, and performance benchmarks is associated with differing conceptions of writing performance (Jeffery, 2009). On the one hand, this variability may produce the discrepancy that is consistently observed between state assessments and NAEP results and make state assessment and NAEP results difficult to reconcile. On the other hand, the variability in the underlying conceptions of writing proficiency raises the concern that teachers are emphasizing different aspects of composition in U.S. classrooms (Jeffery, 2009), because research has shown that tests impact instruction (Hillocks, 2002, Moss, 1994). Hillocks (2002) found that writing instruction in classrooms is often used to help students prepare for high-stakes assessments. In other words, whatever is valued in the assessments students will take is what tends to be taught; the state-to-state variability in the underlying conceptions of writing proficiency in assessment contexts thus leads to the variability of writing instruction found in U.S. classrooms.

What constitutes the writing construct is complex. It can be understood through and approached with multiple theoretical frameworks. A comprehensive perspective ensures a broad presentation of current thinking about writing development, instruction, and assessment; thus, such a perspective is more likely to shed light on the underlying writing construct. Troia and Olinghouse (2010) developed a coding taxonomy to examine writing standards and assessments.

This taxonomy was derived from several theoretical frameworks, including Hayes' cognitive model of writing (Flower & Hayes, 1981; Hayes, 1996), socio-cultural theory (Prior, 2006), genre theories (Dean, 2008), linguistic models of writing (Faigley & Witte, 1981), and motivation theories (Troia, Shankland, & Wolbers, 2012). It consists of seven strands: (1) writing processes, (2) context, (3) purposes, (4) components, (5) conventions, (6) metacognition and knowledge, and (7) motivation.

Adopting this framework allows an in-depth look into the variability of conceptions of the writing construct across states; therefore an analysis that uses it can inform policy makers and test designers about the extant ways the writing construct is defined and proficiency is measured to guide further development of writing assessments. Results from this type of analysis can also advise them on which core characteristics of the writing construct that were valued in the past can continue to be used in the future to supplement the CCSS and the common assessments in each state. Moreover, these results can help the assessment community examine the validity of those large-scale writing assessments.

2. Review of Literature

Dean (1999) conducted content analyses on some popular secondary composition textbooks and studied sample writing tests from Texas, California, and Washington. The study showed that while the textbooks reflected both traditional and current theories of writing, the large-scale writing assessments reflected traditional rhetoric characteristics, which emphasize style, form, and the mechanical aspects of writing. Hillocks (2002) studied writing assessments in five states—New York, Illinois, Texas, Kentucky, and Oregon—and conducted 390 interviews with teachers and administrators. He found that state assessments tended to undermine state standards and encourage writing instruction that helped prepare students for

high-stake assessments. As a result, high stakes testing does not guarantee quality classroom instruction; instead, it encourages ineffective teaching and can come with unintended consequences such as promoting a formulaic approach to writing.

Beck & Jeffery (2007) examined 20 exit-level state writing assessment prompts from Texas, New York, and California, using task analysis of the prompts and genre analysis of the corresponding high-scoring benchmark papers. They found that there was a lack of alignment between the genre demands of the prompts and the genres of the corresponding benchmark papers. The comparison of the genre demands in the prompts with the actual genres produced in the corresponding benchmark papers showed that there was much greater genre variation in the expected responses of Texas and California writing assessments than those from New York. Only 20% of the California benchmark papers and 18% of the Texas benchmark papers were aligned with the prompts, while 42% of the New York benchmark papers were aligned. Jeffery's (2009) study of 68 prompts from 41 state and national exit-level direct writing assessments suggested that national writing assessments were different from state assessments in the degree that they emphasized genre distinctions and provided coherent conceptualizations of writing proficiency. The genre expectations in national writing assessments were consistently associated with rubric criteria whereas this was not true of state assessments.

Studies that have examined how conceptualizations of writing constructs vary among U.S. states have either examined small samples of states (Dean, 1999; Beck & Jeffery, 2007) and their writing assessments (Hillocks, 2002), or targeted exit-level writing assessments for high school students (Jeffery, 2009). Few studies have investigated how conceptualizations of writing construct vary in middle schools among U.S. states. A look into what is emphasized in middle school writing assessments, as well as the various definitions of the writing construct, will shed

light on the expectations of writing competence placed on students. Once deeper understandings of these expectations and differences are developed, more resources can be allocated to help students navigate this important but challenging stage of their writing development.

Middle school is an important stage for students to develop their abstract thinking and more sophisticated ways of using language (De La Paz & Graham, 2002). Students who do not learn to write well are less likely to use their writing to extend their learning, and more likely to see their grades suffer (National Commission on Writing for America's Families, Schools, and Colleges [NCWAFSC], 2003, 2004). As an important transitional step for students between elementary and high school, middle school education lays down a foundation for students' studies in high school and later college. Weak writers in middle school suffer the consequences of the growing trend of using writing proficiency as a factor in grade retention and advancement and continue to be at great disadvantage in high school, and are thus less likely to attend college (Zabala, Minnici, McMurrer, & Briggs, 2008).

The NAEP assesses students' writing at eighth grade, and seventh and eighth graders are also frequently assessed in state writing assessments. Consequently, a large sample can be derived from states' middle school writing assessments to compare with the NAEP's direct writing assessments. Direct writing assessments generally consist of writing prompts to guide the student in writing about a particular topic; for example, a student may be presented with a picture, and asked to write a response to that picture. This study aims to fill in gaps in the research on large-scale writing assessments with a broader comparison by using writing assessment prompts from 27 states and the NAEP 2007 writing prompts to examine the features of states' and NAEP's direct writing assessments, and to explore the similarities and differences between state and national writing assessments at the middle school level. The NAEP 2007 data

was selected because it contained state-level writing data and allowed state-level modeling whereas the NAEP 2011 does not.

Troia and Olinghouse's (2010) coding taxonomy is one analytical tool utilized for this research. The indicators found within the seven strands in the coding taxonomy cover (a) all stages of the writing process and specific composition strategies; (b) circumstantial influences outside the writer that can impact writing performance; (c) a variety of communicative intentions accomplished through different genres; (d) the features, forms, elements, and characteristics of different texts; (e) the mechanics of producing text; and (f) the knowledge resources and (g) personal motivational attributes within the writer that drive writing activity and writing development. In writing assessments, the writer's motivation (i.e., general motivation, goals, attitudes, beliefs, and efforts) does not apply, because states rarely administer assessment documents such as surveys alongside writing assessments to measure writers' personal attributes. Thus, the seventh strand from the original coding taxonomy was not used in this study.

Genre Theories in Composition

Among various theoretical frameworks, genre theories have been used to examine large-scale writing assessments (Beck & Jeffery, 2007; Jeffery, 2009) and thus deserve further mention. Genres thread through all elements of composition, and shape students' ways of thinking about the writing process. Different genres direct students to proceed differently with different stages of writing process. For example, writing a persuasive composition makes planning of certain content more useful than that for a narrative piece (Dean, 2008). Outlines that direct students to begin their essays with thesis statements and explicit arguments and to continue with evidence that supports those arguments and refute the counter-arguments are more appropriate for persuasive compositions than outlines that direct students to begin their paper by

setting the scene and by continuing with a sequence of actions. Thus, knowing how to effectively adopt writing process approaches for different genres in assessment contexts will help students compose their texts more efficiently.

Genres connect texts and contexts. Devitt, Reiff, and Bawarshi (2004) proposed strategies to help students deliberately use genres to make such a connection:

[Teachers] teach students to move from observation of the writing scene and its shared goals, to the rhetorical interactions that make up the situations of this scene (the readers, writers, purposes, subjects, and settings), to the genres used to participate within the situations and scenes. (p. xviii)

In other words, students are taught to observe the context in which the desired writing is expected to fulfill the communicative intent, and then use appropriate genres to fulfill this communicative need; thus, genres bridge texts and contexts. For example, the school is organizing a field trip. Students may have places that they would like to visit; thus, a persuasive letter would be an appropriate genre to fulfill their communicative needs of convincing the audience of their letters—likely school teachers, administrators, and staff—to allow them to visit the places that they would like to visit.

Genres also serve writing purposes. When students study genre, they are “studying how people use language to make their way in the world” (Devitt, 1993). If not taught explicitly what each genre means students will lack the knowledge of genres’ structures, and have a difficult time coming up with appropriate writing content for different purposes. For example, when students are expected to write to persuade, without having genre knowledge of the structural elements and/or information that is canonical for persuasive papers, students may be unable to use argumentation schemes—“ways of representing the relationship between what is stated in the

standpoint and its supporting justificatory structure” (Ferretti, Andrews-Weckerly, & Lewis, 2007, p.277)—such as argument from consequences and argument from example. For example, one prompt asked students to write about whether they think it is a good idea for their school to have candy and soda machines. Those who were against this idea could have argued that these machines would promote unhealthy eating habits among students. This would be an argument from potential negative consequences, which could be further illustrated with examples. For instance, the fact that students purchased candy and soda more frequently and consumed more unhealthy food than before could be cited to illustrate the argument that candy and soda machines’ promoting unhealthy eating habits.

Genres specify the writing content (i.e., features, forms, elements, or characteristics of text) to be included in a text. Donovan and Smolkin (2006) believed that “an important part of ‘doing school’ is mastering the most frequently appearing generic forms” (p.131). Berkenkotter and Huckin (1995) argued that “genres are essential elements of language just as words, syntactic structure, and sound patterns. In order to express one’s individual thoughts, one must use available patterns for speech, that is to say, genres, in one way or another” (p. 160). There are established genres in every language; people choose them and modify them to achieve various purposes by relying on those writing components. For example, to tell a story, writers will have a story line, setting, plot, and characters, as well as dialogue and a climax to elicit an emotional response from the reader.

Genre impacts the mechanics of writing and guides formats. The content requirements specified in a genre, and writers’ consideration of purpose and audience impact their use of vocabulary and/or word choice, which potentially affect spelling (Pasquarelli, 2006). For example, there are differences between vocabularies used in informative writing versus narrative

writing; these differences play out in the spellings of abstract technical vocabulary used in one versus the more colloquial vocabulary used in the other. The sentence structure dictated by a genre also impacts the use of punctuation, such as the often unorthodox use of punctuation in poetry. Also, genres such as poetry have their established formats.

Genre knowledge also serves as an important component of the total writing knowledge students need for successful composition. Genre knowledge is the knowledge about the purposes of writing and the macrostructures of a text including text attributes, elements, and structure common to specific types of writing. Donovan and Smolkin (2006) stated that “genre knowledge develops prior to conventional writing abilities” (p.131). Though genre knowledge does not guarantee successful performance, there is an interactive relationship between genre knowledge and performance; prior genre knowledge can prompt students’ writing performances under new circumstances both in positive and negative ways and expand knowledge of a particular genre to various strategies (Bawarshi & Reiff, 2010; Devitt, 2009; Dryer, 2008; Reiff & Bawarshi, 2011).

Jeffery (2009) used genre theories to explore the writing construct underlying state writing assessments. Jeffery’s (2009) study was based on Ivanic’s (2004) framework of six “discourses of writing”—“skills discourse,” “creativity discourse,” “process discourse,” “genre discourse,” “social practices discourse,” and “social political discourse” (Ivanic, 2004, p. 224). Ivanic defined “discourses of writing” as “constellations of beliefs about writing, beliefs about learning to write, ways of talking about writing, and the sorts of approaches to teaching and assessment which are likely to be associated with these beliefs” (p.224).

In Ivanic’s framework, “skills discourse” describes writing as applying knowledge of sound-symbol relationship and syntactic patterns to compose a text. Thus, a big part of “learning

to write” is learning sound-symbol relationships and syntactic patterns. Likewise, the “teaching of writing” involves the explicit teaching of skills such as phonics, with accuracy emphasized in the assessment criteria (Ivanic, 2004).

“Creativity discourse” views writing as the product of an author’s creativity. “Learning to write” is therefore expected to be achieved by writing on topics that interest writers. The “teaching of writing” also involves implicit teaching of creative self-expression. In this case, “whole language” and “language experience” are emphasized, while interesting content and style are valued in the assessment criteria (Ivanic, 2004).

Ivanic calls the practical realization of the composing processes in the writer’s mind “process discourse.” In this view, “learning to write” is learning both the mental and practical processes in composing a text, and the “teaching of writing” involves explicit teaching of these processes (Ivanic, 2004).

Writing as text-types forged by social context is termed “genre discourse” by Ivanic. In this understanding, “learning to write” is thus to learn the characteristics of different types of writing that serve different purposes in different contexts. Predictably, the “teaching of writing” involves the explicit teaching of genres. The appropriateness of the genre utilized by students is valued in assessment criteria (Ivanic, 2004).

“Social practices discourse” portrays writing as purpose-driven communication in a social context. Consequently, the point of “learning to write” is to write for real purposes in real-life contexts. Therefore the “teaching of writing” involves explicit instruction in functional approaches and the implicit teaching of purposeful communication. Whether writing is effective for the given purpose is valued in assessment criteria in this case (Ivanic, 2004).

Finally, “socio-political discourse” explains writing as a socio-politically constructed practice open to contestation and change. “Learning to write” is therefore the process of understanding why different types of writing have their unique characteristics and to choosing a position from alternatives. “Teaching to write” involves explicit teaching of critical literacy skills, including “critical language awareness.” Social responsibility is highly valued in assessment criteria in this discourse (Ivanic, 2004).

Through an inductive analysis of the rubrics for the exit-level writing assessment prompts, Jeffery (2009) developed a five-criteria coding scheme for rubrics: rhetorical, genre-mastery, formal, expressive, and cognitive. These rubric types represent what different “discourses of writing” value as assessment criteria (Ivanic, 2004). Rhetorical rubrics focus on “the relationship between writer, audience, and purpose across criteria domains” (Jeffery, 2009, p.10). Genre-mastery rubrics emphasize “criteria specific to the genre students are expected to produce” (Jeffery, 2009, p.11). Formal rubrics conceptualize “proficiency in terms of text features not specific to any writing context” (Jeffery, 2009, p.11). Cognitive rubrics target “thinking processes such as reasoning and critical thinking across domains” (Jeffery, 2009, p.12). Expressive rubrics conceptualize “good writing” as “an expression of the author’s uniqueness, individuality, sincerity and apparent commitment to the task” (Jeffery, 2009, p.12).

Meanwhile, through an inductive analysis of exit-level state direct writing assessments, Jeffery (2009) developed a six-genre coding scheme for prompts. The six genres of prompts are argumentative, persuasive, explanatory, informative, narrative, and analytic. Argumentative prompts differ from persuasive prompts by calling abstractly for “support” of a “position” and by not designating a target audience. An example of an argumentative prompt is “many people believe that television violence has a negative effect on society because it promotes violence. Do

you agree or disagree? Use specific reasons and examples to support your response.” In contrast, persuasive prompts require students to convince an identified audience to act on a specific issue. Moreover, persuasive prompts are unlike argumentative prompts because they invite students to take a one-sided perspective on an issue, while argumentative prompts often expect students to consider multiple perspectives on an issue. An example of a persuasive prompt is “you want your parent or guardian to allow you to go on a field trip with your classmates. Convince your parent or guardian to allow you to do this.” In contrast to argumentative and persuasive prompts, “which explicitly identify propositions as arguable and direct students to choose from among positions” (p.9), explanatory prompts anticipate that students will “explain how or why something is so” (p.9). An example of an explanatory prompt is “a good friend plans to visit you for the first time in the U.S. You want to help him/her get ready for the trip. Explain what you would do.” With the above coding frameworks, 68 prompts and 40 rubrics were coded in Jeffery’s (2009) study, and the inter-rater agreement was .87 for prompt coding and .83 for rubric coding.

Jeffery (2009) suggested that one way to illuminate the underlying construct conceptualizations in large-scale writing assessments is to analyze the relationships between genre demands and scoring criteria. Jeffery’s (2009) six genre coding taxonomy can be used to supplement Troia and Olinghouse’s (2010) coding taxonomy by further differentiating the persuasive and argumentative genres. On the other hand, Jeffery’s (2009) five-criteria coding scheme can be used to code rubrics to study how prompts and rubrics are associated, while Troia and Olinghouse’s (2010) coding taxonomy allows an examination of the writing constructs defined by prompts and rubrics together.

3. Research Questions

This study explores how state and national assessments define and measure the writing construct by studying the features of their writing assessments.

More specifically, this study aims to answer the following questions:

1. How do the features of writing prompts and rubrics vary across a sample of states and the NAEP?
2. What are the connections between these prompts and rubrics, especially in terms of their genre demands?
3. What are the similarities and differences between NAEP and state writing assessments?

4. Mode of Inquiry

4.1 State and NAEP Direct Writing Assessments

This study was built upon a prior Institute of Education Sciences (IES)-funded study—the K-12 Writing Alignment Project (Troia & Olinghouse, 2010-2014). In the K-12 Writing Alignment Project, appropriate assessment personnel were located through states' Department of Education websites. Email inquiries and phone calls were conducted to request documents. Because the K-12 Writing Alignment Project examined the alignment between state writing standards and assessments prior to the adoption of the CCSS, the use of the NAEP 2007 assessment ensured that students' NAEP results were an effect of the instruction under state writing standards and assessments current at that time. Also, the NAEP 2007 data contained state-level writing data and allowed state-level modeling, whereas the 2011 data did not. Because the NAEP assessment with which state assessments were compared was from 2007, state direct writing assessments were gathered mainly from between 2001 and 2006 to ensure the representation of the time period. Also, because the study aimed to analyze representative state

writing assessments, while some states had major revisions which changed what their representative writing assessment might be, it was important to identify the number and dates of the major revisions between 2001 and 2006.

After the number and dates were identified, a representative writing prompt, its rubric, and the administrative manual for each genre in each grade being assessed were collected from each time span between major revisions. This resulted in the selection of 78 prompts and 35 rubrics from 27 states¹ in total (See Appendix C for details). There was no NAEP data available for Alaska, Nebraska, Oregon, and South Dakota for the time period in question. There were no state writing standards or writing assessments available for Connecticut, Iowa, Pennsylvania, Montana and New Mexico between 2001 and 2006. Ohio did not assess 7th grade and 8th grade for writing during the period 2001-2006. Therefore, those states' direct writing assessments were not included in this analysis.

Next, state direct writing assessment documents were compiled to include (a) verbal directions from administration manuals for direct writing assessments; (b) actual prompts; (c) supporting materials provided (e.g., dictionary or writer's checklist); (d) sessions arranged for writing tests (e.g., planning session, drafting session, revising session); (e) time given; (f) page limits; and (g) whether (and what kind(s) of) technology was used. The number of compiled documents for each state corresponded with the number of responses expected from students each year. In other words, if students were expected to respond to one prompt with rotated genres each year, prompts from the rotated genres were all compiled into a single document to represent the scope of genres assessed. If students were expected to respond to multiple prompts each year,

¹ The following chose not to participate in the study: Colorado, Delaware, the District of Columbia, Georgia, Hawaii, Maryland, Minnesota, Mississippi, New Hampshire, New Jersey, North Dakota, South Carolina, Utah, and Wyoming.

those prompts were compiled separately into multiple documents. These compiled documents and rubrics were coded with the coding taxonomy.

The publically-released NAEP 2007 writing prompts, scoring guide, and writing framework were collected. There were three NAEP writing prompts from eighth grade included in this analysis: a narrative prompt, an informative prompt, and a persuasive prompt. These three writing prompts were included because they were publicly available and considered representative of the genres assessed. Other writing prompts were not released due to possible future use.

4.2 Coding Taxonomy

This study used the coding taxonomy developed by Troia and Olinghouse (2010) which was modified to accommodate Jeffery's (2009) genre coding scheme for prompts, as well as her criteria coding scheme for rubrics. These two coding frameworks served to provide comprehensive coverage of the underlying writing construct, focused study of the powered genres in state and NAEP direct writing assessments, and the relationships between prompts and rubrics. When used to code the writing prompts, Troia and Olinghouse's (2010) coding taxonomy ensured comprehensive coverage of the writing construct as measured by the 80 indicators under the six strands; thus, not only the genre demands of the writing prompts were examined, but also the writing process, the assessment context, and the required writing knowledge. Jeffery's (2009) coding taxonomy, derived from an inductive analysis of exit-level state direct writing assessments, focused on the genre demands of the writing prompts and could differentiate among similar genres such as the persuasive and argumentative genres, as well as the expository and informative genres. As a result, a seven-category genre coding scheme (see Table 16 in Appendix B) was developed by adapting from the third stand (i.e., purpose) of Troia

and Olinghouse's (2010) coding taxonomy and Jeffery's (2009) genre coding scheme. These seven categories are: descriptive, persuasive, expository, argumentative, informative, narrative, and analytic. When used to code the writing rubrics, Troia and Olinghouse's (2010) coding taxonomy ensured a comprehensive coverage of the writing components and the writing conventions noted in the writing rubrics. Together with the coding from writing prompts, they defined the writing constructs assessed. Jeffery's (2009) coding taxonomy categorized the writing rubrics based on the most prominent features of the rubrics—each rubric could only appear in one of the categories (i.e., rhetorical, formal, genre-mastery, cognitive, and expressive). The taxonomy identified the most dominant rubrics used for each genre of writing; thus, there would be associative patterns between genre demands in the prompts and rubric categories. In summary, Troia and Olinghouse's (2010) coding taxonomy examined the writing construct defined together by prompts and rubrics while Jeffery's (2009) coding taxonomy focuses on the genre demands and the connections between prompts and rubrics. For the current study, these two taxonomies can complement each other to reveal the writing constructs underlying large-scale writing assessments.

4.3 Procedure

In the K-12 Writing Alignment Project, three raters coded state and NAEP writing prompts with the first (writing processes), second (context), third (purposes), and sixth strands (metacognition and knowledge) from Troia and Olinghouse's (2010) coding taxonomy. The first rater, paired with either the second rater or the third rater, coded each compiled assessment document. The inter-rater reliabilities in this study were all calculated using Pearson r absence and presence of agreements. The inter-rater reliability of rater 1 and rater 2 was .97 for prompt coding; the inter-rater reliability of rater 1 and rater 3 was .95 for prompt coding. The reason that

only four strands were coded with prompts was that writing processes and writing contexts were often specified in the verbal directions for test administration, and writing purposes and writing knowledge were often specified in the writing prompts. Two separate raters coded state and NAEP writing rubrics with the fourth (components) and fifth (conventions) strands from Troia and Olinghouse's (2010) coding taxonomy. These last two strands were coded with rubrics because writing components and writing conventions were often specified in scoring rubrics. The inter-rater reliability was .95 for rubrics coding. Differences were resolved through discussion.

Two raters coded state and NAEP writing prompts with the seven category-genre coding scheme adapted from the third strand (purpose) of Troia and Olinghouse's (2010) coding taxonomy and Jeffery's (2009) genre coding scheme. These raters also coded state and NAEP writing rubrics with Jeffery's (2009) criteria coding scheme. The author of this dissertation served as one of the two raters. A graduate student in Digital Rhetoric & Professional Writing served as the second rater. The two raters first practiced coding with a training set. When they reached 85% inter-rater agreement, they moved into coding the actual prompts and rubrics. The inter-rater reliability was .93 for prompt coding and .86 for rubrics coding. Differences were resolved through discussion.

5. Results

5.1 How do the features of writing tasks and rubrics vary across a sample of states and NAEP?

There were direct writing assessments from 27 states in the sample; however, because Rhode Island and Vermont had the same New England Common Assessment Program (NECAP) direct writing assessment, there were 26 distinct sets of prompts and rubrics from state writing assessments. In the sample, 15 states had 7th grade writing assessments, and 18 states, including

Rhode Island and Vermont, had 8th grade writing assessments. There were six states that had both 7th grade and 8th grade assessments (see Appendix C).

According to Troia & Olinghouse's (2010) coding taxonomy, the writing constructs assessed in state and national assessments were defined by prompts and rubrics together and consisted of the writing process, writing context, writing content, writing mechanics, and writing knowledge (see Table 8 in Appendix A).

Writing Process

There were four states that had general references to the writing process in their writing directions and three states that gave students a choice of prompts. Out of 27 states, all but one directed students to plan their compositions before they wrote. However, while the majority of these states gave students planning pages, they did not give students separate planning sessions. Only Kansas and Nevada gave students both pages and sessions for planning. Compared with planning and drafting, revising was a less emphasized stage of the writing process. There were twelve states that did not direct students to revise. Among the other fifteen states, only Kansas and Massachusetts gave students both time and pages for revision. Arizona, Kentucky, Missouri, and Washington gave students pages for revision (but no extra time), and only Nevada gave students 15 minutes for revision (but no extra pages). One possible explanation of why fewer states focused on revision was that some states directed students to edit rather than revise. For example, 18 states directed students to edit. However, there were still seven states that did not direct students to revise or edit their writing. Ten states emphasized the importance of publishing by directing students to write a final product.

There were ten states that offered test-taking strategies to students. The most popular test-taking strategy was about space management—e.g., Massachusetts included the following verbal

directions in their administration manual “YOU MUST LIMIT YOUR WRITING TO THESE FOUR PAGES; BE SURE TO PLAN ACCORDINGLY (originally capitalized)”

(Massachusetts, 2002, Grade 7)— with seven states advising students of that; two states advised students about time management—e.g., Oklahoma included the following verbal directions in their administration manual “Try to budget your time wisely so you will have time to edit and revise your composition” (Oklahoma, 2006, Grade 8); and one state offered students strategies about topic choice—i.e., Kansas’ administration manual contained the instruction that “you will choose the one topic that you like the most and that you feel will allow you to do your best writing. Keep this in mind as you consider each description” (Kansas, 2004, Grade 8).

Writing Context

Seven states gave students at least two writing tasks. New York gave students four integrated writing tasks—short and long listening and responding, and short and long reading and responding. Most states (20 out of 27) had a general mention of audience in their writing prompts. Prior to 2007, only West Virginia had online writing sessions; students in other states wrote on paper with pencils. Students in West Virginia were expected to log on to a website, where they wrote a multiple-paragraph essay equivalent to a one-to-two page handwritten essay. They did not have access to the spell check or grammar check options. Their papers were read and scored by a computer that had been trained with essays written by West Virginia seventh and tenth grade students. Within a few weeks West Virginia students would receive a detailed report of their scores.

Nineteen states provided procedure facilitators for students’ writing; the most popular procedure facilitators were checklists and rubrics. Eleven states allowed students to use dictionaries or thesauri during writing exams. The prompts of Arkansas and Idaho situated

students' writing in other disciplines; for example, "Your social studies class ..." or "As an assignment in your history class," None of the writing prompts required students to consider multiple cultural perspectives on an issue.

Only two states out of the 27 did not specify the response length; the typical length was two pages. Around half of the states in the sample (13/27) did not have a time limit on their writing assessments. Among the fourteen states that had a time limit, ten states had a specified amount of time with an average of 52 minutes; the other four states gave students 45 minutes with an optional extended period of time if needed.

Writing Components

All states evaluated the general organization and content of students' compositions in their rubrics; however, there were seven states that did not emphasize the general structure of students' essays and one state (i.e., Texas) that did not emphasize details. Ten states evaluated the genre-specific information of students' essays including organization, content, and ideas; specifically, five states evaluated narrative components, four states evaluated expository components, six states evaluated persuasive components, and three states evaluated response to writing components. Most states (24/27) evaluated sentence fluency, style, and semantic aspects (e.g., word choice) of students' compositions. Seven states emphasized the use of figurative language, one state (i.e., Kentucky) the use of citations and references, and no states considered the use of multimedia (which is consistent with paper-and-pencil writing tasks).

Writing Mechanics

The majority of states' writing rubrics had general reference to writing conventions (22 states), capitalization (19 states), punctuation (19 states), spelling (18 states), and grammar (24 states). Only Kentucky emphasized specific word-level capitalization and punctuation. Four

states emphasized students' correct use of punctuation regarding sentence ending and clausal linking. Six states emphasized the spelling of high frequency words; among these states, Wisconsin also emphasized the spelling of graphophonemic elements. In addition, Kentucky emphasized the spelling of abbreviations. The most frequently emphasized grammatical aspects were: sentence construction (19 states), verbs and verb phrases (7 states), pronouns (4 states), modifiers (4 states), nouns and noun phrases (3 states), adjectives (3 states), and adverbs (1 state, i.e., West Virginia). Only Arkansas and Kentucky had general reference to formatting; twelve states referred to specific aspects of formatting, e.g., paragraphing or using appropriate spacing between words and sentences.

Writing Knowledge

The majority of states (nineteen states) explicitly directed students to recall their topical knowledge when composing; the prompts in those states often set up situations in ways such as “think about a time” However, none of the states used prompts to evoke students' genre knowledge, linguistic knowledge, procedure knowledge, or self-regulation.

5.2 What are the connections between these prompts and rubrics, especially in terms of their genre demands?

Prompts

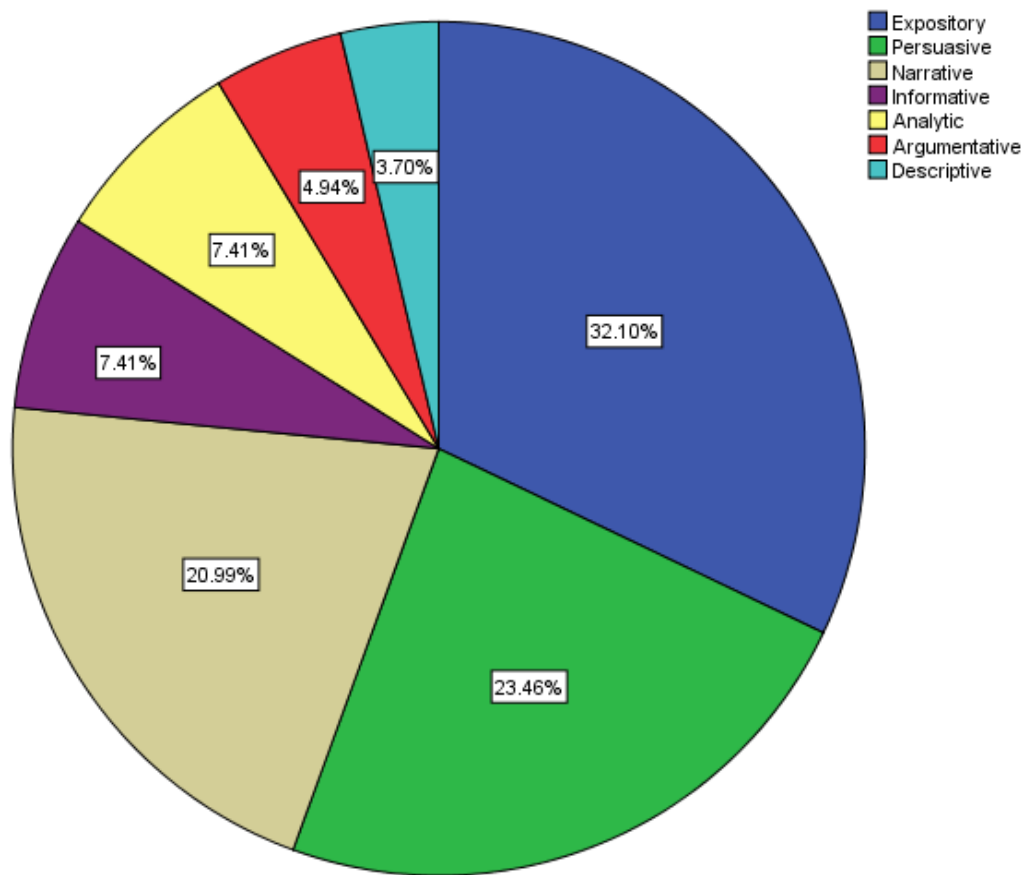


Figure 1 Genre Categories for 81 Prompts

Figure 1 shows the percentages of prompts of each genre in the sample. Out of 81 writing prompts, including three NAEP prompts, there were 26 expository, 19 persuasive, 17 narrative, 6 informative, 6 literary analysis, 4 argumentative, and 3 descriptive prompts. Expository and informative prompts combined comprised a little less than 40% of the prompts in the sample. Expository prompts and informative prompts either assessed students' abilities to "explain" how something worked and why or "provide" facts about more concrete objects. Persuasive essays were the second most used type of prompt and directed students to persuade an audience to agree with their positions on an issue. Similar to persuasive prompts, the four argumentative prompts directed students to provide evidence to support a position; however, they often did not explicitly

direct students to convince an identified audience. Together, persuasive and argumentative prompts were a little less than one-third of the prompts in the sample. Narratives were the third most assessed genre. They asked students to give an account of either an imaginary or actual incident. Narrative prompts often had straightforward directions such as “tell about a time when ...” or “write a story...” The three descriptive prompts differed from the informative prompts by directing students to provide attributes or details about an object, while the informative prompts often asked the students to provide facts.

Rubrics

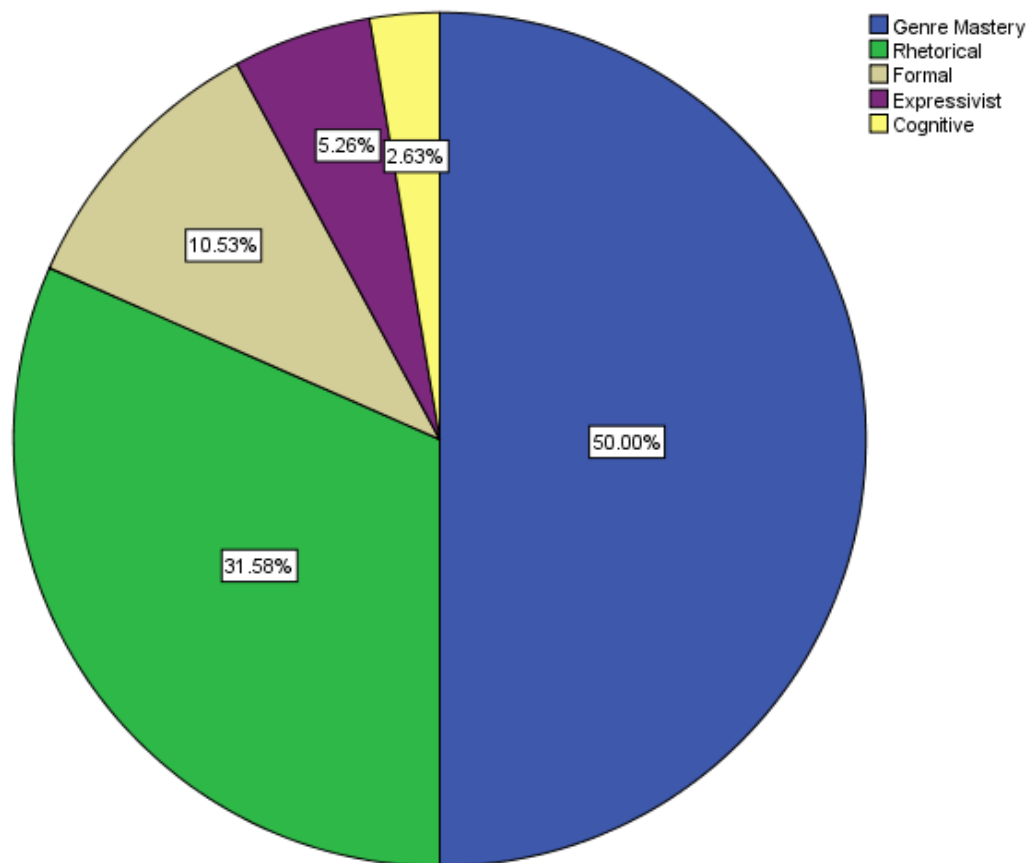


Figure 2 Criteria Categories for 38 Rubrics

Figure 2 shows the percentages of rubrics of each type in the sample. Among 38 rubrics, including three NAEP rubrics, there were 19 genre-mastery rubrics, 12 rhetorical, 4 formal, 2 expressivist, and only 1 cognitive. Genre-mastery rubrics were the most used rubrics in state and national direct writing assessments and comprised half of all the rubrics analyzed, emphasizing students' mastery of genres. Rhetorical rubrics were the second most used rubrics and comprised almost one-third of rubrics examined, emphasizing the importance of addressing the audience and achieving one's writing purposes. There were only a few formal rubrics, which emphasized the general structure and conventions of a paper. The two expressivist rubrics assessed students' creativity in composing their papers, and the single cognitive rubric emphasized students' critical thinking shown through their writing.

Connections between Prompts and Rubrics

Table 1 Prompt-Rubric Contingencies for 81 Prompts

Count		Rubric Category					Total
		Rhetorical	Genre-mastery	Formal	Cognitive	Expressivist	
Genre Categories for Prompts	Persuasive	6	8	2	3	0	19
	Expository	15	8	3	0	0	26
	Narrative	7	8	0	0	2	17
	Argumentative	3	1	0	0	0	4
	Descriptive	0	2	1	0	0	3
	Informative	3	3	0	0	0	6
	Analytic	1	5	0	0	0	6
Total		35	35	6	3	2	81

Table 1 shows the association between prompt genres and rubric types. Out of 81 prompts, there were 35 prompts assessed with rhetorical rubrics and 35 prompts assessed with

genre-mastery rubrics. There were only six prompts assessed with formal rubrics, three with cognitive rubrics, and two with expressivist rubrics.

For informative prompts, the number assessed with rhetorical rubrics and genre-mastery rubrics was the same. For persuasive and narrative prompts, there were slightly more prompts assessed with genre-mastery rubrics than rhetorical rubrics. The majority of analytic prompts were assessed with genre-mastery rubrics. There were only two descriptive prompts, both of which were assessed with genre-mastery rubrics. For expository and argumentative prompts, the majority were assessed with rhetorical rubrics.

Genre-mastery rubrics were used to evaluate all seven genres of writing—persuasive, expository, narrative, argumentative, descriptive, informative, and analytic. Rhetorical rubrics were used to evaluate all genres of writing except descriptive. Formal rubrics were used to evaluate persuasive and expository writing; cognitive rubrics were only used to evaluate persuasive writing; and expressivist rubrics were only used to evaluate narratives.

5.3 What are the similarities and differences between NAEP and state writing assessments?

More than 70% of states' middle school writing assessments involved: directing students to plan before drafting and to write for either a general audience or a specifically-identified audience; providing procedure facilitators such as checklists; specifying the length of the writing; and explicitly directing students to access their topical knowledge. The NAEP 8th grade writing assessments also had these characteristics. For example, NAEP directed students to plan, write, and review their writing, gave students a page for planning, and gave students a brochure of planning and reviewing strategies to facilitate students' writing. Also, NAEP did not give students separate sessions for different stages of writing or specify the length of students' writing.

Over 70% of states' middle school writing assessments evaluated students' texts' quality based on their organization, structure, content, details, sentence fluency, style, semantic aspects, and grammar. More than 60% of states' assessments evaluated students' essays on capitalization, punctuation, spelling, and sentence construction. The NAEP 8th grade writing tasks assessed all of the above aspects. The NAEP 8th grade test also directed students to write in response to two prompts and set a time limit of 25 minutes on each of these writing tasks. Only seven states required two responses from students. Around half of the states in the sample (13/27) did not have a time limit on their writing assessments. The other fourteen states had an average time limit of 52 minutes, or 45 minutes with an optional extended period of time if needed.

While expository, persuasive, and narrative prompts were the most assessed genres in state writing assessments, informative, persuasive, and narrative writing were assessed in the NAEP 2007 direct writing assessments. Expository writing and informative writing were similar because they both required students to explain something. However, they were also different because expository writing directed students to explain more abstract concepts while informative writing often directed students to provide factual information about concrete objects, events, or phenomena. Genre-mastery rubrics were the most-used rubric type in state direct writing assessments. Similarly, all the NAEP's rubrics were genre-mastery rubrics.

5.4 Insights from a combined use of the two approaches

Troia & Olinghouse's (2010) coding taxonomy provided a comprehensive framework to examine writing assessments, as well as details about the components of these assessments (i.e., prompts and rubrics), while Jeffery's (2009) coding taxonomy allowed an analysis of the most dominant genre demands of prompts and the most emphasized features of rubrics. Moreover, Troia & Olinghouse's (2010) coding taxonomy examined writing constructs defined by prompts

and rubrics together, while Jeffery's (2009) coding taxonomy examined the association between prompts and rubrics.

Prompts

The descriptive genre was absent from Jeffery's (2009) genre coding scheme because this genre was not assessed in exit-level writing assessments in that study. The descriptive genre was identified by state contacts during the K-12 Writing Alignment Project's data collection; Troia and Olinghouse's (2010) coding taxonomy included the descriptive genre as one of the purposes. In the K-12 Writing Alignment Project, the genre coding of the prompts was based on states' identification of the prompts' genres if given. For example, if a state identified one of its prompts as expository, then the prompt was coded as expository. As a result, though there were informative and analytical genres in Troia and Olinghouse's (2010) coding taxonomy, few prompts were coded informative or analytical in the K-12 Writing Alignment Project study because these prompts were often identified by states as expository or writing in response to literature. In this study, the genre coding of prompts was determined based on the prompts. When there was ambiguity in prompts, states' identification was taken into consideration. Some responses to literature could be categorized as narrative, expository, or informative, while others invited students to analyze literary elements in the provided literature and were therefore coded as analytic. In the preliminary analysis of state writing prompts, one prompt was identified by its state as summary. However, because summary only appeared once among 76 prompts and was used to provide information about an object or event, it was also categorized as informative.

Rubrics

Table 2 shows those states with genre-mastery rubrics and/or with rubrics containing genre-specific components. According to Jeffery's (2009) criteria coding scheme, NAEP's and

eleven states' writing rubrics were genre-mastery rubrics. However, among these eleven states, five states' writing rubrics were not considered to contain genre-specific components according to Troia and Olinghouse's (2010) coding taxonomy. This occurred because these states' writing rubrics prioritized the assessment of genre and framed other evaluation criteria under it but did not refer to specific genre elements. Also, according to Troia and Olinghouse's (2010) coding taxonomy, ten states' writing rubrics contained genre-specific components. However, among them, four states' writing rubrics were not considered genre-mastery rubrics. Again this was reasonable because though these rubrics contained genre-specific components, the overall orientation or emphasis of these rubrics was not focused on genre mastery. For example, specific genre components might be referred to in rubrics for the purpose of emphasizing the importance of being "effective" with audience. Only NAEP's and six states' (Alabama, California, Illinois, Indiana, New York, West Virginia) writing rubrics were both genre-mastery rubrics and contained genre-specific components.

Table 2 States with Genre-mastery Rubrics and/or States with Rubrics Containing Genre-specific Components

States whose rubrics contained genre-specific components	States whose rubrics were genre-mastery rubrics	States whose rubrics both were genre-mastery rubrics and contained genre-specific components
Alabama	Alabama	Alabama
California	California	California
Illinois	Idaho	Illinois
Indiana	Illinois	Indiana
Kansas	Indiana	New York
Missouri	Kentucky	West Virginia
New York	Rhode Island	
Nevada	Vermont	
Wisconsin	New York	
West Virginia	Virginia	
	West Virginia	

In this way, only these six states' writing assessments placed similar levels of emphasis on genre as NAEP's writing assessments, though the genres they assessed were different from those elicited by the NAEP.

Prompts and Rubrics Associations

For these six states with rubrics that were both genre-mastery rubrics and contained genre-specific components, the following Table 3 shows the genres assessed with these rubrics as well as those genres NAEP assessed.

Table 3 Genre Assessed in States with both Genre-mastery Rubrics and Rubrics Containing Genre-Specific Components

State/NAEP	Genres Assessed
Alabama	Descriptive, Expository, Narrative, Persuasive
California	Narrative, Persuasive, Analytical, Informative
Illinois	Narrative, Persuasive
Indiana	Narrative, Persuasive, Analytical
New York	Analytical, Expository
West Virginia	Descriptive, Persuasive, Narrative, Expository
NAEP	Narrative, Informative, Persuasive

Only California assessed all the genres that NAEP assessed with a similar level of emphasis on the genre demands. However, California also assessed the analytical genre, which NAEP did not.

In summary, a combined use of Troia & Olinghouse's (2010) coding taxonomy and Jeffery's (2009) coding scheme made it possible to examine the genres assessed particularly in middle school writing assessments, as well as to differentiate similar genres such as persuasive and argumentative, and expository and informative. Use of both also allowed a close look at levels of emphasis on genre demands in state and NAEP writing assessments.

6. Discussion

6.1 Prevalent Writing Assessment Practices

The results of this study showed that only three states gave students choices for prompts, thus illustrating it was not a popular practice at least by 2007. Studies of choices in the writing assessment literature have either shown statistically non-significant results regarding students' writing quality (Chiste & O' Shea, 1988; Powers & Fowles, 1998; Jennings, Fox, Graves, & Shohamy, 1999) or mixed results (Gabrielson, Gordon, & Engelhard, 1995; Powers, Fowles, Farnum, & Gerritz, 1992). This may explain why offering a choice of prompts was not a popular practice in state writing assessments.

The results of this study showed that the writing process approach had an impact on the writing assessment, because the majority of states (26 states) directed students to plan, and more than half of the states directed students to revise and edit. However, few states provided separate planning, revision, and editing sessions. Teachers are encouraged to engage students daily in cycles of planning, translating, and reviewing and teach students to move back and forth between various aspects of the writing process as their texts develop (Graham et al., 2012). Though one can argue that assessment should not mimic the entire process, but rather reflect on-the-spot performance, if writing assessments are to measure, function as, and shape writing instructions in schools, the writing procedures in assessments should emulate the process that students are being taught to follow.

To date, it has been unclear exactly what students' writing behaviors actually are under the assessment pressures and time limits: whether students start composing immediately regardless of planning directions when there is not a separate planning session, and whether students revise at the end of their compositions or move back and forth between various aspects

of the writing process while they develop their texts. More research is needed to study students' writing assessment behaviors to provide a solid foundation for designing the testing procedures in direct writing assessments. Also, because assessments have a strong impact on what is taught in schools, if states adopt the writing process approach to text production during testing sessions, instructional practices in schools are more likely to reflect this approach. Hillocks (2002) found that teachers tend to use some stages of the writing process but not others, e.g., some teachers in Illinois, Texas, and New York only incorporated editing. He suggested "the success of the assessment in promoting better teaching of writing is dependent on the character of the assessment" (Hillocks, 2002, p.196).

Olinghouse, Santangelo, and Wilson (2012) found that only limited information about students' writing abilities across a range of skills can be generalized from students' performance on single-occasion, single-genre, holistically scored writing assessments. Chen, Niemi, Wang, Wang, and Mirocha (2007) have shown that three to five writing tasks are required to make a reliable judgment about students' writing abilities. However, the results of this study showed that only seven states gave students even two prompts. The only exception was New York, which gave students four integrated writing tasks that included responding after both listening and reading. The writing tasks from New York's assessment have shown a potential path to increase students' writing opportunities by integrating listening and reading assessments with writing assessments, although this practice has raised the question of how to distinguish students' writing abilities from other abilities.

Another possible way to increase students' writing opportunities is to use writing portfolios to supplement the direct writing assessment (Moss, 1994). Because direct writing assessments are often constrained by time and resources available, when stakes are high a

combined use of direct writing assessments and writing portfolios ensures a more accurate evaluation of students' writing abilities. Because the feasibility and cost of implementing large-scale portfolio assessments remain a challenge, Gearhart (1998) cautioned that the quality of students' portfolios reflects not only students' competence, but also depends on a range of circumstantial variables. They include "teachers' method of instruction, the nature of their assignments, peer and other resources available in the classroom, and home support" (p.50), thus making comparability an issue.

Audience specification has been an extensively researched aspect of prompt design. However, the results of these studies have been mixed (Cohen & Riel, 1989; Chesky & Hiebert, 2001). For example, Redd-Boyd and Slater (1989) observed that audience specification had no effect on scores, but influenced students' motivation and composing strategies. Perhaps because of this, the majority of states (20/27) specified an audience in their state writing prompts, and at least 30% of writing rubrics emphasized the importance of authors' consideration of audience in their compositions. However, these writing prompts incorporated a wide range of audiences including general "readers," pen pals, and students' classes, classmates, or teachers.

Chesky and Hiebert (2001) examined high school students' writing and found that there were no significant differences in the length or quality of students' writing as a function of peers or teachers as a specified audience. Cohen and Riel (1989) compared seventh-grade students' writings on the same topic to peers in other countries and those to their teachers. They found that the quality of students' texts written for their peers was higher than those intended for their teachers, and suggested that contextualization could lead to improvements in the quality of students' classroom writing. However, contextualization of students' writing in direct writing assessments has remained challenging because the audiences are often just the raters. Some

states have tried to construct semi-authentic scenarios for students' writing; for example, two states situated their writing tasks within disciplinary contexts without relying heavily on disciplinary content knowledge, thus illuminating a way to construct a semi-authentic scenario in a setting with which students would be familiar.

In summary, state writing assessments have managed to incorporate extensively researched aspects, but often such incorporations remain only partial. Most state writing assessments only directed students to plan and draft, with less emphasis on revision; most states directed students' writing towards an audience, but contextualization of students' writing still remained a challenge. A few states gave students more than one prompt, but even the second-most-common option of two prompts is not enough to be able to make a generalization about students' global writing abilities. Possible reasons for this partial incorporation dilemma are that a) assessment programs have limited resources, b) the nature of standardized assessments restricts the contextualization of tests to ensure comparability, or c) the understanding of students' assessment behaviors, especially in terms of their interaction with test items, is insufficient. More research is needed on students' assessment behaviors and different methods of assessing students' writing abilities (e.g., integrated writing tasks).

An emphasis on organization, content, and details was a feature for almost all writing rubrics; word choice, sentence fluency, style, and grammar, including sentence construction, were also highly prized aspects of students' papers. General conventions, such as capitalization, punctuation, and spelling, were also assessed by the majority of states. This shows that, regardless the rubric types, these aspects are considered necessary for demonstrating writing proficiency by most states. Only ten states included genre-specific components in their rubrics; persuasive texts' components are most often specified compared with other genres. While

expository is the most assessed genre (16 states), only four states have specified expository texts' components in their rubrics. Genre demands in state writing assessments will be discussed in the next section.

By 2007, only West Virginia had online writing sessions with their state direct writing assessments. However, aligned with the CCSS, the new K-12 assessments developed by the SBAC and the PARCC will be administered via computer. Computer technology has entered most classrooms. In 2009, around 97% of teachers in U.S. public schools had computers in the classroom. The ratio of students to computers in the classroom every day was 5.3 to 1. About 40% of teachers reported that they or their students often used computers in the classroom during instructional time (U.S. Department of Education, 2010). It is possible that many students are now used to composing using computers. However, if the former state writing assessments were taken with paper and pencils, it is important that students are well prepared for the transition.

6.2 Genre Demands in Direct Writing Assessments

The results of this study showed that the most popular prompt genre in middle school assessments was expository, followed by persuasive, narrative, informative, analytic, argumentative, and finally descriptive. Jeffery's (2009) analysis of high school exit-level prompts indicated that the most popular genre was persuasive, followed by argumentative, narrative, explanatory, informative, and analytic. Persuasive and argumentative genres comprised over 60% of all the prompts (Jeffery, 2009). Therefore, the transition from middle school to high school writing assessments signifies an emphasis shift from expository compositions to persuasive and argumentative compositions. This makes sense because argumentative compositions are more abstract and place more cognitive demands on students (Crowhurst, 1988); thus, it might be most suitable for assessments of high school students.

Meanwhile, informative prompts have appeared infrequently both in this study and Jeffery's (2009) study. Given that informative prompts often require students to provide factual information about objects or events and place less cognitive demands on students than even expository prompts (Jeffery, 2009), it might be a genre most suitable for students at grades lower than middle school, unless specified by states' standards. These findings suggest that to ensure a continuum of students' learning and mastery of these genres, it is important that students are provided more opportunities to practice argumentative writing in high school; given that informative and descriptive genres are less emphasized in middle school and exit-level writing assessments, it is important that students are provided more opportunities to master these genres in lower grades.

The results of this study showed that half of the rubrics were genre-mastery rubrics. There were few rubrics that emphasized creativity and critical thinking, which is in accordance with what Jeffery (2009) found with the exit-level writing rubrics. Moreover, the expressivist rubrics, though appearing only two times, corresponded with narrative genres, and the cognitive rubrics corresponded with persuasive prompts, showing a consistency with Ivancic's (2004) framework. Different from Jeffery's (2009) finding that rhetorical rubrics were used with all genres of exit-level prompts, this study found that genre-mastery rubrics were used with all genres, while rhetorical rubrics did not correspond with descriptive prompts. The number of states that used genre-mastery rubrics was about the same as the number of states that used rhetorical rubrics. In a way, this finding confirms the assertion that "the appropriateness of language to purpose is most often prioritized in assessing writing regardless of the task type" (Jeffery, 2009, p.14). Meanwhile, the large number of genre-mastery rubrics suggests that states have started to place more genre-mastery expectations on students. However, as discussed

earlier, only ten states included genre-specific components in their rubrics and only four states included components of the most popular genre, expository texts; as a result, only six states had genre-mastery rubrics that contained genre-specific components. This finding suggests that the genre evaluation criteria that states place on students' writing are either vague or not fully utilized to assess students' genre mastery.

6.3 State and National Alignment

State writing assessments and NAEP seem to align in their adoption of the writing process approach, their attention to audience and students' topical knowledge, their accommodations through procedure facilitators, and their inclusion of organization, structure, content, details, sentence fluency, and semantic aspects as well as general conventions such as punctuation, spelling, and grammar in their assessment criteria.

However, NAEP's writing assessment differs from many states' by having explicit directions for students to review their writing, giving students two timed writing tasks, making the informative genre—which was rarely assessed in state assessments—one of the three genres assessed, and including genre-specific components in their writing rubrics. The fact that all of NAEP's writing rubrics are genre-mastery rubrics with genre-specific components can be considered one of the biggest differences from most of the state writing assessments.

Thus, when state and national writing assessment results are compared, these two assessments differ in the genres they assess, the amount of time and number of tasks they give to students, and the level and specificity of genre demands they emphasize in their evaluation criteria. These differences are observed in this study. When there is a discrepancy between state and national assessment results, can these differences explain some of the discrepancy? Research

with variables that can quantify these differences and model the relationship between these differences and writing assessment results will help answer this question.

7. Implications

More research needs to be done on the interaction between assessment procedures and students' assessment behaviors and performances. For example, further research could examine whether it increases the validity of writing assessments by incorporating explicit directions for different stages of the writing process and providing brochures with tips about planning, drafting, revising and editing.

Under the allowance of time and resources, more writing opportunities should be provided to students during writing assessments so that their writing abilities can be evaluated more accurately. When this is not possible, states should be more explicit about the interpretation of their writing assessments, so that students' performances and results reflect the actual genre assessed and specific measures used (Olinghouse et al., 2012). When states intend to evaluate students' genre-mastery skills, it is helpful to include genre-specific components in their rubrics so that their expectations are made explicit to students, raters, and educators. These recommendations are also applicable to the new K-12 assessments developed by the SBAC and the PARCC.

Students taking the NAEP 2007 were expected to write for three purposes—narrative, informative, and persuasive. It is not clear whether informative writing encompassed expository writing or referred to expository writing in NAEP. However, this study shows that informative writing has rarely been assessed in state writing assessments, while expository writing has been widely assessed in middle school. It is recommended that NAEP clarify and elaborate the categories of persuasive, informative, and narrative in its assessments.

Applebee (2005) suggested that such an attempt for clarification and elaboration has already taken place with the NAEP 2011 writing assessments. For example, the NAEP 2007 writing framework generally suggested that “students should write for a variety of purposes— narrative, informative, and persuasive” (National Assessment Governing Board, 2007, p.11); while the NAEP 2011 writing framework stated that it will “assess the ability: 1. to persuade, in order to change the reader’s point of view or affect the reader’s action; 2. to explain, in order to expand the reader’s understanding; 3. to convey experience, real or imagined” (National Assessment Governing Board, 2010, p.21). Further, the framework explicitly listed how “to explain” looks like for different grade levels:

On the NAEP Writing Assessment, tasks designed to assess students’ ability to write to *explain* at grade 4 might call for a basic explanation of personal knowledge or an explanation of a sequence of pictures and/or steps provided in the task. Grade 8 tasks may ask students to analyze a process or write a response that compares similarities and differences between two events or ideas. Grade 12 tasks may focus on asking students to identify the causes of a problem or define a concept. (p.37)

It is clear that “to explain” in the new framework encompasses both informative writing and expository writing. The framework places more emphasis on informative writing in grade 4, and more on expository writing in grades 8 and 12.

More research is needed to investigate different methods of writing assessments, such as using integrated writing tasks, and study students’ assessment behaviors, such as their interactions with writing prompts and instructions.

8. Limitations

This study only analyzed seventh and eighth grade direct writing prompts. Grade-level expectations for writing performance change from the elementary grades to the high school grades; however, this study could not examine those changes without also examining the elementary grades and the high school grades. Future studies should investigate writing expectations from elementary grades to high school grades because such studies will highlight the changes and help tailor the expectations to appropriate grade level. Indirect writing assessment items also contribute to states' definitions of the writing construct; however, they are beyond the scope of this study.

Because there was no NAEP data available for five states, thirteen states and the District of Columbia chose not to participate the study, and six states did not have 7th grade and 8th grade writing standards and assessments available for the period 2001-2006, only 27 states' direct writing assessments were included in this analysis. Therefore, the writing constructs examined in this study and the comparison between states and NAEP assessments were limited to these 27 states.

The sample of the NAEP examined was limited to publically-released data comprised of three prompts and three rubrics. These prompts represent the genres assessed in the NAEP; but it is possible that they do not showcase all the genres assessed. For example, the informative prompt was coded to assess informative writing in this study; however, it is possible that there were informative prompts that actually assessed expository writing. Without examining the other writing prompts in the NAEP, it is hard to determine how different those writing prompts are from the released sample. Therefore, the writing construct assessed in the NAEP might not be completely captured by this study since the analysis was based only on the publically-released sample.

CHAPTER 2: Predicting Students' Writing Performances on the NAEP from Assessment Variations

1. Introduction

Persistent discrepancies are identified between state and national writing assessment results (Lee, Grigg, & Donahue, 2007; Salahu-Din, Persky, & Miller, 2008). State mandated assessments often report high proficiency levels, but the results of the National Assessment of Educational Progress (NAEP) indicate low proficiency levels. The variation between state and national assessments' definitions of the writing construct and measurements of writing proficiency is one possible explanation of this gap. However, little is known about how these assessments actually vary. Even less is known about how this variation predicts students' performance on the NAEP. One factor contributing to the differences in students' performances between state tests and the NAEP is the differing writing constructs that the state and NAEP tests assess; as a result, students' performances on the NAEP does not only indicate students' writing abilities, it also reflects how well students are prepared for the type of assessments the NAEP utilizes.

Research has shown that high-stakes assessments (i.e., state-mandated assessments) have an impact on classroom instruction (Hillocks, 2002; Moss, 1994). When the content and format of state-mandated assessments are comparable to the national assessment, students are indirectly prepared for the NAEP. However, whether students actually achieve higher scores on the NAEP when their state assessments are more similar to NAEP, and lower scores when their state assessments are less similar, is unknown. In other words, whether this variation between state and national writing assessments predicts students' performance on the NAEP remains unexamined. This study aims to fill this gap in the research.

To examine the impact of the variations between state and national writing assessments on students' performances, it is important to control those variables found in existing research that tend to have an influence on those performances. Students' demographic backgrounds, their writing attitudes and motivations, and their previous experiences with writing have a significant influence on their writing development and performances, which will be discussed next.

Gabrielson, Gordon, and Englehard (1999) studied the effect on writing quality of offering students a choice of writing tasks. To do this, they examined persuasive essay writing tasks administered to 34,200 grade 11 students in the 1993 Georgia state writing assessments. These tasks were organized into packets of single tasks for groups in the assigned-task condition and packets of pairs of tasks for groups in the choice-of-task condition. They found that while the choice condition had no substantive effect, gender, race, and the specific writing tasks given had a significant impact on the writing quality in both the multivariate analysis of variance and the univariate analysis. Female students' essays received higher scores than those of male students. White students' essays received higher scores than those of Black students. The writing task variable had significant interaction with gender and race. Female students were more likely to perform better than male students on some writing tasks rather than others; White students were also likely to perform better than Black students on certain writing tasks. Because the purpose of the study was to investigate the effect on students' writing quality of offering students a choice of writing tasks, and also for test security reasons, the fifteen tasks were not revealed and there was no further illustration of what different characteristics these writing tasks possessed in the study.

Ball's (1999) case study of a sample text written by an African-American high school male sophomore student revealed influence of African-American Vernacular English (AAVE) on

the student's grammatical and vocabulary choices, spelling variations, and discourse style and expressions in his writing. Kanaris (1999) examined writings about a recent excursion by 29 girls and 25 boys in grades 3-4, and found that the girls tended to write longer and more complex texts, with a greater variety of verbs and adjectives and more description and elaboration; the boys were more likely than the girls to use the first person singular pronoun, and less likely to take themselves away from the center of the action.

Research also suggests that students' English language proficiency plays an important role in their writing performances. Research such as Silva's (1993) has examined the nature of English as the First Language (L1) writing and English as a Second Language (ESL/L2) writing, and found that L2 writing is distinct from L1 writing by appearing to be less fluent, less accurate, and less effective with L1 readers than L1 writing. L2 writers' texts are simpler in structure, and include a greater number of shorter T-units and more coordination, as well as a smaller number of longer clauses, less subordination, fewer noun modifications, and minimal passive sentence constructions. They also include more conjunctives and fewer lexical ties, as well as have less lexical control, variety, and sophistication overall (Silva, 1993). ESL students' English proficiency levels greatly influence their writing abilities, so that students with different proficiency levels include a variety of lexical and syntactic features in their writing: number of words, specific lexical classes, complementation, prepositional phrases, synonymy/antonymy, nominal forms, stative forms, impersonal pronouns, passives, relative clauses, deictic reference, definite article reference, coherence features, participial phrases, negation, present tense, adverbials, and 1st/2nd person pronouns (Ferris, 1994).

It is also common for students with special needs to experience substantial difficulty with writing (Graham & Harris, 2005). Gilliam and Johnson (1992) compared the story telling and

writing performance of 10 students with language/learning impairment (LLI) between the ages of 9 and 12 years and three groups of 30 normally-achieving children matched for chronological age, spoken language, and reading abilities using a three-dimensional language analysis system. They found that LLI students produced more grammatically unacceptable complex T-units, especially in their written narratives, than students from the three matched groups.

Newcomer and Barenbaum (1991) reviewed research investigating the written composing ability of children with learning disabilities and concluded that these children struggled with most aspects of mechanics/syntax/fluency, and as a result were less skilled than other children in writing stories and expository compositions. Resta and Eliot (1994) compared the performance of 32 boys between the ages of 8 and 13 years belonging to three groups—those with attention deficits and hyperactivity (ADD+H), those with attention deficits without hyperactivity (ADD-H), and those without attention deficits—on the Written Language Assessment, and found that both ADD+H and ADD-H children had poorer performance on most of the written language subtests than children without attention deficits. They therefore concluded that children with attention deficits possessed significant limitations in their writing and composition.

Students' attitudes and motivation are yet more factors that have a significant impact on their writing development (Mavrogenes & Bezrucko, 1999) and writing achievements (Graham, Berninger, & Fan, 2007). Moreover, students' positive beliefs and attitudes about writing determine their motivations to write (Bruning & Horn, 2000), while difficulties created by lack of knowledge and complexity of writing tasks can adversely influence their motivation levels (Zimmerman & Risemberg, 1997). Meanwhile, motivation is not a unitary construct; rather, it is "a domain-specific and contextually situated dynamic characteristic of learners" (Troia, Shankland, & Wolbers, 2012, p.6). In other words, a student's motivation to write is independent

of their motivation to read, and changes according to the performance contexts. Therefore, performance contexts affect motivation, while in turn “positive motivation is associated with strategic behavior, task persistence, and academic achievement” (Troia et al., 2012, p.6).

Students’ perceptions of prompt difficulties are related to both students’ knowledge about the writing topic (Powers & Fowles, 1998) and prompts’ characteristics such as question type (e.g., compare/contrast, descriptive/narrative, argumentative) (Polio & Grew, 1996; Way, Joiner, & Seaman, 2000) and topic specificity (Chiste & O’Shea, 1988; Polio & Grew, 1996). However, previous research has failed to detect a strong relationship between students’ perception of prompt difficulty and their writing performance (Powers & Fowles, 1998).

Students’ writing activities inside classrooms tend to have a positive effect on students’ writing composition. The meta-analysis conducted by Graham, Kiuahara, McKeown, and Harris (2012) suggested that “four of the five studies that examined the effects of increasing how much students in grades 2 to 6 wrote produced positive effects” (p. 42). The only study that had a negative effect involved English language learners (Gomez & Gomez, 1986). Thus, while students’ writing activities inside classrooms are related to their writing performances, their backgrounds also need to be considered.

Students’ experiences with writing also play a significant role in their writing achievements. In the NAEP 2007 writing assessments, students’ experiences with writing were surveyed through questions asking about the feedback they received from teachers and the use of computers in their daily writing. Research has shown that teachers’ and peers’ feedback tend to improve students’ writing quality and productivity (Rogers & Graham, 2008), while a lack of immediate feedback can negatively impact students’ motivation (Zimmerman & Risemberg, 1997). Meanwhile, students’ use of technology is likely to increase their compositions’ length,

their adherence to conventions, and the frequency of revisions; it also cultivates students' positive attitudes towards writing and improves their writing quality (Bangert-Drowns, 1993; Goldberg, Russel, & Cook, 2003).

In summary, students' writing performance on assessments is closely related to their backgrounds and prior writing experiences. Therefore, a study of the relationships between state and NAEP writing assessment variations and students' NAEP writing performances necessitates controlling for the following variables relating to students' individual characteristics: students' attitudes towards writing and perceptions of prompt difficulty, their demographic backgrounds (i.e., gender, race/ethnicity, English language proficiency, social economic status, and disability status), their writing activities inside classrooms, and their experiences with writing.

2. Research Questions

Through multi-level modeling analysis, this study explores state and NAEP assessment data to answer the following research question: Do students from states that use writing assessments with a higher degree of similarity to NAEP writing assessment features, measured by the Euclidean distance with the multi-dimensional writing construct, perform better on the NAEP, controlling for students' attitudes towards writing and perceptions of prompt difficulty, their demographic backgrounds, their writing activities inside classrooms, and their experiences with writing?

3. Method

3.1 State and NAEP Direct Writing Assessments

This study was conducted upon data from a prior IES-funded study—the K-12 Writing Alignment Project (Troia & Olinghouse, 2010-2014). In the K-12 Writing Alignment Project, states' Department of Education websites were first used to locate appropriate assessment

personnel. Documents were then requested through email inquiries and phone calls. Because the K-12 Writing Alignment Project examined the alignment between state writing standards and assessments prior to the adoption of the CCSS, and the NAEP 2007 data contained state-level writing data allowing state-level modeling, the NAEP 2007 data was used. State direct writing assessments were gathered mainly from between 2001 and 2006, to allow comparisons to be made with the NAEP 2007. The number and dates of the major revisions between 2001 and 2006 were identified for each state to ensure the collection of its representative state writing assessments. From each time span between major revisions, a representative writing prompt, its rubric, and the administrative manual for each genre in each grade being assessed were collected.

In this study, 78 prompts and 35 rubrics from 27 states² were analyzed (see Appendix C for details). NAEP data was not available for Alaska, Nebraska, Oregon, and South Dakota for the time period selected. State writing standards or writing assessments were not available for Connecticut, Iowa, Pennsylvania, Montana and New Mexico between 2001 and 2006. There was no writing assessment for 7th grade and 8th grade in Ohio during the period 2001-2006. Consequently, these states' direct writing assessments were not analyzed in this study.

The state direct writing assessment documents were compiled to be used for coding. The compiled files include the following components: verbal directions from administration manuals for direct writing assessments, actual prompts, supporting materials provided (e.g., dictionary, writer's checklist), sessions arranged for writing tests, time given, page limits, and whether (and what kinds of) technology was used. The number of responses expected from students each year determined the number of compiled files for each state. For example, if students took only one prompt with rotated genres each year, the prompts from the rotated genres were all compiled into

² The following chose not to participate in the study: Colorado, Delaware, the District of Columbia, Georgia, Hawaii, Maryland, Minnesota, Mississippi, New Hampshire, New Jersey, North Dakota, South Carolina, Utah, and Wyoming.

a single document to represent the scope of genres assessed and the number of prompts (i.e., one prompt in this case) assessed in a test administration.

This study included three publically-released NAEP 2007 writing prompts from eighth grade (i.e., a narrative prompt, an informative prompt, and a persuasive prompt), scoring guide, and writing framework. These three writing prompts were released to represent the genres the NAEP assessed; other writing prompts were not released due to test security and possible future use.

3.2 Coding taxonomy

This study used Troia and Olinghouse's (2010) seven-stranded coding taxonomy. The coding taxonomy was derived from several theoretical frameworks—Hayes' cognitive model of writing (Flower & Hayes, 1981; Hayes, 1996), socio-cultural theory (Prior, 2006), genre theories (Dean, 2008), linguistic models of writing (Faigley & Witte, 1981), and motivation theories of writing (Troia, Shankland, & Wolbers, 2012)—to assure a broad representation of current thinking about writing development, instruction, and assessment. The coding taxonomy consisted of seven strands: (1) writing processes, (2) context, (3) purposes, (4) components, (5) conventions, (6) metacognition and knowledge, and (7) motivation. In writing assessments, the indicators in the seventh strand—motivation, which refers to personal attributes within the writer such as general motivation, goals, attitudes, beliefs and efforts—did not apply, because states rarely administered assessment documents such as surveys alongside the writing assessments to measure these personal attributes. The indicators found within those six strands in the coding taxonomy covered: all stages of the writing process; specific composition strategies; circumstantial influences outside the writer; a variety of communicative intentions accomplished through different genres; features, forms, elements, and characteristics of text; the mechanics of

producing text; and knowledge resources within the writer that drive writing activity and writing development.

Meanwhile, Jeffery's (2009) genre and criteria coding schemes, derived from high school exit writing prompts, were used to supplement Troia and Olinghouse's (2010) coding framework. A preliminary frequency analysis of state writing prompts' genres coded with Troia & Olinghouse's (2010) coding taxonomy indicated that only a few genres were assessed in state writing assessments—expository, descriptive, persuasive, response-to-literature, descriptive, narrative, and summary. As a result, the third strand of Troia and Olinghouse's (2010) coding taxonomy was replaced by a seven-category genre coding scheme—descriptive, persuasive, expository, argumentative, informative, narrative, and analytic.

In this coding scheme, persuasive prompts and argumentative prompts were differentiated to represent common and subtle differences between these two genres. Argumentative prompts differ from persuasive prompts by calling abstractly for “support” of a “position,” and by not designating a target audience. In contrast, persuasive prompts require students to convince an identified audience to act on a specific issue. Moreover, persuasive prompts are unlike argumentative prompts because they invite students to take a one-sided perspective on an issue, while argumentative prompts often expect students to consider multiple perspectives on an issue. A new strand evaluating rubrics' most dominant features was created by using Jeffery's (2009) criteria coding scheme. Rubrics in the sample were categorized into one of the five criteria coding schemes: rhetorical, genre-mastery, formal, cognitive, and expressivist.

The result of this was a coding taxonomy containing seven strands and 90 indicators. For each compiled document, all the indicators could only be coded 0 or 1 (absent or present). The

exception was that indicators for planning, drafting, and revising in the first strand could have up to three points each to accommodate information, including whether students were directed to plan, draft, and revise, as well as the time and pages or writing space given for each step. For example, Kansas directed eighth grade students to plan, draft, and revise and gave students the time and space to do each step, thus, it received a maximum score of nine for these three indicators: plan, draft, and revise. Louisiana directed eighth grade students to draft and gave students the time and space to do so, but did not direct students to plan and revise, nor did it give students time or space for these activities, thus it received a minimum score of three for only one indicator—draft. When there were multiple compiled assessment documents for either seventh grade or eighth grade, a sum score of these coded compiled assessment documents was used for each indicator for a state. When a state had both 7th and 8th grade writing assessments, an average score of the 7th grade and 8th grade coded compiled assessment documents was used for each indicator for the state.

3.3 Coding Procedure

In the K-12 Writing Alignment Project, the first (writing processes), second (context), third (purposes), and sixth (metacognition and knowledge) strands from Troia and Olinghouse's (2010) coding taxonomy were used to code state and NAEP writing prompts by three raters, because writing processes and writing contexts were often specified in the verbal directions of test administrations, and writing purposes and writing knowledge were often specified in writing prompts. The first rater was paired with either the second rater or the third rater to code each compiled assessment document. The first rater and the second rater reached an inter-rater reliability of .97; the first rater and the third rater reached an inter-rater reliability of .95. Because writing components and writing conventions were often specified in the scoring rubrics, the

fourth (components) and fifth (conventions) strands from Troia and Olinghouse's (2010) coding taxonomy were used to code state and NAEP writing rubrics by two separate raters. They reached an inter-rater reliability of .95 and resolved differences through discussion.

In this study, two raters coded state and NAEP writing prompts with the seven-category genre coding scheme adapted from the third strand (purpose) of Troia and Olinghouse's (2010) coding taxonomy and Jeffery's (2009) genre coding scheme. These raters also coded state and NAEP writing rubrics with Jeffery's (2009) criteria coding scheme. The inter-rater reliability was .93 for prompt coding and .86 for rubrics coding. Differences were resolved through discussion.

Once the coding of prompts and rubrics was finished, each state's writing assessments were characterized by the 90 indicators under the seven strands, including Jeffery's (2009) criteria coding scheme and the six strands from Troia and Olinghouse's (2010) coding taxonomy. These indicators were used to calculate the distance between state assessments and the NAEP in the next step.

3.4 Distance between State Assessments and the NAEP

Because state and NAEP direct writing assessments were coded with the above taxonomy, the writing constructs in these assessments were examined in multiple dimensions. As a pure mathematical concept, Euclidean distance measures the distance between two objects in Euclidean n-spaces. More specifically, state X's writing construct could be defined by the 90 indicators in the coding taxonomy as $(x_1, x_2, \dots, x_{90})$, and NAEP Y's writing construct could be defined by the 90 indicators in the coding taxonomy as $(y_1, y_2, \dots, y_{90})$. Euclidean distance can be calculated as:

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_{90} - y_{90})^2} = \sqrt{\sum_{i=1}^{90} (x_i - y_i)^2}$$

where $d(X, Y)$ indicates the amount of difference between state and NAEP direct writing assessments. A small $d(X, Y)$ means that state and NAEP direct writing assessments are similar. A large $d(X, Y)$ means that state and NAEP direct writing assessments are different. The Euclidean distance is unstandardized because most of the indicators are coded as 0 or 1; thus, it is less likely that some indicators carry much more weights than other indicators and dominate the distance.

Because the number of compiled documents are the same number of prompts that students should write in response to in state writing assessments, states with more compiled documents were given more codes as each compiled document was coded with the coding taxonomy once. NAEP gave students two writing prompts, thus, states which gave students two prompts had writing assessments more similar to NAEP's based on the Euclidean distance. The value of $d(X, Y)$ for each state can be found in the last column in Table 4 below. The value is in the range of 7.48 and 15.2, with a mean of 9.97 and a standard deviation of 1.53.

3.5 NAEP Sample

A total of 139,910 eighth grade students participated in the NAEP 2007 writing assessments. From this total, 85,437 students from the 27 states where direct assessments were gathered were selected. When weighted, this represented a population of 2,415,129 (see Table 9 in Appendix A for descriptive statistics). Because this sample was missing some data, the sample used in Hierarchical Linear Modeling (HLM) analysis was reduced. There were 73,754 eighth grade students in the HLM sample (see Table 4 below for descriptive statistics). The demographics of the HLM sample and the 27-state NAEP sample were very similar (see Table 10 in Appendix A for comparisons between the 27-state NAEP sample and HLM sample).

3.6 Students' NAEP Composition Performance

Table 4 Sample Sizes, Achievement, and Student Demographics, 27 State Grade 8 HLM Sample

State	N	Weighted N	Mean Student Achievement	SE(Mean)	% Black	% Hispanics	% Asian	% American Indian	% Female	% LEP	% With IEPs	% Free/ reduced-price lunch	Distance between state and NAEP
Alabama	2360	48406	150.877	1.335	33.2%	1.9%	0.8%	0.4%	51.1%	1.1%	9.2%	47.6%	12.845
Arizona	2199	57486	150.436	1.426	5.4%	37.9%	2.6%	6.5%	49.8%	8.7%	7.1%	42.6%	10.428
Arkansas	2081	29300	152.304	1.21	22.3%	7.1%	1.0%	0.3%	48.5%	3.7%	11.2%	51.6%	9.138
California	6361	366387	151.844	0.997	6.4%	46.4%	12.9%	1.3%	50.3%	18.4%	6.6%	47.4%	11.314
Florida	3302	157639	160.332	1.4	21.7%	23.2%	2.4%	0.3%	49.9%	4.7%	11.5%	41.9%	8.124
Idaho	2460	17890	155.447	1.079	1.0%	12.8%	1.5%	1.6%	48.9%	5.1%	8.0%	38.8%	9.327
Illinois	3337	128181	162.029	1.508	17.7%	17.7%	4.6%	0.1%	49.4%	2.5%	11.8%	38.7%	10.954
Indiana	2309	67987	156.499	1.247	11.5%	5.9%	1.2%	0.2%	50.4%	2.3%	10.6%	33.5%	7.483
Kansas	2380	28803	157.12	1.385	7.7%	11.7%	1.9%	1.5%	50.1%	3.7%	10.3%	35.7%	9.274
Kentucky	2251	38972	152.067	1.376	9.9%	1.6%	1.0%	0.0%	50.9%	0.9%	8.0%	46.5%	9.314
Louisiana	2059	41170	148.265	1.24	41.7%	2.2%	1.2%	1.0%	49.1%	0.8%	11.2%	59.1%	9.925
Maine	2243	12942	162.335	1.106	1.5%	0.7%	1.4%	0.2%	49.9%	1.5%	14.1%	33.0%	9.274
Massachusetts	2944	57051	168.863	1.524	8.4%	9.7%	5.4%	0.2%	48.9%	3.1%	13.5%	25.4%	10.770
Michigan	2195	100740	153.185	1.286	17.1%	2.6%	2.4%	0.9%	50.3%	1.5%	10.8%	31.3%	9.925
Missouri	2495	62339	154.508	1.126	17.6%	2.7%	1.6%	0.1%	50.2%	1.6%	10.6%	36.1%	9.381
Nevada	2136	22842	146.746	1.063	9.4%	33.3%	8.8%	1.6%	51.0%	8.4%	9.2%	36.7%	8.944
New York	3050	170662	157.207	1.273	16.9%	17.3%	6.8%	0.3%	50.9%	3.6%	13.3%	46.1%	15.199
North Carolina	3452	86993	154.978	1.266	28.0%	6.9%	2.4%	1.3%	50.2%	3.7%	13.7%	42.5%	9.220
Oklahoma	2233	36291	153.877	1.161	8.9%	8.2%	2.2%	20.0%	50.0%	3.2%	12.6%	47.5%	8.832
Rhode Island	2248	10034	156.225	0.832	7.6%	16.6%	3.0%	0.5%	50.4%	2.2%	16.0%	30.5%	10.050
Tennessee	2436	64043	157.487	1.398	23.9%	4.7%	1.5%	0.0%	50.7%	1.7%	8.2%	43.9%	10.440
Texas	5951	246259	153.128	1.16	15.3%	43.3%	3.1%	0.2%	49.9%	5.7%	6.6%	49.3%	9.899
Vermont	1744	5956	162.968	1.174	1.6%	1.0%	1.6%	0.4%	47.9%	2.3%	16.2%	26.7%	10.050
Virginia	2301	74430	157.838	1.257	27.3%	5.6%	4.6%	0.2%	49.7%	2.9%	9.7%	26.7%	8.944
Washington	2418	62506	160.472	1.453	5.3%	12.7%	9.4%	2.3%	48.9%	4.4%	8.0%	33.4%	9.000
West Virginia	2537	19100	147.663	1.082	4.8%	0.9%	0.7%	0.2%	50.8%	0.7%	13.7%	46.7%	8.307
Wisconsin	2272	52385	159.204	1.435	8.4%	6.4%	3.3%	1.2%	49.3%	3.4%	11.7%	28.9%	9.539
Total	73754	2066794											

Note. The means and percentages reported are for the samples weighted to represent U.S. students.

Eighth grade students' writing performances on 20 NAEP writing prompts were used for this analysis. In the NAEP database, each student wrote in response to two prompts; five plausible values were generated from students' conditional distributions. These five plausible values were used as the outcome variable—students' NAEP performance. The NAEP 2007 writing assessment was designed with six overarching objectives. Students were expected to write (a) for three purposes (i.e., narrative, informative, and persuasive); (b) on a variety of tasks and for diverse audiences; (c) from a variety of stimulus materials and within various time constraints; (d) with a process of generating, revising, and editing; (e) with effective organization, details for elaborating their ideas, and appropriate conventions of written English; and (f) to communicate (National Assessment Governing Board, 2007). All students' writing products were first evaluated by NAEP for legibility, staying on task, and ratable. If they passed the above evaluation, they were then scored based on a six-point rubric, where 1 was Inappropriate, 2 was Insufficient, 3 was Uneven, 4 was Sufficient, 5 was Skillful, and 6 was Excellent. If they did not pass the initial evaluation and thus did not receive a score, they were not included in this study.

3.7 Students' Characteristics in NAEP

The dataset used for analysis was from the NAEP 2007 eighth grade student database. Student characteristics data were gathered through student and teacher surveys. There were 34 student characteristic variables. They were categorized into six groups for the convenience of reporting results. These six groups did not suggest six factors nor should those variables be considered as indicators of these factors. Because the main purpose of this study is to investigate the effect of state-level variables on students' writing performances by controlling for the comprehensive coverage of the students' characteristics variables, all related students'

characteristics variables were included and scale reduction was not considered necessary. The six groups were employed to allow reporting variables similar in meaning to NAEP survey descriptions. First, there was students' demographic background, which consisted of students' ELL status, free/reduced lunch eligibility status, with or without Individualized Education Plans (IEP), gender, race or ethnicity, as well as location states. Second, students' attitudes towards writing were measured by whether they considered writing stories or letters as a favorite activity, and whether they found writing helpful in sharing ideas. Third were students' perceptions of the difficulty of the NAEP writing tests. Fourth, students' levels of motivation for taking the NAEP writing assessments were evaluated by measuring their perceptions of their efforts on the NAEP writing tests and the importance of success on the tests.

Fifth, students' writing activities inside classrooms included (a) the frequency and types of writing they did in school, including writing that was used to express their thoughts/ observations on their in-school writing activities, a simple summary of what they read, a report based on what they studied, an essay they analyzed, a letter/essay, a personal or imagined story, or business writing; (b) the aspects of writing they had worked on in school, including how often they brainstormed, organized papers, made changes, or worked with other students; and (c) their writing in content areas, including how often they wrote one paragraph in their English, science, social studies, history, and math class.

Sixth, students' experiences with writing consisted of (a) their computer use, i.e., whether they had used a computer from the beginning, for changes, or for the internet when writing papers for school; and (b) their teachers' expectations and feedback, such as how often teachers talked to students about their writing or asked them to write more than one draft, and whether

teachers graded students more heavily for spelling, punctuation, or grammar, paper organization, quality and creativity, and length of paper.

3.8 Structure of the Data Set and Statistical Analyses

The NAEP 2007 writing assessments used stratified multi-stage cluster sampling. Schools in the nation were grouped into strata based on their locations, sizes, percentages of minority students, student achievement levels, and area incomes. Then schools were selected randomly within each stratum, and students were selected randomly within schools. Selected schools and students were assigned weights to represent a national sample. To reduce NAEP testing time, the NAEP used “matrix sampling”—students only took a portion of the full NAEP battery of potential items. This sampling method ensured an accurate estimate of the population’s performance but resulted in large intervals of individual estimates of abilities. Instead of a single score indicating a students’ writing ability, five plausible values were drawn from the conditional distribution of student writing ability estimates based on the student’s background characteristics and the patterns of responses to the items that were administered to the student. Therefore, an analysis of NAEP achievement data required that statistical analyses be conducted for each of the five plausible values and the results synthesized (Rubin, 1987).

This study used appropriate weights and statistical procedures to address the special characteristics of the NAEP data set. Data management was mostly done using SPSS. AM statistical software is designed with procedures to handle the weighting and jackknifing needs of complex data sets such as the NAEP’s. This study used AM to calculate achievement means and standard errors as well as generate descriptive statistics of the 27-state NAEP reporting sample and the HLM sample.

Given the hierarchical organization of the NAEP data set, in which students were nested within states, a multi-level analysis was most suitable because it ensured more precise parameter estimation and allowed more accurate interpretation (Goldstein, 1987). HLM software is designed with features to use weights at level-1, level-2, or both levels to produce correct HLM estimates, as well as features to run analyses with each of the five plausible values and synthesize the results of these analyses by averaging values and correcting standard errors (Raudenbush & Bryk, 2002). This study used HLM 7.0 to create a sequence of two-level models—state level and student level—to examine the research question.

The overall weight was used at level 1—student level—because it adjusted for the unequal probability for both the student to be selected and the school that the student was enrolled in to be selected. No weight was used at level 2—state level. All binary variables such as demographic variables were uncentered; all continuous variables such as students' writing-experience variables were grand mean centered; and the state-level variable—the distance between a state's and the NAEP's writing assessments (i.e., $d(X,Y)$)—was uncentered. The uncentering of the binary variables allowed interpretations to be made about differences in performances between students in separate categories for all of the binary variables such as female and male. The grand mean centering of students' writing experience variables afforded understandings about students with average writing experience for each variable. Finally, the uncentering of the state-level variable made it possible to interpret the results of states with same writing assessments as the NAEP (i.e., no distance between state and NAEP writing assessments).

3.9 Statistical Models

To answer the research questions, this study utilized four statistical models. Similar to Lubienski & Lubienski's (2007) data analysis design, this study first ran an unconditional model, then added demographic variables, then students' writing-experience variables, and finally the state-NAEP distance variable. This procedure allowed the researcher to examine the extent of additional variance in the outcome variable that the inclusion of each group of variables explained. The total variance in students' NAEP performances was decomposed into a between-states component (state level) and a between-students component (student level).

Unconditional model (Model 1). State- and student-level variables did not enter the model. The unconditional model measures whether there was a significant difference between states' mean scores on the NAEP.

$$Y = \beta_0 + e + \varepsilon$$

where

Y is one of the students' five plausible values,

e is the random error between states, and

ε is the random error between students

When discussing the results, special attention was paid to $\text{var}(e)$, to see whether it was significant. A significant $\text{var}(e)$ means that there are significant differences among states in terms of students' performance. Therefore, the differences among states can be further modeled.

Main effect model (Model 2). Student-level demographic variables entered the model as fixed effects.

$$\text{Level 1: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$\text{Level 2: } \beta_0 = \gamma_{00} + e$$

$$\text{Combined model: } Y = \gamma_{00} + \beta_1 X_1 + \dots + \beta_k X_k + e + \varepsilon$$

where

Y is one of the students' five plausible values,

X_k is the students' demographic variables,

e is the random error between states, and

ε is the random error between students

Main effect model (Model 3). Both student-level demographic variables and writing-experience variables entered the model as fixed effects.

$$\text{Level 1: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$\text{Level 2: } \beta_0 = \gamma_{00} + e$$

$$\text{Combined model: } Y = \gamma_{00} + \beta_1 X_1 + \dots + \beta_k X_k + e + \varepsilon$$

where

Y is one of the students' five plausible values,

X_k is the students' demographic variables and writing-experience variables,

e is the random error between states, and

ε is the random error between students

Main effect model (Model 4). Both state-level variables (i.e., the distance between NAEP and state writing assessments) and student level variables (i.e., demographic variables and writing-experience variables) entered the model as fixed effects.

$$\text{Level 1: } Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$\text{Level 2: } \beta_0 = \gamma_{00} + \gamma_{01}d + e$$

$$\text{Combined model: } Y = \gamma_{00} + \gamma_{01}d + \beta_1 X_1 + \dots + \beta_k X_k + e + \varepsilon$$

where

Y is one of the students' five plausible values,

X_k is the students' demographic variables and writing-experience variables,

d is the distance between states' assessments and the NAEP,

e is the random error between states, and

ε is the random error between students

When discussing the results, special attention was paid to γ_{01} , to determine whether it was significant. A negative γ_{01} indicates that the more state assessments differ from the NAEP, the lower students' NAEP performances will be. A positive γ_{01} indicates that the more different state assessments are from the NAEP, the higher students' NAEP performances will be.

4. Results

This study utilized four hierarchical linear models to examine whether the distance between NAEP and state writing assessments can predict students' performances on NAEP. Table 11 in Appendix A shows the raw and unweighted descriptive statistics for all the variables used in the HLM analyses.

The HLM results can be found in Table 5 below. Because whether the difference between state and NAEP writing assessments can predict students' NAEP performances is the main interest of this study, standard errors are provided for the intercept and the state and NAEP difference variable. The unconditional model (model 1) showed that the average writing performance of all students was 155.5. It also showed that 54.863% of the variance was between states, and 45.137% of the variance was within states. Because the between-state variance was very significant, it made a multi-level model necessary.

Model 2 added student-level demographic variables. It showed that the student-level demographic variables were all significant. The intercept of 160.387 was the estimated mean achievement of a student who was at the level of 0 on all the binary predictors (i.e., male, White,

non-ELL, without IEPs, and not eligible for free/reduced lunch). Except Asian students, students from other minority ethnicities (i.e., Black, Hispanic, and American Indians) had an average score lower than the estimated mean achievement of a student with the above level-0 characteristics. Similarly, students who were ELLs, had IEPs, or were eligible for free/reduced-price lunch also had lower average scores. Female students had higher average scores than male students. Student-level demographics explained an additional 33.185% of the variance between states and an additional 33.151% of the variance within states. The between-state variance remained very significant.

Model 3 also included student-level demographic variables, and added student-level writing experience variables. It showed that almost all student writing experience variables were significant except the following: how often students wrote a letter or essay for school, and their perception of the importance of success on the writing test they were undertaking. The intercept of 161.692 was the estimated mean achievement of a level-0 student on all the binary predictors (i.e., male, White, non-ELL, without IEPs, and not eligible for free/reduced lunch) and at the mean of all the continuous predictors (i.e., students' writing experience variables). Students' attitudes towards writing and their perceptions of the difficulty of the NAEP writing test were positively related to their NAEP performance. More specifically, students who enjoyed writing, thought that writing helped to share ideas, and considered the NAEP writing assessment easier than other tests tended to get higher scores. However, students' perceptions of their efforts and the importance of success on the NAEP writing test were negatively related with students' NAEP performance. More specifically, students who believed that they tried harder and considered their successes on the NAEP writing assessments more important tended to get lower scores. Student-level writing-experience variables explained an additional 10.397% of the

variance between states (43.582% instead of 33.185%) and an additional 10.285% of the variance within states (43.436% instead of 33.151%). The between-state variance remained very significant.

Model 4 included both student-level demographic variables and writing experience variables, and added the variable of primary interest—the state and NAEP difference variable. It showed that when differences in students’ backgrounds and writing experiences were controlled, state and NAEP direct writing assessment differences were significant. The intercept of 163.148 was the estimated mean achievement of a level-0 student on all the binary predictors (i.e., male, White, non-ELL, without IEPs, and not eligible for free/reduced lunch), at the mean of all the continuous predictors (i.e., students’ writing experience) and from a state with same writing assessment as the NAEP (i.e., no distance between state and NAEP writing assessments). More specifically, 163.148 was the predicted mean achievement of a White, non-IEP, non-ELL, subsidized lunch-ineligible male student with average frequency of certain writing practices, average amount of feedback from teachers, and average perception of difficulty and importance of the NAEP writing assessments from a state which had the same writing assessment as NAEP.

The state and NAEP distance/difference variable was found to be statistically significant with a coefficient of -0.143 and a standard error of 0.067. With every difference of one unit between states’ writing assessments and the NAEP writing assessment, the predicted achievement of such an above-mentioned student would be a significant 0.143 points lower than the average.

All student-level demographic variables remained very significant. Almost all the student-level writing experience variables were significant, except the two that were insignificant in model 3. The variables that were positively related to students’ NAEP performances in model

Table 5 HLM Model Results

Variable	Model 1: Unconditional Model	Model 2: Student Demographics	Model 3: Student Demographics +Writing Experience	Model 4: Student Demographics+ Writing Experience +State Difference
Fixed effects				
Intercept	155.5***	160.387***	161.692***	163.148***
(S.E.)	0.144	0.197	0.189	0.693
State Level				
Distance between NAEP and State Assessments				-0.143*
(S.E.)				0.067
Student level				
Demographics				
Black		-13.436***	-13.843***	-13.832***
Hispanic		-8.951***	-8.07***	-8.016***
Asian		10.438***	7.978***	8.076***
American Indian		-11.885***	-9.234***	-9.276***
Female		18.143***	12.23***	12.229***
ELL		-25.766***	-22.208***	-22.189***
IEP		-33.929***	-30.253***	-30.238***
free/reduced lunch		-13.059***	-10.488***	-10.468***
Writing Experience in school				
Writing stories/letters is a favorite activity			1.64***	1.646***
Writing helps share ideas			2.562***	2.566***
How often teacher talk to you about writing			0.654*	0.658*
How often write thoughts/observation			0.569***	0.568***
How often write a simple summary			1.534***	1.547***
How often write a report			-0.905***	-0.897***
How often write an essay you analyze			1.97***	1.983***

Table 5 (cont'd)

How often write a letter/essay for school			-0.024	-0.05
How often write a story personal/imagine			-0.453**	-0.459**
How often write business writing			-2.55***	-2.552***
How often when writing-get brainstorm			-0.982***	-0.977***
How often when writing-organize papers			0.71***	0.698***
How often when writing-make changes			6.031***	6.034***
How often when writing-work with other students			-1.449***	-1.452***
Write paper-use computer from begin			-0.951***	-0.941***
Write paper for school-use computer for changes			3.615***	3.627***
Write paper for school-use computer for internet			1.677***	1.681***
How often write one paragraph in English class			4.204***	4.209***
How often write one paragraph in science class			0.959***	0.946***
How often write one paragraph in social studies/history class			0.926***	0.942***
How often write one paragraph in math class			-2.735***	-2.738***
How often teacher asks to write more than 1 draft			1.174***	1.181***
Teacher grades important for spelling/ punctuation/ grammar			-0.863***	-0.871***
Teacher grades important for paper organization			2.743***	2.739***
Teacher grades important for quality/creativity			3.1***	3.105***
Teacher grades important for length of paper			-1.257***	-1.263***
Difficulty of this writing test			-2.644***	-2.644***
Effort on this writing test			-0.378*	-0.389*
Importance of success on this writing test			-0.269	-0.278
Random effects				
Intercept (variance between states)	638.408	426.552	360.18	360.137
Level 1 (variance within states)	525.226	351.111	297.089	297.062
Intraclass correlation (proportion of variance between states)	0.548633	0.548505	0.547995	0.547988
Variance in achievement between states explained (%)	NA	33.185%	43.582%	43.588%
Variance in achievement within states explained (%)	NA	33.151%	43.436%	43.441%

Note. *p<.05. **p<.01. ***p<.001.

3 remained positively related in model 4: whether students considered writing stories or letters a favorite activity and thought writing helped share ideas; the frequency with which teachers talked to students about writing; how often students wrote thoughts or observations, simple summaries, and analyses of essays; how frequently students organized papers, and made changes when writing for school; the frequency of students' use of computers for changes and for accessing the internet when writing papers for school; how frequently students wrote one paragraph in English, science, and social studies or history classes; how often teachers asked students to write more than one draft; and whether teachers in their grading emphasized the importance of paper organization and quality or creativity. The variables that were negatively related to students' NAEP performance in model 3 remained negatively related in model 4: how frequently students wrote a report for school, a personal or imagined story, and business writing; the frequency with which students brainstormed or worked with other students when writing for school; how often students used a computer from the beginning when writing; the frequency of students writing one paragraph in math class; whether teachers in their grading emphasized the importance of spelling, punctuation, or grammar, and of length of paper; and students' perceptions of their efforts and the importance of success on the NAEP writing assessment.

A few students' writing experience variables consistently had large, statistically significant coefficients in both model 3 and model 4. These variables were the frequency with which students made changes when writing for school, used computers for changes when writing papers for school, wrote one paragraph in English class, and had teachers who in their grading emphasized the importance of quality or creativity and paper organization, as well as whether students thought that writing helped share ideas. State-NAEP differences explained an additional 0.006% of the variance between states (43.588% instead of 43.582% in model 3) and an

additional 0.005% of the variance within states (43.441% instead of 43.436% in model 3). The between-state variances remained significant.

5. Discussion

The main finding of this study is that students' preparedness for the NAEP tasks, namely their home states' assessments' similarity to the NAEP, also plays a role in students' performance on the NAEP. Students from those states with writing assessments more similar to the NAEP perform significantly better than students from states with writing assessments that are less similar to the NAEP. However, this predictor only explains a little of the variance in the outcome variable—students' NAEP performances; thus, it does not negate the interpretation of NAEP scores as an indicator of students' writing abilities.

Research has shown that students' demographic backgrounds have a significant impact on students' writing quality (Garielson, Gordon, & Englehard, 1999; Ball, 1999; Kanaris, 1999, Silva, 1993; Ferris). This study's results confirm these assertions. All of the students' demographic variables were found to be statistically significant in all models. More specifically, students who were ELLs, had IEPs, or were eligible for free/reduced priced lunch performed significantly poorer than students who were without those characteristics. Students who were Black, Hispanic, or American Indian performed significantly poorer than students who were White. Asian students performed significantly better than White students, and female students performed significantly better than male students.

Research has shown that students' attitudes and motivations have a significant impact on their writing achievements (Graham, Berninger, & Fan, 2007). More specifically, students' positive beliefs and attitudes about writing contribute to their levels of motivation to write (Bruning & Horn, 2000). This study's results confirm this assertion by finding that students who

thought that writing helped to share ideas performed better than students who did not. However, this study also finds that students' perceptions of the importance of the NAEP writing test were not significantly related to their writing performances. Moreover, students who believed that they exerted more effort on the NAEP writing test did not perform as well as those who did not. It is possible that students who found they needed to devote more effort were also those students who found the writing test more difficult, which would explain why they did not perform as well.

Research has also shown that students' writing activities inside classrooms, such as how often they write, have a positive effect on students' compositional quality (Graham, Kiuahara, McKeown, & Harris, 2012). In this study, almost all students' writing activities inside the classroom were found to be significantly related to their writing performance except the frequency with which students wrote letters or essays for school. However, some of the students' writing activities were found to be negatively related to their writing performance, including how frequently students wrote reports, personal/imaginative stories, and business writing; the frequency with which they brainstormed and worked with other students when writing; and the frequency of writing wrote one paragraph in math class. It is unclear why these activities were negatively related to students' writing performances. Among the positively related variables, how often students revised and wrote in English class was consistently associated with large coefficients in all models. This finding seems to confirm the assertion that the frequency with which students write has a positive effect on their writing quality.

Research has also shown that students' writing experiences have a significant impact on their writing quality. All variables regarding students' writing experiences were found to be significantly related to their performance. However, some of the students' writing experiences were found to be negatively related to their writing performance, including the frequency of

using computers from the beginning when writing papers, and whether teachers emphasized the importance of spelling/punctuation/grammar and length of papers in their grading. Perhaps teachers' overemphasis on the mechanics of students' compositions distracted them from improving the organization and overall quality of their compositions. Among the positively related variables, whether teachers emphasized quality or creativity and paper organization in their grading was consistently found to have large coefficients in all models. This finding suggests that though teachers' feedback tends to improve students' writing quality (Rogers & Graham, 2008), the things teachers emphasize in their feedback also matters.

6. Implications

The results of this study show that state and NAEP assessment differences play a role in students' performances on the NAEP. This finding has three implications: First, it should raise awareness that students' NAEP performances are a result of many factors, including the similarity of students' home state assessments to the NAEP. Because the NAEP is a low-stakes assessment, students are unlikely to prepare for it; however, high-stakes assessments in students' home states tend to impact the instruction and writing experience students get in school. When states' assessments are more similar to the NAEP, students have indirectly prepared for it; as a result, their performance on the NAEP is slightly better than that of students whose home state assessments are more dissimilar. Therefore, when students' performances on the NAEP are compared, we have to be aware of their different levels of preparedness as a result of their home states' writing assessments' similarities and differences with the NAEP.

Second, this finding does not suggest that state and NAEP assessments should be designed to be more similar. Instead, both the NAEP and states' assessments can move forward by incorporating more evidence-based writing assessment practices, which are likely to shrink

the differences between the NAEP and states' assessments. As a result, students' performances on the NAEP are less likely to be impacted by their different levels of preparedness for the NAEP's tasks. Third, the large amount of unexplained variance remaining between states suggests that there are still more state-level variables to be explored, such as the alignment between states' standards and assessments, and the stringency of states' accountability policies.

7. Limitations

This study only controlled for students' characteristics in the multilevel modeling. It did not study teacher characteristics and school characteristics. Teachers' characteristics (such as their educational backgrounds and teaching experiences) and schools' characteristics (e.g., staff opportunities for professional development in writing, and the existence of and extent to which writing was a school-wide initiative) are both likely to impact students' performances on the NAEP. However, investigation of these groups of characteristics was beyond the scope of this project.

In this study, the main variable of interest was at the state level, and the outcome variable was at the student level, thus, state- and student-level were two essential levels to investigate the research question of this study. It is assumed that compared with the impact of states' assessment characteristics and students' backgrounds and experiences in writing, the impact as a result of differences among teachers and schools is relatively small on students' NAEP performances. While there has been limited research done to study the teachers' and schools' effects on students' achievements using NAEP data, Lubienski and Lubienski (2006) examined the NAEP 2003 data with hierarchical linear models to study whether the disparities in mathematics achievement were a result of schools' performances or student demographics. Their study found that when students' demographic differences are controlled for, private school advantages no

longer exist. This suggests that students' demographic variables have more impact on students' performances than one of the central characteristics of schools.

The assumption referred to above is also made for two computational reasons. First, it simplifies the model and increases the precision and efficiency of estimation, as well as allowing a focused investigation of the research question. Second, unless there is strong evidence supporting teacher-level and school-level effects, it is better not to include these two levels because it causes computational difficulties and can produce meaningless and inaccurate estimation as a result of small variances. Nevertheless, it is acknowledged that teachers' and schools' characteristics are important components of students' experiences with schooling. Therefore, future research should be conducted to investigate state-level differences when teachers' and schools' characteristics are accounted for in addition to students' characteristics.

CHAPTER 3: Genre Demands in State Writing Assessments

1. Introduction

Since implementation of the No Child Left Behind Act of 2001, state assessments have been a heated topic for discussion given their important role in states' accountability systems. As state assessments tend to influence curricula, student promotion and retention, and ratings of teacher effectiveness (Conley, 2005), their validity has also been explored (Beck & Jeffery, 2007; Carroll, 1997).

A validity concern raised regarding state writing assessments is the level of ambiguity in prompts. Beck and Jeffery (2007) examined 20 state exit-level writing assessment prompts from Texas, New York, and California, and found that the terms “discuss” and “explain” appeared in 20% of the prompts. However, words like “discuss” do not necessarily align with conventional genre categories. For example, a prompt may ask a student to “discuss” something; depending on what follows “discuss,” however, such a prompt can be requesting either an explanation or an argument. Because “discuss” can be used for eliciting a range of rhetorical purposes, it becomes “an ambiguous directive that does little to help students understand what is expected of them” (Beck & Jeffery, 2007, p.65).

Meanwhile, besides the traditional meaning of “explain,” which asks the writer to explain how something works and often leads to an expository essay, “explain” has been used in two other ways: as an indication that students should take a position and argue for it, which can be classified as argumentative, and that they should give the definition and classification of something, which can be considered descriptive. Thus, there is a lack of precision in these writing prompts. Jeffery's (2009) study of 68 prompts from 41 state exit-level direct writing assessments, in which students produced texts in response to prompts, also suggested that verbs

such as “explain” generated more than one genre category depending on the objects of “explain.” These objects “varied with respect to the degree of abstraction and the extent to which propositions were presented as arguable” (Jeffery, 2009, p.8).

Moreover, in Beck and Jeffery’s (2007) study, 14 prompts out of the 20 examined specified multiple rhetorical purposes. For example, one prompt asked students to “discuss two works of literature,” choose to “agree or disagree with the critical lens,” and then “support” their opinions (p.68). Beck and Jeffery (2007) suggested that, although this prompt was categorized as “argumentation,” the expectation that students should produce an argument was implicit, thus making the prompt ambiguous.

Ambiguity in prompts and implicit expectations of prompts can be viewed as two separate problematic features, rather than the unitary concept in Beck and Jeffery’s (2007) study. Ambiguity is defined in this paper as the presence of two or more conflicting genre demands in a prompt. For example, consider the following prompt: “You find something special. Describe what it is and what you do with it.” The initial statement that “You find something special” can be understood as setting the stage for a narrative account. “Describe what it is” suggests a descriptive text is expected. “Describe ...what you do with it” can be interpreted in two ways. The first interpretation is that the writer should “explain what you do with it,” which suggests an expository text is expected; the second interpretation is the meaning of “tell us what you decide to do with it,” which along with “you find something special” again suggests a narrative text is expected. Because these three genre demands compete for an examinee’s attention, this prompt can be considered ambiguous.

Critics in the genre studies and writing rhetoric communities may argue that there are very few “pure” genre structures invoked in real communicative contexts; rather, there is often

blending. In that case, perhaps we should encourage students to do this kind of blending. This might be a valid approach to prepare students for real communicative tasks; however, there are often high stakes involved in the large-scale assessments and time constraints imposed on students during testing. Therefore, we have to be aware of the additional cognitive demands we place on students, as well as threats to the validity of the assessments when prompts can be interpreted from multiple perspectives.

The second potentially problematic feature of writing assessment prompts is implicit expectations. A prompt's implicit expectation is defined in this paper as the prompt's lack of verbs (e.g., "argue," "convince") or nouns (e.g., "story") that explicitly signal the genre desired in response to the writing prompt. For example, consider the following prompt: "Write about an important lesson that children should learn." This prompt can be also phrased, "Explain an important lesson that children should learn," which suggests an expository text is expected. However, none of the words in either version of the prompt explicitly signal the desired genre. Thus, this prompt would be considered to have an implicit rather than explicit genre expectation.

When discussing possible reasons for the confusing signals about genre expectations in the prompts they examined, Beck and Jeffery (2007) suggested that test designers may assume that students have limited experience with different genres, and thus lack sufficient vocabulary knowledge to associate these key verbs, nouns, and phrases with responding using specific genres. As a result, test designers resort to terminology they feel will be familiar to students, such as "support." However, practice is ahead of research in this area. Little research has been done to examine the thinking processes that students adopt when reading writing prompts. Students' vocabulary precision is one potential area for future research using procedures such as think-aloud protocols and interviews.

A prompt can be ambiguous, or contain implicit expectations, or both. Therefore, tools are needed to examine prompts for ambiguity and lack of explicit genre expectations. Glasswell, Parr, and Aikman (2001) have outlined conventional genre classifications with six genres: “to explain,” “to argue or persuade,” “to instruct or lay out a procedure,” “to classify, organize, describe, or report information,” “to inform or entertain through imaginative narrative,” and “to inform or entertain through recount” (p.5). They also specified these genres’ purposes, functions, types, features, text organization/structure, and language resources. Their work can serve as a reference for identifying genre demands in prompts.

Meanwhile, by identifying demand verbs and corresponding objects (e.g., “convince” and “your friend” in “convince your friend to try something new”), syntax analysis (Jonassen, Hannum, & Tessmer, 1999) can be used to spot words that signal rhetorical processes that can be matched with genre demands (Beck & Jeffery, 2007; Jeffery, 2009). The basis of syntactic analysis is the sentence, in which each word is assigned a label (e.g., subject, verb, object of that verb). Such labeling allows the key verbs and objects of the verbs to be spotted and matched with genre demands.

The ambiguities and implicit expectations in writing prompts may be attributable to the following factors: (a) test designers using terminology that they consider most familiar to students, such as “support,” rather than adopting more explicit verbs for genres, such as “argue,” and (b) test designers purposefully including conflicting genre demands to give students choices in their compositions (Beck & Jeffery, 2007). However, such ambiguities and implicit expectations pose threats to the validity of state writing assessments for the following reasons: (a) different interpretations of writing prompts can lead to students producing compositions that are not representative of their writing abilities, and (b) a lack of consensus among test designers

as well as scorers of the responses may lead to unclear expectations of writing from students, which will result in unfair judgments of students' writing competence. This is especially problematic when a prompt is ambiguous or has implicit expectations while it has a rubric that emphasizes genre mastery. Therefore, it is important to examine this phenomenon.

Jeffery's (2009) five-criteria coding scheme provides just such a tool for examining rubrics for genre mastery. This coding scheme was developed through an inductive analysis of rubrics for exit-level writing assessment prompts. The coding scheme includes rhetorical, genre-mastery, formal, expressive, and cognitive rubrics. Rhetorical rubrics focus on "the relationship between writer, audience, and purpose across criteria domains" (p.10). Genre-mastery rubrics emphasize "criteria specific to the genre students are expected to produce" (p.11). Formal rubrics conceptualize proficiency "in terms of text features not specific to any writing context" (p.11). Cognitive rubrics target "thinking processes such as reasoning and critical thinking across domains" (p.12). Expressive rubrics portray "good writing" as "an expression of the author's uniqueness, individuality, sincerity and apparent commitment to the task" (p.12). Jeffery (2009) suggested that one way to illuminate the underlying proficiency conceptualizations in large-scale writing assessments is to analyze the relationships between genre demands and scoring criteria. Using the above coding framework, 40 rubrics were coded in Jeffery's (2009) study with inter-rater agreement of .83.

When state writing assessment prompts are ambiguous or contain implicit expectations, it brings into question whether students are expected to demonstrate mastery of the demands of the genre(s) presented in prompts; if not, what genres are students expected to master? State standards provide an answer by specifying what students are expected to learn. Moreover, state standards tend to have a significant impact on classroom instruction—teachers have been

reported to increase their instructional emphasis on writing for specific genres in response to changes in standards (Stecher, Barron, Chun, & Ross, 2000).

For these reasons, an examination of genre expectations in state standards that correspond with state writing assessments will help identify the range of genres middle school students are expected to master in different states. It will not only present the state of alignment between genre expectations in standards and assessments using a representative sample, but also help provide an answer to what genres are expected to be mastered when ambiguity and implicit expectation arise. Troia and Olinghouse's (2010) coding taxonomy with comprehensive coverage of 21 genres provides just such a tool for identifying the genre expectations in state standards. Their taxonomy was derived from several theoretical frameworks, including Hayes' cognitive model of writing (Flower & Hayes, 1981; Hayes, 1996), socio-cultural theory (Prior, 2006), genre theory (Dean, 2008), linguistic models of writing (Faigley & Witte, 1981), and motivation theories of writing (Troia, Shankland, & Wolbers, 2012). The indicators found within the "writing purpose" strand in their coding taxonomy cover a variety of communicative intentions accomplished through different genres.

While a small number of studies have been conducted to examine the ambiguity or genre demands of high school exit-level writing prompts (Beck & Jeffery, 2007; Jeffery, 2009), no research has been done to examine the genre demands of middle school state writing assessment prompts, or issues with ambiguity and implicit expectations in those prompts. Nevertheless, writing in middle school is important because middle school students start to be able to think abstractly and use language in more complex ways (De La Paz & Graham, 2002). A study of genre expectations in the prompts for middle school students thus becomes necessary because it will make an important part of the writing expectation explicit and thus help better prepare

students for writing tasks. As the NAEP assesses students' writing at grade 8, seventh graders and eighth graders are also frequently assessed in state writing assessments. It is therefore important that these large-scale assessments are examined in terms of their writing constructs to ensure their validities. The fact that both NAEP and many states assess students' writing at grade 8 also provides a large sample to compare national and state writing assessment at the same grade level, which has not yet been extensively studied.

This study aims to fill that gap by examining genre expectations in seventh and eighth grades. In addition to classifying state writing assessment prompts into different genre categories, this study will use syntactic analysis to investigate multiple competing or conflicting genre demands within each prompt to shed light on the problems of ambiguities and implicit expectations in writing prompts for middle school students. For each prompt, the demand verbs and corresponding objects will be identified and the rhetorical purposes signaled will be matched with the existing genre demands outlined in Glasswell, Parr, and Aikman (2001). This study will also highlight the connection between genre demands in writing prompts and genre-mastery expectations in rubrics and state standards to discuss the validity of state writing assessments.

2. Research Questions

Through analyses of state writing assessment prompts, writing rubrics, and state writing standards, this paper aims to answer the following questions:

1. How many state writing prompts possess the problematic features of ambiguity and/or implicit genre expectations? Which key words in prompts are associated with ambiguity and implicit genre expectations, and how frequently do they appear?
2. What is the relationship between prompts' genre specification and rubrics' genre-mastery expectations?

3. What is the relationship between genre expectations in state standards and writing assessment prompts?

3. Method

3.1 State Direct Writing Assessments and Standards

This study was carried out using data from a prior IES-funded study—the K-12 Writing Alignment Project (Troia & Olinghouse, 2010-2014). In the K-12 Writing Alignment Project, email inquiries and phone calls were conducted to request documents from appropriate assessment personnel located through states' Department of Education websites. Because the K-12 Writing Alignment Project examined the alignment between state writing standards and assessments prior to the adoption of the CCSS and used the NAEP 2007 assessment for its inclusion of state-level data, state direct writing assessments were gathered mainly from between 2001 and 2006, to ensure the representation of the time period. Representative state writing assessment documents including a representative writing prompt, its rubric, and the administrative manual for each genre in each grade being assessed were collected from each time span between major revisions of state assessments.

This study examined 78 prompts and 35 rubrics from 27 states³ (see Appendix C for details). No NAEP data existed for Alaska, Nebraska, Oregon, and South Dakota for the chosen time period. State writing standards or writing assessments were not available for Connecticut, Iowa, Pennsylvania, Montana and New Mexico between 2001 and 2006. No 7th grade and 8th grade writing assessment existed in Ohio during the period 2001-2006. As a result, this study did not include these states' direct writing assessments.

³ The following chose not to participate in the study: Colorado, Delaware, the District of Columbia, Georgia, Hawaii, Maryland, Minnesota, Mississippi, New Hampshire, New Jersey, North Dakota, South Carolina, Utah, and Wyoming.

The collected state direct writing assessment documents were compiled. In each compiled file, there are verbal directions from administration manuals for direct writing assessments, actual prompts, supporting materials provided (e.g., dictionary, writer's checklist), sessions arranged for writing tests, time given, page limits, and whether (and what kinds of) technology was used. There were as many compiled documents for each state as the written responses expected from students each year. In other words, if students took only one prompt with rotated genres each year, there would be a single compiled document for that state containing a representative prompt from each rotated genre to represent the scope of genres assessed. These compiled documents and rubrics were later coded with the coding taxonomy.

Similar procedures were applied to gathering state standards. Within each state and grade, all standards closely related to writing were coded. To ensure the reliability of coding within and across states, the unit of content analysis (i.e., the smallest grain size for a set of standards) was determined to be lowest level at which information was presented most consistently in a set of standards and designated level A. The next level of organization was designated level B, the next, level C, and so forth. Each individual code was applied within level A only once to avoid duplication, but multiple different codes could be assigned to any given unit. To accommodate the potential for additional information presented at higher levels of organization for a set of standards, unique codes were assigned at these superordinate levels (levels B, C, and so on), but duplication of codes from the lower levels was not allowed. Therefore, states' writing standards were rendered comparable regardless of their different organizations. In this study, genre expectations in state standards were only examined for grades 7 and 8.

3.2 Data Coding

Genre demands in prompts. To distinguish genre demands within each prompt, this study used syntactic analysis (Jonassen, Hannum, & Tessmer, 1999) to identify demand verbs and their corresponding objects in prompts. Key words such as main verbs were recorded, tallied, and considered as signals for rhetorical purposes. These signals were compared with the conventional genre classifications as outlined in Glasswell, Parr, and Aikman (2001). When there were two or more genre demands within a prompt, the prompt was recorded as ambiguous. When there were no explicit verbs/nouns for genres, the prompt was recorded as containing an implicit expectation. All explicit verbs/nouns for genres (e.g., “argue,” “convince”) were recorded. Concordance software was used to count the frequencies for all explicit verbs/nouns.

Genres of prompts. This study used a seven-category genre coding scheme adapted from the third strand (purposes) of Troia and Olinghouse’s (2010) coding taxonomy and Jeffery’s (2007) genre coding scheme to code the genres of the prompts. Troia and Olinghouse’s (2010) coding taxonomy ensured comprehensive coverage of the writing purposes with 21 indicators. A preliminary frequency analysis of state writing prompts’ genres coded with this coding taxonomy indicated that only seven genres were assessed in state writing assessments—expository, descriptive, persuasive, response-to-literature, descriptive, narrative, and summary. Jeffery’s (2009) coding taxonomy was derived from an inductive analysis of state exit-level direct writing assessments and differentiated similar genre categories such as persuasive and argumentative prompts and expository and informative prompts. Such differentiations were helpful in distinguishing similar genres in this study. Therefore, a seven-category genre coding scheme was used. These seven categories were: descriptive, persuasive, expository, argumentative, informative, narrative, and analytic. The author of this dissertation served as one of the two raters. A graduate student in Digital Rhetoric & Professional Writing served as the

second rater. The two raters first practiced coding with a training set. When they reached 85% inter-rater agreement, they moved into coding the actual prompts and reached an inter-rater reliability of .93. Differences were resolved through discussion.

Genre expectations in rubrics. This study used the five-criteria coding scheme developed by Jeffery (2009) to examine rubrics for genre-mastery expectations. While the coding scheme includes rhetorical, genre-mastery, formal, expressive, and cognitive rubrics, special attention was paid to the connection of genre demands in prompts and the genre-mastery category as coded in rubrics. Genre-mastery rubrics emphasized criteria specific to the genre expected in the prompts; though these rubrics might contain descriptions that also signify other categories such as expressive, cognitive, or formal, all the descriptions were “framed by the specific communicative purpose that characterizes the genre” (Jeffery, 2009, p.11). Jeffery (2009) gave this example from a 6-point rubric in Nevada: “clarifies and defends or persuades with precise and relevant evidence.” This example signified a genre-mastery category because of the expectation of effective persuasive writing. These rubric types represented what different “discourses of writing”—“constellations of beliefs about writing, beliefs about learning to write, ways of talking about writing, and the sorts of approaches to teaching and assessment which are likely to be associated with these beliefs” (Ivanic, 2004, p.224)—value as assessment criteria. The relationships between genre demands in prompts and rubric types illuminated the underlying proficiency conceptualizations contained in large-scale writing assessments (Jeffery, 2009). The two raters who coded the prompts followed the same procedure and coded the rubrics. They reached an inter-rater reliability of .86 and resolved differences through discussion.

Genre expectations in state standards. Genre expectations in state standards have been coded with the third strand (purposes) of Troia and Olinghouse’s (2010) seven-strand coding

taxonomy in the K-12 Writing Alignment Project. The genre expectations that appeared in those 27 states' grade 7 and grade 8 standards were recorded. The inter-rater reliability was .87 for standards coding. To also allow genre expectations in state standards and writing prompts to be comparable using Jeffery's (2009) genre coding taxonomy, when the persuasive and expository genres were coded in the writing standards according to Troia and Olinghouse's (2010) coding taxonomy, they were further categorized as either persuasive or argumentative and either expository or informative as in Jeffery's (2009) genre coding taxonomy. As a result, genre expectations in state standards were coded with the third strand (purposes) of Troia and Olinghouse's (2010) seven-strand coding taxonomy modified to accommodate Jeffery's (2009) genre coding scheme.

In the current study, the "purposes" strand of Troia and Olinghouse's (2010) was modified by breaking out persuasion and argumentation to accommodate Jeffery's (2009) genre coding taxonomy; the 21 writing purposes in the strand were changed into 22 purposes. The author of this dissertation and a doctoral student in English Literature served as raters. The two raters coded standards following the same procedure of coding prompts and rubrics. They reached an inter-rater reliability of .86 and resolved differences through discussion.

3.3 Data Analyses

The percentages of prompts that were either ambiguous or contained implicit expectations were recorded. The key verbs/nouns associated with ambiguity and implicit expectations and their frequencies were also recorded. The connections between the ambiguity and implicit expectations of prompts and their rubrics' categories were examined, with special attention to the genre-mastery category. Genre expectations in standards were obtained from the coding of standards using Troia & Olinghouse's (2010) coding taxonomy modified to

accommodate Jeffery's (2009) genre coding scheme. Ambiguity and implicit genre expectations in prompts were determined through the syntactic analysis of prompts in the former data coding step. Genre expectations from state standards were presented with the genres assessed in state writing prompts. The genres assessed in state writing prompts were identified by the two raters using the seven-category genre coding scheme adapted from the third strand (purposes) of Troia and Olinghouse's (2010) coding taxonomy and Jeffery's (2009) genre coding scheme. When there was ambiguity in a prompt, states' identification of the genre of the prompt was taken into consideration.

4. Results

4.1a. How many state writing prompts possessed the problematic features of ambiguity or implicit genre expectations?

Among 78 prompts, 11 prompts from seven states were considered ambiguous, and seven prompts from four states were determined to have implicit genre expectations. In other words, 14% of prompts were ambiguous, and 9% of prompts had implicit genre expectations. Together, 23% of prompts possessed one of the two problematic features.

Ambiguous prompts were mostly expository, narrative, argumentative, and informative prompts. The genre coding was based on the syntactic analysis of the prompts; however, in the case of ambiguity, states' identification of the prompts' genres was taken into consideration. There were six expository prompts that were ambiguous. For example, the Massachusetts 2002 prompt asked students to "think of someone who is [their] personal hero," "describe this person," and "explain two qualities they most admire about him or her" in "a well-developed composition." If students were only expected to "describe this person," the prompt could be easily categorized as descriptive, or if students were only expected to "explain two qualities," the

prompt could be easily categorized as expository; however, when the two demand verbs were used in a parallel way, without any specific noun (e.g., “descriptive”, “expository”) to indicate the genre, it was hard to determine which genre was expected. However, a state contact in Massachusetts helped us identify that the genre the prompt was written to assess was expository.

Narrative prompts often had explicit directions for students; for example, “write a story”, or “tell about a time when...” However, there were three cases when narrative prompts employed demand verbs in a way that made the genre expectation ambiguous. For example, in a response-to-literature prompt from Indiana, students were provided with the situation that “if Bessie had kept a journal about her flying experiences, how might she have described her thoughts and emotions?” and directed to “write an essay in which you describe one of Bessie’s flying experiences.” Though “describe” might appear to suggest a descriptive text, “one of Bessie’s flying experiences” indicated a particular experience; moreover, the “journal” context seemed to suggest a narrative retelling of what had happened. Furthermore, because “describe” was used in many different genres, it was hard to make a judgment about the expected genre based on the verb “describe” alone. Consequently, this prompt may have made it difficult for students to figure out whether they should spend more time describing Bessie’s thoughts and emotions from her flying experience or telling a story about one of her flying experiences. Similarly, in the other two cases, “describe” and “explain” were used in an ambiguous way to prompt students’ narrative skills.

There were only four argumentative prompts in this sample. None of the prompts used “argue” as a demand verb; instead, these prompts used “explain” and “describe.” Moreover, the way in which a prompt from Virginia used the demand verb “explain” could potentially lead students to interpret it as looking for expository composition. This prompt read, “Your school is

planning to issue laptop computers to ninth graders next year. Do you think this is a good idea? Write to explain why or why not.” Different from expository prompts, which often asked students to select or identify an item, an event, or a phenomenon to be explained, this prompt asked students to take a position on a two-sided issue and use reasons to support their positions. Therefore, this was classified as an argumentative prompt; however, the use of “explain” as the demand verb made this prompt’s genre expectation ambiguous.

There were only five informative prompts. These informative prompts also often used “explain” and “describe” as the demand verbs, with one exception. The prompt from Arizona read, “Your class has joined a pen pal program. You have selected a pen pal who lives in another state. Write a letter to your new pen pal introducing yourself and telling about your interests.” The verb “tell” is a rhetorical term that is often used in narrative genre writing to mean entertaining through the course of recounting an experience and happenings. In this prompt, however, “tell” was used as a synonym of “inform,” which directed students to provide information about their interests, rather than constructing or reconstructing a view of the world like a narrative often does.

Prompts with implicit expectations were mostly persuasive, expository, and argumentative prompts. Persuasive prompts often had explicit verbs such as “convince” or “persuade.” However, there was one persuasive prompt that did not have any explicit verbs. This Kentucky prompt read, “Select one current issue that you feel people should be concerned about. Write a letter to the readers of the local newspaper regarding this issue. Support your position with specific reasons why the readers should be concerned about this issue.” This prompt did not have a demand verb that explicitly indicated any genre. However, “support” and “position” were often employed by persuasive and argumentative prompts. In contrast to argumentative prompts,

persuasive prompts often contain an explicit reference to their audience. In this case, it was the readers of the local newspaper. However, the lack of demand verbs made this prompt's genre expectation implicit rather than explicit.

Two argumentative prompts also lacked explicit demand verbs. These two response-to-literature prompts from Michigan had very similar structures. One prompt read, "Is this a good example of seventh-grade writing? Why or why not? Use details from the student writing sample to support your answer." In this prompt, there was no demand verb that indicated explicitly the genre to which the students' writing should conform. However, students were expected to take a position either arguing that it was a good example or it was not, and use details to support their positions. Such a genre expectation was considered to be implicit.

Though the majority of the expository prompts used the demand verb "explain," there were still cases where students were given a topic and directed to write about the topic without a clear indication of the genre. For example, a prompt from Arkansas read, "What advice would you consider the best? Why? Write an essay about the best advice. Give enough detail." In this prompt, there was no explicit verb indicating the genre of the prompt. The noun "essay" also did not specify the genre because it could be used to refer to all kinds of writing, including persuasive, narrative, argumentative, and literary analysis essays. Though this prompt might be categorized as expository because when one writes about a topic one frequently has to explain information about the topic, without an explicit demand verb the genre expectation remained implicit.

4.1b. Which key words in prompts were associated with ambiguity and implicit genre expectations, and how frequently do they appear?

The key words mentioned in the previous section associated with ambiguity and implicit genre expectations were: “explain,” “describe,” “essay,” “support,” “discuss,” and “tell.” Table 6 below is a table of the frequencies of words discussed in this section and the percentages of prompts in each genre in which they were used.

“Explain” was widely used in 69% of expository prompts and 83% of literary analysis prompts. It was also used in 22% of persuasive, 6% of narrative, 25% of argumentative and 40% of informative prompts. In other words, “explain” was used in every one of the seven genre prompts except descriptive prompts. Some of these uses evoke unconventional meanings of “explain”. For example:

(1) Write a fictional story about a day during your favorite season. Create a main character or characters and describe the action that takes place during that day. Explain where and when the story takes place (Indiana 2002 8th grade).

(2) Explain how someone lost a privilege as a result of not being responsible (Michigan 2006 8th grade).

(3) Compare your social life as a teenager with your social life as a young child. Explain how it is different and how has it remained the same? Support your main points with examples (Kansas 2004 8th grade).

In these three examples, “explain” was used in several different ways. In the first case, “explain” was a synonym of “describe;” in the second case, it could be interpreted as “give an account of how someone loses a privilege,” while “lost” in the past tense also seemed to suggest students should “tell a story of how someone lost a privilege;” in the third case, it was used in a traditional sense, meaning providing information about the given topic.

“Describe” was also widely used in all genres except persuasive prompts. It was used in

Table 6 Frequency (F) and Percentage (P) of Key Words Usage in Genres

	Persuasive (n=18)		Expository (n=26)		Narrative (n=16)		Argumentative (n=4)		Descriptive (n=3)		Informative (n=5)		Analysis (n=6)	
	F	P	F	P	F	P	F	P	F	P	F	P	F	P
explain	4	22%	18	69%	1	6%	1	25%	0	0%	2	40%	5	83%
detail	6	33%	11	42%	5	31%	3	75%	0	0%	2	40%	3	50%
support	7	39%	8	31%	0	0%	3	75%	0	0%	0	0%	5	83%
describe	0	0%	6	23%	6	38%	1	25%	3	100%	1	20%	2	33%
essay	4	22%	9	35%	1	6%	1	25%	0	0%	0	0%	3	50%
reason	8	44%	4	15%	0	0%	1	25%	0	0%	0	0%	0	0%
convince	8	44%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
story	0	0%	0	0%	7	44%	0	0%	0	0%	0	0%	1	17%
tell	0	0%	1	4%	7	44%	0	0%	0	0%	1	20%	0	0%
persuade	6	33%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
answer	0	0%	2	8%	0	0%	2	50%	0	0%	0	0%	2	33%
position	3	17%	0	0%	0	0%	1	25%	0	0%	0	0%	0	0%
idea	0	0%	1	4%	1	6%	0	0%	0	0%	1	20%	1	17%
conclusion	3	17%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
persuasive	3	17%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
response	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	1	17%
opinion	3	17%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
compare	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%
discuss	1	6%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%
justify	1	6%	0	0%	0	0%	0	0%	0	0%	0	0%	1	17%
argue	2	11%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
point	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%
evidence	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	17%
theme	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	1	17%

100% of descriptive, 23% of expository, 38% of narrative, 25% of argumentative, 20% of informative, and 33% of literary-analysis prompts. “Describe,” when used alone as the only demand verb in a prompt, often indicated a descriptive prompt; however, some states also used it by itself to indicate a narrative prompt. It was also used in a combination with other demand verbs such as “explain” to indicate genres other than descriptive and narrative. Consider the following examples:

(1) Describe a time when you or someone you know had a difficult experience but learned a valuable lesson from it (Michigan 2006 7th grade).

(2) Think of a teacher that you will always remember. Describe this teacher (Alabama 2004 7th grade).

(3) Think of someone who is your personal hero. In a well-developed composition, describe this person and explain two qualities you most admire about him or her (Massachusetts 2002 7th grade).

In examples (1) and (2), the meaning of “describe” was different. In example (1), “describe” was used as the equivalent of “tell a story about a time when ...,” while in examples (2) and (3), “describe” was used in the traditional sense to mean “provide details and attributes about something.” However, different from example (2), in example (3), “describe” was used in conjunction with “explain” to indicate another genre. When “describe” was used alone to indicate genres other than descriptive or used with other demand verbs in a parallel manner, ambiguity in genre expectations often happened.

“Essay” was another popular word used in prompts; however, its lack of genre specification made it similar to other abstract nouns such as “writing,” “composition,” or “answer.” Among its eighteen occurrences, there were only two times when words that explicitly

indicated genres such as “expository” and “persuasive” were used ahead of it. There were other times when “essay” was used with demand verbs to clearly indicate genres. However, when “essay” was used alone—such as in the example, “What advice you would consider the best? Why? Write an essay about the best advice. Give enough detail”—it did not add much to the genre specification of the prompts.

“Support” was another word that was widely used with all kinds of genres; it was used with persuasive, expository, argumentative, and literary-analysis writing. The term “support” was traditionally used in persuasive or argumentative prompts in combination with words such as “position,” “points,” and “evidence.” However, this study showed that “support” was used in 31% of expository prompts. Among these uses, there were a variety of things that “support” was supposed to reinforce—“opinion,” “ideas,” “position,” “theme,” “points,” “response,” “answer,” “details,” “reasons,” and “conclusions.” The use of “opinion,” “conclusions,” and “position” was strongly associated with persuasive and argumentative essays. The use of “reasons” was strongly associated with persuasive writing; however, it was also used with expository writing. Surprisingly, “points,” which traditionally is more associated with persuasive writing, was only used in expository prompts. “Answer” and “details” were used more often with expository writing than any other genre.

“Discuss” was only used three times in the 73 prompts. “Discuss” did not signify any specific genre itself. However, in each case it was used in conjunction with other demand verbs. Here are the three examples:

(1) Describe a special privilege or right that people your age are sometimes given and discuss the responsibilities that go with it (Michigan 2006 8th grade).

(2) Your teacher has asked you to write an essay discussing what you would do if you could be President for one day... Now write an essay about what you would do if you could be the President for one day... Explain your ideas clearly so that your teacher will understand (Arkansas 2007 8th grade).

(2) The Television Advertisers Association is sponsoring an essay contest for students. Students are invited to submit essays that discuss ONE thing about television advertising they believe should be changed. Write an essay for the contest identifying the change that should be made and persuading your reader why this change is important (Wisconsin 2007 8th grade).

In these three prompts, “discuss” was used with “describe,” “explain,” and “persuade.” The use of “discuss” alone did not indicate the genre of the prompts; thus, the genre specification of the prompt depended on the interaction between “discuss” and other demand verbs. In examples (2) and (3), “explain” and “persuade” were used in the traditional sense and “discuss” was used to reinforce the rhetorical purpose expected, so the prompt could be easily categorized as expository. In example (1), however, “discuss” added another task besides “describe” without specifying the genre, which made the prompt ambiguous.

“Tell” was a verb that often explicitly indicated narrative writing. However, in this study, “tell” was also found to be used in expository and informative prompts. Consider the following examples:

(1) Write a narrative composition telling about ONE time you observed something that was really strange or weird (Illinois 2010 8th grade).

(2) Write an editorial for the local newspaper about the importance of being kind to others. Tell about a time when you observed or participated in an act of kindness. Support your response with details or examples (Kentucky 2007 8th grade).

(3) Write a letter to your new pen pal introducing yourself and telling about your interests (Arizona 2005 7th grade).

(4) Think about a person who has had an influence on your and your life ... Write an essay telling who this person is and explaining why he/she has had such an influence on you (Alabama 2004 7th grade).

In examples (1) and (2), “tell” was used in the conventional way. Example (1) was an explicit narrative prompt. Contrasted with example (1), in example (2) students were expected to explain the importance of being kind to others; however, students were also expected to tell about an event. The expectation that the event should be “told” to support the explanation was implicit, which resulted in ambiguity. In example (3), “tell” was used as a synonym for “provide details.” In example (4), “tell” was used as a synonym for “identify.”

These results show that these genre-associated key words were often utilized in ambiguous ways. There was little consensus about how they should be used to make genre expectations clear and explicit for students.

4.2. What is the relationship between prompts’ genre specification and rubrics’ genre-mastery expectations?

Among the 32 prompts that were used with genre-mastery rubrics, five prompts from three states possessed problematic features (i.e., ambiguity or implicit expectations). In other words, among the 15 prompts that possessed problematic features, five prompts were used with genre-mastery rubrics. These genre-mastery rubrics directed raters to evaluate students’ compositions in terms of whether they demonstrated students’ mastery of the genres as the most important criteria. Table 7 shows the five prompts with problematic features used with genre-

mastery rubrics. The table includes the five prompts' rhetorical processes, key words, genres assessed as a result of prompt coding, and problematic features.

Table 7 Prompts with Problematic Features and Used with Genre-mastery Rubrics

	Rhetorical Purposes	Key Words	Genre Assessed	Problematic Feature
IN 2002 G8	write a fictional story; create a main character or characters; describe the action; explain where and when; details; an event or series of events	fictional, story, character, describe, action, explain, detail, event	Narrative	Ambiguity
IN 2003 G8	write an essay; describe one of Bessie's flying experiences; include two ideas from the poem	essay, describe, experience, idea	Narrative	Ambiguity
KY 2008 G8	select one current issue; write a letter to the readers of the local newspaper; support your position with specific reasons	select, issue, letter, support, position, reason	Persuasive	Implicit Genre Expectation
KY 2007 G8	write an editorial for the local newspaper about the importance of being kind to others; tell about a time; support your response with details or examples	editorial, tell, time, support, response, detail, example	Expository	Ambiguity
VA 2011 G8	write to explain why or why not	Explain	Argumentative	Ambiguity

The rubrics used with the above five prompts all encompassed the genres to be assessed in an implicit or explicit way. However, the interplay between the ambiguity in prompts and the criteria in rubrics might further complicate the writing assessments as illustrated below.

In Indiana's writing rubrics, students were assessed with regard to whether their compositions fully accomplished tasks such as supporting an opinion, summarizing, storytelling, or writing an article. They were also assessed on whether they "included vocabulary to make explanations detailed and precise, description rich, and actions clear and vivid." In other words, the writing rubrics included a range of genres. However, the ambiguity in prompts can interfere with students' understanding of what the task entailed.

One example of this was identified in the 2002 prompt, which was intended to assess students' storytelling ability. It asked students to include key elements of narrative composition such as "main character or characters," "actions," "where and when," and "event." It could be implied that for students to fully accomplish this storytelling task, they had to include these elements. The prompt used language emphasizing "describe actions" and "explain when and where." However, the loose use of "explain" as a synonym of "describe" might have led students to interpret it as if they were expected to provide reasons for choosing the place and the time of the event instead of just describing the place and the time. This ambiguity could have interfered with students' capability to accomplish the task as assessed in the rubrics. When students were instructed to provide "detailed and precise" explanations of "where and when" in their compositions, which was also assessed in the rubrics, how should their compositions be evaluated?

Similarly, the 2008 prompt also tried to assess students' storytelling ability. The main demand verb "describe" could possibly have distracted students from telling about one of Bessie's flying experiences using key elements of narrative compositions, and instead directed students to provide a "rich" description of Bessie's flying experiences in general. In this case, how should their compositions have been evaluated? Could students' "detailed and precise" explanations and "rich" descriptions compensate for students' seemingly off-task performance? These two examples illustrate that the ambiguity in writing prompts can lead students to write compositions in an unexpected way yet still meet the evaluation criteria of the rubrics. This complicates the evaluation of students' writing abilities.

In Kentucky's 8th grade writing rubrics, students were assessed regarding whether they skillfully applied characteristics of the genre; the writing rubrics did not identify the specific

genres corresponding to the prompts. The 2008 prompt did not explicitly specify the genre to be assessed, which left interpretation of the intended genre to students and raters. The 2007 prompt directed students to tell about an event while explaining the importance of being kind to others. Such an arrangement is atypical for the expository genre that students were assessed on; thus, it would pose a challenge for raters to assess whether students skillfully applied characteristics of the assessed genre. This example shows that when prompts are ambiguous, there is little agreement on the genre that was meant to be assessed; even though the rubrics directed raters to assess students' genre-mastery skills, it is impossible for raters to know what characteristics of the genres they should look for in students' writing. Therefore, the ambiguity in prompts undermines the rubrics' emphasis on assessing students' genre-mastery skills.

In Virginia's writing test composing rubrics, students' narrative organization was expected to be intact: minor organizational lapses may be permissible with other modes of writing, but in all types of writing, a strong organizational plan was expected to be apparent. The rubrics included a range of genres, but did not identify what those "other modes of writing" were. The rubrics still expected an apparent strong organizational plan, while they gave students some flexibility to structure their texts. The 2011 prompt assessed students' argumentative writing. The prompt's use of "explain" rather than "argue" might have led students to interpret it as an expository prompt. As a result, students might have organized their texts to make their explanations detailed and precise instead of focusing on employing strong and relevant evidence to support their positions regarding whether they considered it a good idea for their schools to "issue laptop computers to ninth graders next year." Depending on their interpretations of the prompt, students' organizational plans would differ. Meanwhile, students were assessed on these organizational plans. This example echoes the examples from Indiana and Kentucky and shows

that 1) the ambiguity in writing prompts might lead students to write compositions in an unexpected way or in a different mode that nevertheless still meet certain criteria of the rubrics, thus complicating the evaluation of students' writing abilities; and 2) the ambiguity in prompts undermines the rubrics' emphasis on assessing students' genre-mastery skills.

4.3. What is the relationship between genre expectations in state standards and writing prompts?

Table 12 in Appendix A shows the relationship between genre expectations in state standards and writing prompts. It includes the genres expected to be mastered at grades 7 and 8 in state standards, the percentage of each genre out of the total genre occurrences in that state's standards, the genres assessed, and the percentage of the genres in the state standards that were actually assessed (e.g., if a state's standards included five genres but only two were assessed, then the percentage would be 40%). Those genres in standards that appeared in more than 10% of all the genre occurrences were bolded to highlight those more frequently mentioned genres.

Among the seven genres assessed, the most widely-referred-to genre was narrative; it was referenced in 25 states' writing standards; this was followed by persuasive (24 states), expository (23 states), informative (22 states), descriptive (12 states), analysis (7 states) and argumentative (4 states). There were another 12 states whose standards implicitly referred to argumentative genre by describing argumentative genre features without distinguishing argumentative from persuasive texts and 11 states whose standards implicitly referred to features of literary analysis without labeling it as such.

Among the 27 states evaluated in this study, 12 covered all the genres they assessed in their writing standards. Moreover, they also referred to those genres more frequently than other genres in their standards. Another nine states covered all the genres they assessed in their writing

standards. However, they referred to those genres less frequently than some other genres in their standards.

Most importantly, there were six states whose writing standards did not cover all the genres they assessed. Alabama and North Carolina included persuasive writing in their writing assessments; however, persuasive writing was not covered in their writing standards. Maine covered descriptive writing in their writing assessments; descriptive writing was not addressed in their writing standards. Oklahoma assessed expository writing; expository writing was not included in their writing standards. Virginia assessed argumentative writing; argumentative writing was not covered in their writing standards. Finally, West Virginia's writing assessments contained both descriptive and narrative compositions; however, neither of these two genres was covered in their writing standards.

The percentage of those genres in state standards assessed was in the range of 0-60% with an average of 18%. For example, North Carolina included the following writing purposes in their standards: narrate, express, explain, inform, analyze, reflect, and evaluate; however, none of these purposes were assessed; instead persuasive composition was assessed in their writing assessments, thus, 0% of those genres in North Carolina's writing standards were assessed. Vermont included the following writing purposes in their standards: respond to literature potentially covering literary analysis, direct, narrate, persuade, and inform; among them, literary analysis, persuasive, and informative composition were assessed in New England Common Assessment Program (NECAP) direct writing assessment, thus, 60% of these genres in Vermont's writing standards were assessed.

5. Discussion

5.1 Ambiguity in prompts

The findings in the results section identified five scenarios that create ambiguity and implicit expectations in state writing prompts: a) the meanings of “demand verbs” are evoked in unconventional ways, such as “describe a time;” “explain where and when;” b) demand verbs are absent in prompts, for example, “write an essay about the best advice and give enough detail;” c) two “demand verbs” that signal different genres are used in a way that competes for writers’ attention, e.g., “describe a person and explain two qualities;” d) demand verbs, such as “describe,” “explain,” “support,” and “discuss,” which are widely used in a variety of genres, are used on their own without other supplemental information to specify the genre; and e) nouns like “writings,” “responses,” “essays,” “paragraph” are used by themselves to denote the type of writing expected without any other genre-specific demand verbs or nouns.

The findings suggest that “explain,” “describe,” “essay,” “support,” “discuss,” and “tell” were often used in ambiguous ways or were employed to refer to genre implicitly. These findings have confirmed Beck & Jeffery’s (2007) assertion about the lack of consensus of the use of the demand verbs “explain” and “discuss,” as well as terms such as “support.” These findings further suggest unspecified uses of terms such as “describe,” “essay,” and “tell.” “Discuss” appeared much less frequently in middle school prompts than exit-level high school prompts; instead, “describe” appeared much more frequently.

The introduction section of this chapter discussed possible reasons for why the above five scenarios occurred—test designers use terminologies considered most familiar to students, or they purposefully include conflicting genre demands to give students a choice in their compositions (Beck & Jeffery, 2007). These possible reasons cannot justify the threats such ambiguities and implicit expectations pose to the validity of state writing assessments. When writing prompts can be interpreted in multiple ways, students may produce compositions that are

not representative of their writing abilities. This may also lead to unclear expectations about writing performance, resulting in unfair judgments of students' writing abilities.

5.2 Genre Expectation in Standards, Rubrics, and Prompts

About 45% of state writing standards covered all the genres they assessed and referred to those genres more frequently than other genres; 33% of state writing standards covered all the genres they assessed but referred to those genres less frequently than some other genres. The alarming fact is that 22% of state writing standards did not cover the genres that were assessed in the corresponding state writing assessments. When state writing standards do not cover the genres to be covered in state writing assessments, it leaves teachers to determine whether those genres are important enough to be taught. As a result, students receive different levels of preparation to write in those genres. The consequence is that state writing assessments may not only assess students' writing abilities but also assess students' preparedness for the tests.

Genre-mastery rubrics allow states to emphasize students' mastery of genres in their evaluation criteria. When attention is given to genres and explicit direction is included in these rubrics, writing expectations are more likely to be concrete. Thus, utilizing genre-mastery rubrics with explicit genre-component directions for the raters is helpful. If genres are well-specified in prompts and evaluated with genre-mastery rubrics, students' abilities to accomplish the tasks are more likely to be fairly assessed. When a prompt is ambiguous or has implicit expectations with a rubric that emphasizes genre mastery (five prompts out of 68 in this study), this is especially problematic. In this scenario, not only are students given a prompt that can be interpreted in multiple ways, but also their compositions are assessed using criteria about which students are not provided enough information or explicit directions. Therefore, when students' mastery of

genre is an important criterion in rubrics, it is even more important that prompts are explicit with respect to genre expectations.

5.3 Validity of State Writing Assessments

The above aspects posed potential threats to the validity of state writing assessments—the standards do not cover what is to be assessed, and the prompts do not explicitly specify genres while their rubrics assess students’ mastery of genres. Standards of test development emphasize that “the instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended” (AERA/APA/NCME, 2011, p.47). When writing rubrics assess students’ mastery of genres but the prompts do not explicitly specify the genres being assessed, students lack sufficient information to respond to those prompts in the way that test developers intended. If test designers purposefully include conflicting genre demands to give students choices in their compositions, there is little evidence to suggest that this practice actually helps “increase students’ engagement and allow them to demonstrate their best possible writing performance” (Beck & Jeffery, 2007, p.76). Therefore, aligning assessments with state standards, aligning rubrics’ criteria with prompts’ genre expectations, and making prompts’ genre expectations explicit will help ensure the valid interpretation of state writing assessments.

6. Implications

State assessments should be aligned with state standards to ensure that those genres being assessed are also covered in state standards. This is important because state standards specify what students are expected to learn. Teachers have been reported to increase their instructional emphasis on writing for specific genres in response to changes in standards (Stecher, Barron, Chun, & Ross, 2000). When genres are assessed without being specified in state standards, it

leaves to teachers to decide whether those genres are important for students to learn; as a result, students receive different levels of preparation to write for those genres and have to shoulder the consequences of high-stakes testing. Prompts should make their assessed genres explicit. To avoid the five scenarios that tend to cause ambiguity and/or implicit genre expectations in state writing prompts, the following recommendations should be considered:

- a) Try to include relevant demand verbs in prompts. For example, use “tell” in narrative prompts; use “persuade” in persuasive prompts (along with an explicit audience), use “argue” in argumentative prompts, and so forth;
- b) Make sure that the meanings of “demand verbs” such as those above are evoked in conventional ways;
- c) When two or more “demand verbs” which signal different genres have to be used in the same prompt, their relationships should be explicit. In other words, it should be explicit how those rhetorical processes should work together to achieve a specified purpose. For example, if it is expected that students will explain the importance of being kind to others, tell about a time when they observed or participated in an act of kindness, and support their response with details or examples; then the prompt should specify the role of the narrative event in students’ compositions such as “Write to explain the importance of being kind to others. In your expository essay, include details and an example in which you tell about a time when you observed or participated in an act of kindness to elaborate your idea;”
- d) When demand verbs such as “describe,” “explain,” “support,” and “discuss,” which are widely used in a variety of genres, are used on their own, there should be other supplemental information giving more details about genre expectations; and

e) More concrete nouns that signify genres, such as “story,” “description,” “exposition,” “persuasion,” and “argument” should be used in prompts to indicate the expected responses.

These practices will help make genre expectations in prompts explicit. Future research can be conducted to investigate whether state writing assessments are more likely to be fair assessments of students’ writing abilities under these circumstances—when those genres explicitly assessed in prompts are covered by state writing standards and genre-mastery rubrics are used to evaluate whether students’ compositions accomplish the specified task demands.

More research is needed to examine the thinking processes that students adopt when reading writing-assessment prompts. Students’ vocabulary precision is also a potential area for future research using procedures such as think-aloud protocols and interviews.

7. Limitations

This study only explored the coverage of genres in prompts, rubrics, and state standards. It did not explore the attributes of those genres students are expected to master, though a study of those would contribute to our understanding of genre knowledge specified in schooling. Meanwhile, genre expectations in state standards were only examined at grades 7 and 8. On the one hand, this might have caused underrepresentation of genre expectations in some states, when genres expected and assessed at lower grades did not appear again in seventh and eighth grade state standards. On the other hand, a rationale for including only seventh and eighth grade state standards was that if states intended to emphasize certain genres to be mastered by seventh and eighth graders, they should include those genres in the state standards for those grades regardless of whether those genres had appeared in earlier grades. It would be even more important for those genres to be specified in the state standards for those grades if those genres were further assessed in the state’s writing assessments.

CHAPTER 4: Summary and Moving Forward

The three pieces of research presented in this dissertation have investigated the writing constructs underlying state and national writing assessments, explored the relationship between the differences in state and national assessments and students' NAEP performances, and examined important components of writing assessments in depth. This chapter will review major findings, highlight implications for state writing assessments and the NAEP, as well as for writing prompt design, and offer some future directions for research.

1. Major Findings

1.1 Prevalent Writing Practices

Among the 27 states examined, only three states gave students choices of prompts, thus illustrating it was not a popular practice (at least by 2007). The writing process approach had an impact on the writing assessment because the majority of states (26/27) directed students to plan, and more than half of the states directed students to revise and edit. However, few states provided separate planning, revision, and editing sessions. Only seven states gave students two prompts. The only exception was New York, which gave students four integrated writing tasks that included responding after both listening and reading activities. The integrated writing tasks in New York's assessment suggest a potential path for increasing students' writing opportunities by integrating listening and reading assessments with writing assessments.

The majority of states (20/27) specified an audience in their writing prompts, and at least 30% of writing rubrics emphasized the importance of authors' consideration of the intended audience in their compositions. However, the writing prompts incorporated a wide range of audiences including general "readers," pen pals, and students' classes, classmates, or teachers.

An emphasis on organization, content, and detail was a feature in almost all writing rubrics; word choice, sentence fluency, style, and grammar, including sentence construction, were also highly prized aspects of students' papers. General conventions, such as capitalization, punctuation, and spelling were also assessed by the majority of states. This shows that, regardless of the rubric types, these aspects are considered necessary for demonstrating writing proficiency by most states. Only ten states included genre-specific components in their rubrics; persuasive essays' components are most often specified compared with other genres. While expository is the most assessed genre (16/27 states), only four states specified expository essays' components in their rubrics. By 2007, only West Virginia had online writing sessions for their state direct writing assessments.

1.2 Genre Demands in Direct Writing Assessments

The most popular prompt genre in middle school assessments was expository, followed by persuasive, narrative, informative, analytic, argumentative, and lastly descriptive. Half of the rubrics were genre-mastery rubrics. Few rubrics emphasized creativity and critical thinking. Genre-mastery rubrics were used with all genres, while rhetorical rubrics were not used with descriptive prompts. About the same number of states used either genre-mastery rubrics or rhetorical rubrics. Only six states had genre-mastery rubrics that contained genre-specific components. This finding suggests that the genre evaluation criteria that states place on students' writing are either vague or not fully utilized to assess students' genre mastery.

1.3 State and National Alignment

State writing assessments and the NAEP align in their adoption of the writing process approach, their attention to audience and students' topical knowledge, their accommodations through procedure facilitators, and their inclusion of organization, structure, content, details,

sentence fluency, and semantic aspects as well as general conventions such as punctuation, spelling, and grammar in their assessment criteria. However, the NAEP writing assessment differs from many states' writing assessments by having explicit directions for students to review their writing, giving students two timed writing tasks, making the informative genre—rarely assessed in state assessments—one of the three genres assessed, and including genre-specific components in their writing rubrics. One of the biggest differences between the NAEP and most of the state writing assessments is that all of NAEP's writing rubrics are genre-mastery rubrics with genre-specific components. Thus, when state and national writing assessment results are compared, these two assessments differ in the genres they assess, the time and the number of tasks they give to students, and the level and specificity of genre demands they emphasize in their evaluation criteria.

1.4 The Relationship between the Variability between State and National Assessments and Students' NAEP Performance

Students' preparedness for the NAEP tasks, namely their home states' assessments' similarity to NAEP, is found to play a marked role in students' performance on the NAEP. Students from those states with writing assessments more similar to the NAEP perform significantly better than students from states with writing assessments more different from the NAEP. However, this predictor only explains a small amount of the variance in the outcome variable (students' NAEP performance); consequently, it does not negate the interpretation of NAEP scores as an indicator of students' writing abilities.

1.5 The Relationship between Students' Characteristics and their NAEP Performance

All of the students' demographic variables were found to be statistically significant in all models. More specifically, students who were English Language Learners, had IEPs, or were

eligible for free/reduced priced lunch performed significantly poorer than students who were without those characteristics. Black, Hispanic, or American Indian students performed significantly poorer than students who were White. Asian students performed significantly better than White students, and female students performed significantly better than male students.

Students who thought that writing helped share ideas performed better than students who did not. Students' perceptions of the importance of the NAEP writing test were not significantly related to their writing performances. Moreover, students who believed that they exerted more effort on the NAEP writing test did not perform as well as those who did not.

Almost all students' writing activities inside the classroom were found to be significantly related to their writing performance, the exception being the frequency with which students wrote letters or essays for school. However, some of the students' writing activities were found to be negatively related to their writing performance. These included the frequency students wrote reports, personal/imaginative stories, and business writing, how regularly they brainstormed and worked with other students when writing, and how often they wrote one paragraph in math class. The frequency of students' revision and writing in English class was consistently found to be strongly positively related to their writing performance.

All variables regarding students' writing experiences were found to be significantly related to their performances. However, some of the students' writing experiences were found to be negatively related to their writing performances, including how frequently they had used computers from the beginning when writing papers, and whether teachers emphasized the importance of spelling/punctuation/grammar and length of papers in their grading. Among the positively related variables, whether teachers emphasized papers' quality or creativity and paper

organization in their grading was consistently found to have a strong positive relationship with students' writing performance

1.6 Ambiguity in Prompts and Genre-mastery Rubrics

Among 78 prompts, 11 prompts from seven states were considered ambiguous, and seven prompts from four states were considered to have implicit genre expectations. In total, 23% of prompts possessed one of the two problematic features: 14% of prompts were ambiguous, and 9% of prompts had implicit genre expectations.

Ambiguous prompts were mostly expository, narrative, argumentative, and informative prompts. Prompts with implicit expectations were mostly persuasive, expository, and argumentative prompts. Key words associated with ambiguity and implicit genre expectations include “explain,” “describe,” “essay,” “support,” “discuss,” and “tell.”

Among the 15 prompts that possessed these problematic features (i.e., ambiguity and implicit expectations), five prompts from three states were used with genre-mastery rubrics. In other words, these three states expected students to show their mastery of genres assessed (but not clearly or directly explained) in the prompts.

1.7 Genre Expectation in Standards and Genres Assessed

Among the seven genres assessed, the most widely referred to genre was narrative; it was referred to in 25 states' writing standards; this was followed by persuasive (24 states), expository (23 states), informative (22 states), descriptive (12 states), analytic (7 states) and argumentative (4 states). There were another 12 states whose standards implicitly referred to the argumentative genre by describing argumentative genre features without distinguishing argumentative from persuasive writings, and 11 states whose standards implicitly referred to features of literary analysis without labeling it as such. About 45% of state writing standards (12/27 states) covered

all the genres assessed in those states and referred to those genres more frequently than other genres; 33% of state writing standards (9/27 states) covered all the genres those states assessed but referred to those genres less frequently than some other genres. Around 22% of state writing standards (6/27 states) did not cover all of the genres that were assessed in the corresponding state writing assessments.

2. Implication for Writing Assessment Practices

2.1 For State Writing Assessment and NAEP

State assessments should be aligned with state standards to ensure that those genres assessed are also covered in state standards. Prompts should make their assessed genres more explicit. When states intend to evaluate students' genre-mastery skills, it is helpful to include specific genre components in their rubrics so that their expectations are more explicit to students, raters, and educators. Under the allowance of time and resources, more writing opportunities should be provided to students so that their writing abilities can be assessed more accurately. These recommendations are also applicable to the new CCSS-aligned K-12 assessments developed by the SBAC and the PARCC.

State and NAEP assessment differences play a role in students' performance on the NAEP. Students' NAEP performances are a result of many factors, including the similarity of students' home state assessments to the NAEP. When students' performances on NAEP are compared, we have to be aware of their different levels of preparedness as a result of their state writing assessments' similarities and differences with the NAEP.

Instead of focusing on the differences between state and NAEP assessments, both NAEP and states' assessments can move forward by incorporating more evidence-based writing assessment practices, which are likely to shrink the differences between states' and NAEP

assessments. As a result, students' performances on the NAEP are less likely to be impacted by their different levels of preparedness for NAEP tasks.

2.2 Writing Prompt Design

To make the assessed genres more explicit in writing prompts, the following practices are recommended:

- a) Include relevant demand verbs in prompts whenever possible.
- b) Make sure that the meanings of “demand verbs” in a) are evoked in conventional ways;
- c) When two or more “demand verbs” that signal different genres have to be used in the same prompt, how those rhetorical processes should work together to achieve a specified purpose should be specified;
- d) When demand verbs, which are often widely used in a variety of genres, are used on their own, there should be other supplemental information giving more details about genre expectations; and
- e) Use more concrete nouns and adjectives that signify genres in prompts.

3. Implication for Writing Instruction

Research has shown that process writing instruction, including gathering information, prewriting or planning, drafting, and editing, has a positive impact on students' writing qualities. Because some writing assessments also directed students to follow part of the writing process, teachers should continue to adopt a writing process approach for their instruction.

In addition to the writing process, teachers should also pay attention to the contextual factors in writing instruction. By 2007, only West Virginia had online writing sessions for their state direct writing assessments. However, the new generation of the assessments, which the two multi-state consortiums developed, is on computer. Thus, teachers can provide students with

more computer-based writing opportunities, as well as use research to inform their awareness of the impact of the word-processing software on students' writing qualities. The results of the research suggest that prompts did not often specify the genre expectations and rubrics tended to emphasize different aspects of writing construct. As a result, teachers can utilize rubrics in their writing instruction so that not only students have more explicit understanding of the writing expectations, but students can also learn to use rubrics to inform their planning of writing.

In terms of writing components, organization, structure, content, and detail were emphasized in almost all writing rubrics. Teachers can provide paragraph structure instruction and text structure instruction because research has shown that this kind of instruction is effective on students' writing qualities. Because word choice, sentence fluency, style, and grammar, including sentence construction, were also highly prized aspects of the students' papers, teachers can use text models to direct students to examine specific attributes of the texts and use sentence combining exercises to improve students' sentence construction and writing performance. Teachers should generally avoid traditional grammar instruction involving worksheets and decontextualized practice (Graham & Perin, 2007; Hillocks, 1984), but, instead, use students' own writing as examples in their instruction and provide students authentic editing opportunities.

General conventions, such as capitalization, punctuation, and spelling, were also assessed by the majority of states. These conventions should be taught in developmentally and instructionally appropriate ways. In terms of spelling, previously taught words should be reinforced in written work and reviewed periodically to promote retention. Students should be encouraged to correct their own capitalization, punctuation, and spelling mistakes after practice and assessment occasions.

Certainly, all these discussions do not suggest that teachers should teach to the test because large-scale writing assessments can only incorporate the measurable portion of writing constructs. Some expectations for writing performance in real life demands cannot be addressed in large-scale writing assessments due to all kinds of constraints. Those expectations that are addressed might still raise the question whether they can provide a valid and reliable assessment of students' writing abilities. For example, the integrated writing tasks are celebrated for its similarity to the writing tasks that students are likely to encounter in real life, but issues exist with their psychometric properties as how to distinguish students' reading and writing abilities in such tasks. Therefore, a constant struggle in test design is to balance the content dimension of the test and its psychometric dimension. Because of this limitation with large-scale assessments, teachers' instruction should definitely not only include the large-scale assessments' content and format, but also provide students with learning opportunities as a way to prepare for real life writing demands.

4. Next Steps for Research

More research is needed to investigate different methods of writing assessment, such as using integrated writing tasks. More research is also needed to study students' assessment behaviors, such as their interactions with writing prompts, especially the thinking processes that students adopt when reading writing prompts. Students' vocabulary understanding could be a potential area for future research using procedures such as think-aloud protocols and interviews.

Future research can be done to investigate the state-level difference when school- and teacher-level variables are entered as part of a multi-level model. The remaining large amount of unexplained variance between states found in this study suggests that there are still more state-level variables to be explored, such as alignments between states' standards and assessments, and

the stringency of states' accountability policies. Future research can also be conducted to examine how subgroups are affected by alignment variability and whether other factors in NAEP database might explain higher than expected achievement for students in subgroups.

Another potentially fruitful area for future research is to investigate whether state writing assessments are more likely to be fair assessments of students' writing abilities under the recommended circumstances—when those genres explicitly assessed in prompts are covered by state writing standards and genre-mastery rubrics are used to evaluate whether students' compositions accomplish the specified task demands. Moreover, experimental research can be conducted to examine connections between prompt design and student outcome within states using generalizability theory by varying aspects of prompt design.

It is hoped that these findings can advise test designers about what central characteristics of the writing construct have been valued in the past and can continue to be incorporated into future assessments, and what pitfalls are to avoid when designing writing prompts. It is also hoped that these findings can raise the general public's awareness that students' performances on the NAEP reflect both their writing abilities and how well they are prepared for the type of assessments the NAEP uses. Furthermore, it is hoped that these findings will draw the assessment and writing research communities' attention to validity-related issues in large-scale writing assessments and encourage more research to investigate components of these large-scale writing assessments in-depth.

APPENDICES

Appendix A Tables

Table 8 NAEP Coding & Frequency Counts and Percentage of States

	States' Frequency Counts (n) and Percentage (p)	G7 (N=15)		G8 (N=18)		Total (N=27)		G8
Strand	Indicators	n	p	n	p	n	p	NAEP
Writing Process	101 General Writing Process	1	0.067	3	0.167	4	0.148	
	102 Topic/Genre Selection	2	0.133	3	0.167	3	0.111	
	103 Gather Information	2	0.133	4	0.222	5	0.185	
	104 Pre-Writing/Planning	13	0.867	18	1	26	0.963	X
	105 Drafting Text	15	1	18	1	27	1	X
	106 Revising	9	0.6	9	0.5	15	0.556	X
	107 Editing	9	0.6	12	0.667	18	0.667	
	108 Publishing	8	0.533	4	0.222	10	0.37	
	109 Strategies	2	0.133	9	0.5	10	0.37	
Writing Context	201 Purpose	15	1	18	1	27	1	X
	202 Task	15	1	18	1	27	1	X
	203 Audience	14	0.933	13	0.722	20	0.741	X
	204 Collaboration		0		0		0	
	205 Sharing		0		0		0	
	206 Feedback		0		0		0	
	207 Text Models		0		0		0	
	208 Guidance/Support		0		0		0	
	209 Computer Technology	1	0.067		0	1	0.037	
	210 Procedural Facilitator	12	0.8	13	0.722	19	0.704	X
	211 Reference Materials	8	0.533	6	0.333	11	0.407	
	212 Source Materials	4	0.267	5	0.278	7	0.259	
	213 Disciplinary Context	1	0.067	1	0.056	2	0.074	
	214 Writing In/Writing Out of School		0		0		0	
	215 Length of Writing	13	0.867	16	0.889	25	0.926	
	216 Quantity of Writing	3	0.2	6	0.333	7	0.259	X
	217 Time for Writing	6	0.4	10	0.556	14	0.519	X
	218 Sophistication		0		0		0	
	401 General Organization	15	1	18	1	27	1	X
	402 General Structure	11	0.733	14	0.778	20	0.741	X

Table 8 (cont'd)

Writing Component	403 General Content	15	1	18	1	27	1	X
	404 Elaboration/Detail	14	0.933	18	1	26	0.963	X
	405 Genre Specific Organization & Content/Ideas				0		0	
	405A Narrative	3	0.2	3	0.167	5	0.185	X
	405B Expository	3	0.2	1	0.056	4	0.148	
	405C Persuasive	4	0.267	3	0.167	6	0.222	X
	405D Poetic		0		0		0	
	405E Response to Writing	2	0.133	2	0.111	3	0.111	
	406 Sentence Fluency	12	0.8	17	0.944	24	0.889	X
	407 Style	13	0.867	17	0.944	24	0.889	
	408 Figurative Language	4	0.267	6	0.333	7	0.259	
	409 Semantic Aspects	14	0.933	16	0.889	24	0.889	X
	410 Citations and References	1	0.067	1	0.056	1	0.037	
	411 Multimedia		0		0		0	
Writing Convention	501 General Conventions	9	0.6	16	0.889	22	0.815	
	502 Capitalization-General	11	0.733	12	0.667	19	0.704	
	503 Capitalization-Specific		0		0		0	
	503A Sentence Beginning		0		0		0	
	503B Word Level		0	1	0.056	1	0.037	
	503C Text Level		0		0		0	
	504 Punctuation-General	11	0.733	12	0.667	19	0.704	X
	505 Punctuation-Specific				0		0	
	505A Sentence Ending	1	0.067	4	0.222	4	0.148	
	505B Clausal Linking	1	0.067	4	0.222	4	0.148	
	505C Parenthetical		0		0		0	
	505D Word Level		0	1	0.056	1	0.037	
	506 Quotes/Dialogue		0		0		0	
	507 Handwriting-General		0		0		0	
	508 Handwriting-Manuscript		0		0		0	
	509 Handwriting-Cursive		0		0		0	
	510 Keyboarding		0		0		0	

Table 8 (cont'd)

	511 Spelling-General	10	0.667	12	0.667	18	0.667	X
	512 Spelling-Specific				0		0	
	512A Graphophonemic Elements		0	1	0.056	1	0.037	
	512B High-Frequency Words	2	0.133	5	0.278	6	0.222	
	512C Graphomorphemic Elements		0		0		0	
	512D Common Spelling Rules		0		0		0	
	512E Other Elements		0	1	0.056	1	0.037	
	513 Grammar-General	13	0.867	15	0.833	24	0.889	X
	514 Grammar-Specific				0		0	
	514A Nouns & Noun Phrases	2	0.133	2	0.111	3	0.111	
	514B Verbs & Verb Phrases	5	0.333	6	0.333	7	0.259	
	514C Pronouns & Pronominal Phrases	2	0.133	2	0.111	4	0.148	
	514D Adjectives	1	0.067	2	0.111	3	0.111	
	514E Adverbs		0	1	0.056	1	0.037	
	514F Modifiers	2	0.133	4	0.222	4	0.148	
	514G Sentence Construction	11	0.733	13	0.722	19	0.704	X
	515 Formatting-General	2	0.133	2	0.111	2	0.074	
	516 Formatting-Specific	6	0.4	8	0.444	12	0.444	
Writing Knowledge	601 Topic Knowledge	9	0.6	13	0.722	19	0.704	X
	602 Genre Knowledge		0		0		0	
	603 Linguistic Knowledge		0		0		0	
	604 Procedural Knowledge		0		0		0	
	605 Self-Regulation		0		0		0	

Table 9 Sample Sizes, Achievement, and Student Demographics, 27 State Grade 8 NAEP Reporting Sample

State	n	Weighted N	Mean Student Achievement	SE(Mean)	% Black	% Hispanics	% Asian	% American Indian	% Female	% LEP	% With IEPs	% Free/reduced-price lunch
Alabama	2710	55739	147.579	1.346	35.9%	2.2%	0.8%	0.3%	50.5%	1.4%	10.7%	50.1%
Arizona	2644	69384	148.227	1.441	5.8%	38.8%	2.7%	6.6%	49.1%	9.3%	7.9%	45.6%
Arkansas	2369	33196	150.634	1.162	23.8%	7.3%	1.2%	0.5%	48.2%	3.7%	11.3%	52.7%
California	8121	461402	147.889	0.971	7.3%	48.0%	11.8%	1.2%	48.3%	20.1%	7.9%	48.8%
Florida	3903	186141	158.042	1.313	23.0%	23.8%	2.5%	0.3%	49.5%	5.2%	12.3%	42.9%
Idaho	2807	20291	154.248	1.177	1.0%	12.8%	1.4%	1.5%	47.3%	4.9%	8.1%	38.7%
Illinois	3870	146929	159.927	1.489	19.1%	17.8%	4.5%	0.1%	48.7%	2.6%	12.2%	40.1%
Indiana	2623	77274	154.758	1.339	12.6%	6.4%	1.0%	0.2%	50.2%	2.5%	11.1%	34.9%
Kansas	2660	32160	156.263	1.386	8.1%	11.6%	1.9%	1.5%	49.8%	3.7%	10.3%	36.1%
Kentucky	2491	43056	151.443	1.373	10.2%	1.6%	1.0%	0.1%	50.4%	1.1%	8.1%	46.5%
Louisiana	2336	46721	146.693	1.258	43.6%	2.3%	1.2%	0.9%	48.2%	0.9%	11.5%	59.9%
Maine	2520	14596	161.034	1.066	1.6%	0.8%	1.5%	0.2%	49.4%	1.6%	15.0%	33.7%
Massachusetts	3437	64751	166.754	1.567	9.0%	10.5%	5.4%	0.2%	47.9%	3.5%	13.9%	26.5%
Michigan	2526	116199	151.058	1.338	18.6%	2.8%	2.3%	0.9%	49.8%	1.6%	11.2%	32.5%
Missouri	2776	69320	152.83	1.201	18.8%	2.6%	1.5%	0.2%	49.5%	1.7%	11.0%	37.5%
Nevada	2525	27139	143.094	1.046	10.5%	34.8%	8.2%	1.5%	49.2%	9.4%	10.7%	38.0%
New York	3647	199919	154.181	1.262	19.0%	17.9%	6.7%	0.3%	49.8%	4.1%	13.8%	48.0%
North Carolina	4042	101678	152.833	1.242	30.2%	7.1%	2.4%	1.3%	49.3%	3.7%	13.5%	44.0%
Oklahoma	2527	41091	152.789	1.2	9.4%	8.3%	2.2%	19.8%	49.5%	3.2%	12.8%	48.4%
Rhode Island	2566	11446	153.816	0.768	7.9%	17.2%	3.0%	0.5%	49.7%	2.8%	17.0%	31.4%
Tennessee	2725	71516	156.156	1.301	25.6%	4.7%	1.5%	0.1%	49.2%	1.8%	8.6%	45.3%
Texas	6783	278798	151.059	1.165	15.8%	43.8%	2.9%	0.2%	49.4%	6.4%	7.2%	50.4%
Vermont	1955	6679	161.534	1.078	1.6%	1.0%	1.4%	0.5%	47.2%	2.2%	16.7%	27.6%
Virginia	2631	84978	156.931	1.259	27.8%	5.7%	4.4%	0.2%	49.1%	2.8%	9.9%	26.9%
Washington	2840	73881	157.735	1.417	6.1%	12.7%	9.6%	2.4%	48.0%	4.7%	8.6%	34.4%
West Virginia	2818	21229	146.265	1.177	5.3%	0.8%	0.6%	0.1%	50.1%	0.6%	14.2%	47.4%
Wisconsin	2585	59616	157.71	1.411	9.6%	6.2%	3.3%	1.2%	49.0%	3.4%	11.3%	29.9%
Total	85437	2415129										

Note. The means and percentages reported are for the samples weighted to represent U.S. students.

Table 10 Comparison of Sample Sizes and Student Demographics for 27 State Grade 8 NAEP Reporting Sample and HLM Sample

State	n	Full Sample								n	HLM Sample							
		% Black	% Hispanics	% Asian	% American Indian	% Female	% ELLs	% With IEPs	% Free/reduced-price lunch		% Black	% Hispanics	% Asian	% American Indian	% Female	% ELLs	% With IEPs	% Free/reduced-price lunch
AL	2710	35.9	2.2	0.8	0.3	50.5	1.4	10.7	50.1	2360	33.2	1.9	0.8	0.4	51.1	1.1	9.2	47.6
AZ	2644	5.8	38.8	2.7	6.6	49.1	9.3	7.9	45.6	2199	5.4	37.9	2.6	6.5	49.8	8.7	7.1	42.6
AR	2369	23.8	7.3	1.2	0.5	48.2	3.7	11.3	52.7	2081	22.3	7.1	1.0	0.3	48.5	3.7	11.2	51.6
CA	8121	7.3	48.0	11.8	1.2	48.3	20.1	7.9	48.8	6361	6.4	46.4	12.9	1.3	50.3	18.4	6.6	47.4
FL	3903	23.0	23.8	2.5	0.3	49.5	5.2	12.3	42.9	3302	21.7	23.2	2.4	0.3	49.9	4.7	11.5	41.9
ID	2807	1.0	12.8	1.4	1.5	47.3	4.9	8.1	38.7	2460	1.0	12.8	1.5	1.6	48.9	5.1	8.0	38.8
IL	3870	19.1	17.8	4.5	0.1	48.7	2.6	12.2	40.1	3337	17.7	17.7	4.6	0.1	49.4	2.5	11.8	38.7
IN	2623	12.6	6.4	1.0	0.2	50.2	2.5	11.1	34.9	2309	11.5	5.9	1.2	0.2	50.4	2.3	10.6	33.5
KS	2660	8.1	11.6	1.9	1.5	49.8	3.7	10.3	36.1	2380	7.7	11.7	1.9	1.5	50.1	3.7	10.3	35.7
KY	2491	10.2	1.6	1.0	0.1	50.4	1.1	8.1	46.5	2251	9.9	1.6	1.0	0.0	50.9	0.9	8.0	46.5
LA	2336	43.6	2.3	1.2	0.9	48.2	0.9	11.5	59.9	2059	41.7	2.2	1.2	1.0	49.1	0.8	11.2	59.1
ME	2520	1.6	0.8	1.5	0.2	49.4	1.6	15.0	33.7	2243	1.5	0.7	1.4	0.2	49.9	1.5	14.1	33.0
MA	3437	9.0	10.5	5.4	0.2	47.9	3.5	13.9	26.5	2944	8.4	9.7	5.4	0.2	48.9	3.1	13.5	25.4
MI	2526	18.6	2.8	2.3	0.9	49.8	1.6	11.2	32.5	2195	17.1	2.6	2.4	0.9	50.3	1.5	10.8	31.3
MO	2776	18.8	2.6	1.5	0.2	49.5	1.7	11.0	37.5	2495	17.6	2.7	1.6	0.1	50.2	1.6	10.6	36.1
NV	2525	10.5	34.8	8.2	1.5	49.2	9.4	10.7	38.0	2136	9.4	33.3	8.8	1.6	51.0	8.4	9.2	36.7
NY	3647	19.0	17.9	6.7	0.3	49.8	4.1	13.8	48.0	3050	16.9	17.3	6.8	0.3	50.9	3.6	13.3	46.1
NC	4042	30.2	7.1	2.4	1.3	49.3	3.7	13.5	44.0	3452	28.0	6.9	2.4	1.3	50.2	3.7	13.7	42.5
OK	2527	9.4	8.3	2.2	19.8	49.5	3.2	12.8	48.4	2233	8.9	8.2	2.2	20.0	50.0	3.2	12.6	47.5
RI	2566	7.9	17.2	3.0	0.5	49.7	2.8	17.0	31.4	2248	7.6	16.6	3.0	0.5	50.4	2.2	16.0	30.5
TN	2725	25.6	4.7	1.5	0.1	49.2	1.8	8.6	45.3	2436	23.9	4.7	1.5	0.0	50.7	1.7	8.2	43.9
TX	6783	15.8	43.8	2.9	0.2	49.4	6.4	7.2	50.4	5951	15.3	43.3	3.1	0.2	49.9	5.7	6.6	49.3
VT	1955	1.6	1.0	1.4	0.5	47.2	2.2	16.7	27.6	1744	1.6	1.0	1.6	0.4	47.9	2.3	16.2	26.7
VA	2631	27.8	5.7	4.4	0.2	49.1	2.8	9.9	26.9	2301	27.3	5.6	4.6	0.2	49.7	2.9	9.7	26.7

Table 10 (cont'd)

WA	2840	6.1	12.7	9.6	2.4	48.0	4.7	8.6	34.4	2418	5.3	12.7	9.4	2.3	48.9	4.4	8.0	33.4
WV	2818	5.3	0.8	0.6	0.1	50.1	0.6	14.2	47.4	2537	4.8	0.9	0.7	0.2	50.8	0.7	13.7	46.7
WI	2585	9.6	6.2	3.3	1.2	49.0	3.4	11.3	29.9	2272	8.4	6.4	3.3	1.2	49.3	3.4	11.7	28.9
Total	85437									73754								

Note. The means and percentages reported are for the samples weighted to represent U.S. students.

Table 11 Raw Unweighted Descriptive Statistics of Variables in HLM Models

VARIABLE NAME (N=73754 Students from 27 States)	MEAN	SD	MIN	MAX
State level				
Distance between NAEP and state writing assessments	9.97	1.53	7.48	15.2
Student level				
Plausible Value 1	154.8	34.05	4.72	285.24
Plausible Value 2	154.8	34.12	0	284.44
Plausible Value 3	154.9	34.08	0	300
Plausible Value 4	154.9	34.18	0	283.28
Plausible Value 5	155.1	34.15	0	293.82
Black	0.16	0.36	0	1
Hispanic	0.18	0.38	0	1
Asian	0.04	0.2	0	1
American Indian	0.02	0.13	0	1
Female	0.5	0.5	0	1
English Language Learners (ELLs)	0.05	0.22	0	1
Individualized Education Plan (IEPs)	0.1	0.3	0	1
Free/Reduced-priced Lunch	0.44	0.5	0	1
Writing stories/letters is a favorite activity	2.17	0.94	1	4
Writing helps share ideas	2.6	0.89	1	4
How often teacher talk to you about writing	2.4	0.6	1	3
How often write thoughts/observation	2.34	1.24	1	4
How often write a simple summary	2.6	1.07	1	4
How often write a report	2.55	0.84	1	4
How often write an essay you analyze	2.53	0.93	1	4
How often write a letter/essay for school	2.38	0.92	1	4
How often write a story personal/imagine	2.43	0.96	1	4
How often write business writing	1.6	0.81	1	4
How often when writing-get brainstorm	1.9	0.62	1	3
How often when writing-organize papers	2.21	0.74	1	3
How often when writing-make changes	2.6	0.59	1	3
How often when writing-work with other students	2.09	0.68	1	3
Write paper-use computer from begin	1.97	0.74	1	3
Write paper for school-use computer for changes	2.24	0.75	1	3
Write paper for school-use computer for internet	2.49	0.63	1	3
How often write one paragraph in English class	3.56	0.76	1	4
How often write one paragraph in science class	2.86	1.01	1	4
How often write one paragraph in social studies/history class	3.13	0.95	1	4
How often write one paragraph in math class	1.98	1.1	1	4
How often teacher asks to write more than 1 draft	2.26	0.63	1	3
Teacher grades important for spelling/ punctuation/ grammar	2.59	0.57	1	3

Table 11 (cont'd)

Teacher grades important for paper organization	2.55	0.59	1	3
Teacher grades important for quality/creativity	2.6	0.57	1	3
Teacher grades important for length of paper	2.09	0.65	1	3
Difficulty of this writing test	1.47	0.71	1	4
Effort on this writing test	2.05	0.81	1	4
Importance of success on this writing test	2.67	1	1	4

Table 12 Genre Expectations in Standards and Genre Assessed

State	Grade	Genre Expectations in Standards	% Total Genre Occurrences	Genre Assessed	% Genre Assessed
AL	7	Respond Narrative Poetic Express Exchange Expository Describe Research	9.1% 9.1% 9.1% 27.3% 9.1% 9.1% 9.1% 9.1%	Descriptive Expository Narrative Persuasive ^a	38%
AR	7	Respond Narrative Poetic Persuade Expository Describe Summarize Reflect Research	10% 20% 5% 10% 20% 15% 5% 5% 10%	Persuasive Expository	22%
	8	Respond Narrative Poetic Persuade Expository Describe Reflect Research	12.5% 12.5% 6.3% 18.8% 18.8% 12.5% 6.3% 12.5%	Expository	13%
AZ	7	Record Respond Direct Narrative Poetic Exchange Persuade Expository Inform Describe Summarize Functional	3.7% 3.7% 3.7% 14.8% 3.7% 14.8% 11.1% 11.1% 14.8% 7.4% 7.4% 3.7%	Informative	8%
	8	Record Respond Direct Narrative Poetic Exchange Persuade Expository Inform Describe Summarize Functional	3.8% 3.8% 3.8% 15.4% 3.8% 15.4% 7.7% 11.5% 15.4% 7.7% 7.7% 3.8%	Narrative	8%

Table 12 (cont'd)

CA	7	Respond	14.3%	Narrative Persuasive Analysis Informative (Summary)	57%
		Narrative	14.3%		
		Persuade	14.3%		
		Expository	14.3%		
		Describe	14.3%		
		Summarize	14.3%		
		Research	14.3%		
		Argumentative*	-----		
FL (1996)	6-8	Record	13.6%	Expository Persuasive	20%
		Respond	22.7%		
		Direct	4.5%		
		Narrative	9.1%		
		Express	9.1%		
		Exchange	4.5%		
		Persuade	9.1%		
		Expository	13.6%		
		Inform	9.1%		
		Reflect	4.5%		
		Argumentative*	-----		
		Analysis*	-----		
FL (2007)	8	Record	5.6%	Expository Persuasive	17%
		Remind	5.6%		
		Direct	16.7%		
		Narrative	11.1%		
		Poetic	5.6%		
		Express	5.6%		
		Exchange	5.6%		
		Persuade	5.6%		
		Expository	11.1%		
		Inform	16.7%		
		Summarize	5.6%		
		Research	5.6%		
		Argumentative*	-----		
ID	7	Respond	7.7%	Expository	11%
		Direct	7.7%		
		Express	7.7%		
		Persuade	15.4%		
		Expository	15.4%		
		Inform	23.1%		
		Analyze	7.7%		
		Evaluate	7.7%		
		Research	7.7%		
IL	8	Record	1.6%	Narrative Persuasive	13%
		Remind	1.6%		
		Direct	1.6%		
		Narrative	15.6%		
		Poetic	7.8%		
		Express	1.6%		
		Exchange	10.9%		
		Persuade	14.1%		
		Expository	10.9%		
		Inform	17.2%		
		Analyze	3.1%		
		Synthesize	3.1%		

Table 12 (cont'd)

IN	7	Evaluate	4.7%	Narrative Persuasive Analysis	30%
		Research	1.6%		
		Functional	3.1%		
		Argumentative	1.6%		
		Remind	7.1%		
		Respond	7.1%		
		Narrative	14.3%		
		Exchange	14.3%		
		Persuade	7.1%		
		Expository	14.3%		
		Inform	14.3%		
		Describe	7.1%		
		Summarize	7.1%		
		Research	7.1%		
	8	Argumentative*	-----	Expository Narrative	17%
		Analysis*	-----		
		Remind	5.9%		
		Respond	5.9%		
		Direct	5.9%		
		Narrative	11.8%		
		Exchange	17.6%		
		Persuade	5.9%		
		Expository	11.8%		
		Inform	11.8%		
		Describe	5.9%		
		Synthesize	5.9%		
		Summarize	5.9%		
		Research	5.9%		
KS	8	Argumentative*	-----	Expository Informative	33%
		Analysis*	-----		
		Direct	21.4%		
		Narrative	7.1%		
		Exchange	7.1%		
		Persuade	35.7%		
		Expository	7.1%		
KY (1999)	7	Inform	21.4%	Persuasive ^a Narrative ^a	0%
		Argumentative*	-----		
		Record	16.7%		
		Respond	16.7%		
		Express	16.7%		
	8	Summarize	16.7%	Persuasive ^a Expository ^a	0%
		Reflect	33.3%		
		Respond	25%		
		Synthesize	25%		
		Reflect	50%		
KY (2006)	7	Respond	4.8%	Persuasive Narrative	13%
		Narrative	4.8%		
		Poetic	4.8%		
		Express	14.3%		
		Exchange	4.8%		
		Persuade	4.8%		
		Expository	9.5%		
		Inform	14.3%		
		Describe	4.8%		

Table 12 (cont'd)

		Analyze	4.8%		
		Synthesize	4.8%		
		Summarize	4.8%		
		Reflect	9.5%		
		Research	4.8%		
		Functional	4.8%		
	8	Respond	5.3%	Persuasive	13%
		Narrative	5.3%	Expository	
		Poetic	5.3%		
		Express	15.8%		
		Exchange	5.3%		
		Persuade	5.3%		
		Expository	5.3%		
		Inform	10.5%		
		Analyze	5.3%		
		Synthesize	5.3%		
		Summarize	5.3%		
		Reflect	10.5%		
		Evaluate	5.3%		
		Research	5.3%		
		Functional	5.3%		
LA	7	Respond	18.8%	Narrative	18%
		Narrative	12.5%	Expository	
		Exchange	6.3%		
		Persuade	12.5%		
		Expository	6.3%		
		Inform	6.3%		
		Describe	6.3%		
		Analyze	6.3%		
		Evaluate	6.3%		
		Research	12.5%		
		Functional	6.3%		
		Argumentative*	-----		
	8	Respond	7.1%	Narrative	22%
		Narrative	14.3%	Expository	
		Exchange	7.1%		
		Persuade	14.3%		
		Expository	14.3%		
		Describe	7.1%		
		Analyze	14.3%		
		Evaluate	7.1%		
		Research	14.3%		
		Argumentative*	-----		
MA	7	Respond	10.0%	Expository	17%
		Narrative	20.0%		
		Poetic	10.0%		
		Expository	30.0%		
		Inform	20.0%		
		Research	10.0%		
		Analysis*	-----		
ME	5-8	Narrative	27.3%	Persuasive	22%
		Express	9.1%	Expository	
		Exchange	9.1%	Descriptive ^a	
		Persuade	9.1%		

Table 12 (cont'd)

		Expository	9.1%		
		Inform	9.1%		
		Summarize	9.1%		
		Reflect	9.1%		
		Research	9.1%		
		Argumentative*	-----		
MI	6-8	Respond	4.8%	Persuasive	44%
		Narrative	19.0%	Narrative	
		Poetic	14.3%	Argumentative	
		Persuade	23.8%	Expository	
		Expository	4.8%		
		Inform	19.0%		
		Synthesize	4.8%		
		Reflect	4.8%		
		Research	4.8%		
		Argumentative*	-----		
MO	7	Respond	12.5%	Expository	14%
		Narrative	12.5%		
		Exchange	25%		
		Persuade	12.5%		
		Expository	12.5%		
		Describe	12.5%		
		Summarize	12.5%		
		Argumentative*	-----		
		Analysis*			
NC	7	Narrative	12.5%	Persuasive ^a	0%
		Express	12.5%		
		Expository	12.5%		
		Inform	12.5%		
		Analyze	12.5%		
		Reflect	12.5%		
		Evaluate	25%		
RI	8	Respond	7.4%	Analysis	33%
		Direct	11.1%	Persuasive	
		Narrative	18.5%	Informative	
		Poetic	11.1%		
		Persuade	11.1%		
		Expository	7.4%		
		Inform	22.2%		
		Describe	7.4%		
		Reflect	3.7%		
		Analysis*	-----		
VT	8	Respond	15.4%	Analysis	60%
		Direct	23.1%	Persuasive	
		Narrative	15.4%	Informative	
		Persuade	23.1%		
		Inform	23.1%		
		Analysis*	-----		
NV	8	Respond	9.1%	Narrative	10%
		Narrative	9.1%		
		Exchange	9.1%		
		Persuade	9.1%		
		Inform	9.1%		
		Describe	9.1%		

Table 12 (cont'd)

NY	8	Summarize	9.1%	Expository Analysis	18%
		Evaluate	9.1%		
		Research	18.2%		
		Functional	9.1%		
		Argumentative*	-----		
		Analysis*	-----		
		Record	4.2%		
		Respond	20.8%		
		Narrative	8.3%		
		Poetic	8.3%		
OK	8	Exchange	12.5%	Argumentative Expository ^a	8%
		Expository	4.2%		
		Inform	12.5%		
		Analyze	16.7%		
		Summarize	4.2%		
		Research	4.2%		
		Argumentative	4.2%		
		Record	5.9%		
		Respond	5.9%		
		Direct	5.9%		
TN	8	Narrative	11.8%	Expository	6%
		Exchange	11.8%		
		Persuade	5.9%		
		Inform	5.9%		
		Synthesize	5.9%		
		Summarize	17.6%		
		Reflect	5.9%		
		Evaluate	5.9%		
		Research	11.8%		
		Argumentative*	-----		
TX	7	Analysis*	-----	Narrative	6%
		Draw	2.9%		
		Record	2.9%		
		Respond	11.8%		
		Direct	5.9%		
		Narrative	8.8%		
		Poetic	5.9%		
		Express	2.9%		
		Exchange	5.9%		
		Persuade	8.8%		
TX	7	Expository	14.7%	Narrative	6%
		Inform	5.9%		
		Describe	5.9%		
		Synthesize	5.9%		
		Reflect	5.9%		
		Research	2.9%		
		Functional	2.9%		
		Argumentative*	-----		
		Analysis*	-----		
		Draw	5.6%		
TX	7	Record	16.7%	Narrative	6%
		Respond	2.8%		
		Direct	5.6%		
		Request	2.8%		

Table 12 (cont'd)

		Narrative	8.3%		
		Poetic	5.6%		
		Express	11.1%		
		Exchange	8.3%		
		Persuade	2.8%		
		Expository	2.8%		
		Inform	11.1%		
		Describe	2.8%		
		Summarize	2.8%		
		Reflect	2.8%		
		Evaluate	2.8%		
		Research	2.8%		
		Argumentative	2.8%		
		Analysis*	-----		
VA	8	Narrative	25%	Argumentative ^a	0%
		Persuade	25%		
		Expository	25%		
		Inform	25%		
WA	7	Record	4.4%	Expository	11%
		Remind	1.5%	Persuasive	
		Respond	2.9%		
		Direct	2.9%		
		Narrative	11.8%		
		Poetic	8.8%		
		Express	4.4%		
		Exchange	4.4%		
		Persuade	20.6%		
		Expository	10.3%		
		Inform	13.2%		
		Describe	1.5%		
		Analyze	1.5%		
		Reflect	2.9%		
		Evaluate	1.5%		
		Research	4.4%		
		Functional	1.5%		
		Argumentative	1.5%		
WI	5-8	Respond	16.7%	Persuasive	20%
		Narrative	33.3%		
		Exchange	16.7%		
		Persuade	16.7%		
		Expository	16.7%		
		Argumentative*	-----		
		Analysis*	-----		
WV	7	Poetic	11.1%	Descriptive ^a	29%
		Express	11.1%	Persuasive	
		Exchange	11.1%	Narrative ^a	
		Persuade	11.1%	Expository	
		Expository	11.1%		
		Inform	22.2%		
		Research	22.2%		

Note: *genres potentially covered by state standards.

^aassessed genres not covered by state standards.

Appendix B Coding Taxonomies

Table 13 Prompt Coding—Troia & Olinghouse’s (2010) Coding Taxonomy

100s Writing Processes: Any aspect of the stages or specific strategies that one uses when producing a piece of writing Guiding Question: Is this something that relates to the writer’s <i>actions</i> in composing the text? Actions are things that the writer does. Actions are differentiated from the purpose guiding those actions, the products of those actions, or the knowledge required to initiate those actions.		
Indicator	Definition	Examples
101	General Writing Process: A general reference to the writing process	proceed through the writing process, produce a well written paper using the writing process, the process of writing
102	Topic/Genre Selection: The process of determining the general topic, theme, focus, point of view, or genre of the writing	[Prewrite] establish a controlling idea or focus, generate and narrow topics,
103	Gather Information: The process of collecting relevant information as it pertains to the topic	Develop a comprehensive and flexible search plan, selecting appropriate information to set context, research (for the purpose of gathering information)
104	Pre-Writing/Planning: The process of using activities prior to writing to generate, structure, or organize content	outlining, brainstorming, [Prewrite] generating ideas, [Prewrite] organize ideas,
105	Drafting Text: The process of producing written text that is later expected to be altered	Draft: complete a draft demonstrating connections among ideas,
106	Revising: The process of altering existing text in order to better achieve communicative aims with content, organization, and style	Revise, rewrite (if clear that changes are being made to draft),
107	Editing: The process of altering existing text to better match expectations for writing conventions	Proofreading, revise for spelling, revise for capitalization, revise for punctuation,
108	Publishing: The process of preparing the final form of a text possibly for public distribution	final copy, final draft, final product
109	Strategies: The process of using steps or supports in order to problem solve during the writing process	re-reading, time management, test-taking
200s Writing Context: The social, physical, or functional circumstances outside the writer that influence text production. Guiding Question: Is this something that is located <i>outside</i> the writer’s text and <i>outside</i> the writer’s mind?		
201	Purpose: General reference to the objective or intent in creating a piece of writing	
202	Task: General reference to the writing task	given the writing task, writing is appropriate for the task at hand, writing in different genres,

Table 13 (cont'd)

		appropriate for the given topic, format requirements, context
203	Audience: General reference to a reader or readers for a piece of writing	
204	Collaboration: Cooperatively working with others to produce a piece of writing	
205	Sharing: Telling or showing ideas, plans, or a piece of writing to others that may or may not elicit a response; sharing can occur at any point during the writing process	tell a peer ideas for writing
206	Feedback: Verbal or written information in response to an author's work at any point in the writing process received from peers or adults	peer conferencing to elicit suggestions for improvement
207	Text Models: Examples of structures, forms, or features used as explicit cues for text production	Use literary models to refine writing style.
208	Guidance/Support: Verbal or written assistance, aside from feedback, tailored to the needs of students during writing from peers or adults	with the help of peers, with teacher modeling, with assistance, in response to a prompt or cue, using dictation
209	Computer Technology: Using a computer as a tool in the process of writing	digital tools, use appropriate technology to create a final draft.
210	Procedural Facilitator: External material used to support the process of writing,	rubric, checklist, graphic organizer, story map
211	Reference Materials: Sources of information consulted to support writing mechanics and formatting	dictionaries, thesauruses, style manual
212	Source Materials: Reference to source materials that are integrated into the written content	web sites, articles, texts, documents, encyclopedic entries
213	Disciplinary Context: The general or particular academic setting (content area) in which a piece of writing is produced is specified	writing across the curriculum, writing for a range of discipline specific tasks, writing a procedural text in science, writing in the content areas
214	Writing In/Writing Out of School: The general place in which a piece of writing is produced is specified	
215	Length of Writing: Length of a piece of writing is specified	Brief, multi-page, short, long, # paragraphs specified
216	Quantity of Writing: The number of pieces of writing is specified	portfolio, several, numerous
217	Time for Writing: Duration and/or frequency of sustained student writing is specified	60 minutes, over two sessions, routinely
218	Sophistication: Expectations for complexity in a given text	multiple perspectives, sensitivity to cultural diversity
300s Writing Purposes: The variety of communicative intentions that can be accomplished through many different genres. Guiding Question: Is this something that relates to why the writer is writing and does not appear in the actual text?		

Table 13 (cont'd)

301	Draw: Producing a picture or diagram for the purpose of communicating	illustration, picture, diagram, drawing
302	Record: Copying text or taking notes on information	note taking, copy
303	Remind: Bringing attention to something for the purpose of recall	reminder, list
304	Respond: Responding to a stimulus , such as a question, prompt, or text, through writing	response, personal response, on-demand writing, text as stimulus for writing something new, response to literature
305	Direct: Giving directions, commands, or procedures	how-to, procedure, instructions, manual, technical text
306	Request: Asking for information or action	request, solicitation
307	Entertain/Narrate: Giving an account, either fictional or factual that often provides amusement and enjoyment	narrative, personal narrative, story, memoir, recount, biography, autobiography, fiction, fantasy, fable, folktale, myth, legend, adventure, mystery, tall tale, fairytale, drama, short story
308	Poetic: Evoking imagination or emotion through intentional manipulation of form and language	poetry, free verse, haiku, lyric, ballad, rhyme, sonnet, couplet, cinquain, limerick, dactyl, ode
309	Express: Conveying thoughts, feelings, or beliefs for personal reasons	journal writing, diary writing
310	Exchange: Conveying thoughts, feelings, or beliefs for social reasons	email, blog, letter, editorial
311	Persuade: Convincing an identified audience to act on a specific issue	persuasive essay,
312	Exposit/Explain: Explaining, clarifying, or expounding on a topic ; this may be done generally or in depth through elaboration	explanation, essay, exposition
313	Inform: Giving facts about a subject which may or may not be integrated	informational piece, article, report
314	Describe: Giving details/attributes about an object or event	description, descriptive text
315	Analyze: Systematically and intentionally examining something through details and structure	critique, literary criticism
316	Synthesize: Combining various things into one coherent, novel whole	synthesis, lab report,
317	Summarize: Using a brief statement or paraphrase to give the main points	summary, synopsis, paraphrase
318	Reflect: Thinking deeply and carefully about something after the fact, often using written text to learn	reflections, reflective writing, writing-to-learn
319	Evaluate: Examining the match between others' writing intent and form using criteria	book review
320	Research: Using systematic investigation to obtain information/knowledge for a piece of writing	experiments
321	Functional: Completing forms, applications , and other fill-in types of documents	checks, resumes

600s

Writing Metacognition & Knowledge: Knowledge resources **within the writer** that are drawn upon to compose a written text and/or knowledge that is the focus of development during

Table 13 (cont'd)

instruction (explicit reference to knowledge, recognition, distinguishing, identifying, recognizing, learning, or understanding must be made) or reflection on the knowledge one possesses. Guiding Question: Is this something that is happening in the student's mind (e.g., thinking about or analyzing their writing)? If it is something that the student is doing, or that is revealed in their writing, it cannot be a 600.		
601	Topic Knowledge: Knowledge of facts, information, or experiences pertaining to a particular subject that are within the writer and used to compose a written text	use personal experience to develop content for an essay, through experimentation, develop knowledge about natural phenomena for writing
602	Genre Knowledge: Knowledge about the purposes of writing and/or the macrostructures of a text that are within the writer and used to compose a written text	text attributes, elements, structure common to specific types of writing
603	Linguistic Knowledge: Knowledge of the microstructures of a text that are within the writer and used to compose a written text	sound-symbol relationships, spelling rules, grammatical rules, vocabulary
604	Procedural Knowledge: Knowledge of the procedures or processes of writing that are within the writer and used to compose a written text	knowledge of how to plan or revise, knowledge of how to use specific things during the writing process (e.g., knowing how to use a dictionary)
605	Self-Regulation: The process of explicitly managing, reflecting upon, and/or evaluating one's behaviors, performance, thoughts, or feelings	

Table 14 Rubric Coding—Troia and Olinghouse’s (2010) Coding Taxonomy

400s Writing Components: Features, forms, elements, or characteristics of text observed in the written product Guiding Question: Is this something that you can observe in the text itself? Is this something you can see without the writer(s) being present?		
Indicator	Definition	Examples
401	General Organization: How written content for a whole text is organized to achieve an intended purpose	• Order and Organization ○ out of order ○ writing progresses in an order that enhances meaning ○ logical organization ○ progression of text may be confusing or unclear • Unifying theme • Clear structure • Coherence • Central idea • Controlling idea
402	General Structure: Portions of a text that bridge content and organization through structural representation	• Introduction • Beginning • Middle • End • Conclusion ○ beginning, middle, end may be weak or absent
403	General Content: Topical information or subject matter presented within the text or content that is a more specific example of a structural representation	• Ideas and content ○ topic/idea <i>development</i> ○ ideas are fresh, original, or insightful ○ content goes beyond obvious • References to the topic ○ the writer defines the topic ○ topic may be defined, but not developed • Main idea ○ the writer states main idea ○ writing lacks main idea • Topic sentence • Information is very limited • Control of topic • Establishing a context for reading
401/403	Rubric descriptors that will receive both a general organization code [401] and a general content code [403].	• References to addressing the task ○ fully accomplishes the task ○ accomplishes the task ○ minimally accomplishes the task ○ does not accomplish the task ○ addresses all parts of the writing task • References to addressing the prompt ○ addresses all of the specific points in the prompt

Table 14 (cont'd)

		○ addresses most of the points in the prompt • References to purpose ○ demonstrates a clear understanding of purpose ○ demonstrates a general understanding of purpose ○ demonstrates little understanding of purpose • References to addressing/awareness of genre ○ response is appropriate to the assigned genre ○ uses genre-appropriate strategies ○ response does not demonstrate genre awareness ○ organization appropriate to genre ○ awareness of genre/purpose • Organizing Ideas ○ ideas are organized logically ○ meaningful relationships among ideas ○ related ideas are grouped together ○ ideas go off in several directions ○ ideas may be out of order ○ writing does not go off on tangents • Focus ○ stays focused on topic and task ○ may lose focus ○ lapse of focus ○ writing may go off in several directions ○ the writing is exceptionally clear and focused ○ consistent focus on the assigned topic, genre, and purpose ○ sustained focus and purpose ○ stays fully focused on topic/purpose ○ sustained or consistent focus on topic ○ clarity, focus, and control ○ sustained focus on content ○ maintains consistent focus on topic ○ clear focus maintained for intended audience
404	Elaboration/Detail: Information that illustrates, illuminates, extends, or embellishes general content	• Details ○ supporting details are relevant ○ writer makes general observations without specific details ○ examples, facts, and details ○ concrete details ○ minimal details ○ omits details ○ includes unrelated details ○ list of unrelated specifics without extensions ○ anecdotes

Table 14 (cont'd)

		○ Elaborate/ elaborated/ elaboration/ elaborating ideas that are fully and consistently elaborated ○ minimal elaboration
405	Genre Specific Organization & Content/Ideas: Structural elements and/or information that is canonical for a specific genre	
405A	Narrative	• Story line • Plot • Dialogue • Setting • Characters • Goals • Tells a story • Events • Sequence of events ○ thoroughly developed sequence of significant events ○ lacks a sequence of events • Reactions • Structure showing a sequence through time • Chronology • Chronological sequence of ideas
405B	Expository/ Procedural/ Descriptive/ Informational	• References to canonical text structures of the genre ○ cause/effect ○ similarity and difference ○ compare/contrast
405C	Persuasive	• Thesis • Anticipates reader's questions • Supports an opinion • Question and answer • Reasons • Points • Sub-points • Position ○ maintains position/logic throughout ○ subject/position (or issue) is clear, identified by at least an opening statement ○ subject/position is vague ○ subject/position (or issue) is absent ○ defends a position • Evidence
405D	Poetic	• Rhyme
405E	Response to Writing	• Connections to experience or texts • Interpretation • Connects text to self, the outside world, or another text • Supports a position in response to the text

Table 14 (cont'd)

		• Demonstrates understanding of literary work ○ demonstrates clear understanding of literary work ○ demonstrates a limited understanding of literary work ○ demonstrates little understanding of literary work • Supports judgments about text ○ provides effective support for judgments through specific references to text and prior knowledge ○ provides some support for judgments through references to text and prior knowledge ○ provides weak support for judgments about text ○ fails to provide support for judgments about text Interpretation
406	Sentence Fluency: The variety, appropriateness, and use of sentences in the text	• Sentence variety ○ variety of sentence structures ○ sentences vary in length and structure ○ uses an effective variety of sentence beginnings, structures, and lengths ○ includes no sentence variety ○ writer uses varied sentence patterns ○ sentences are purposeful and build upon each other
407	Style: Language intentionally used to enhance purposes, forms, and features	• Style • Voice • Tone • Register ○ writer chooses appropriate register to suit task • Repetition ○ writing is repetitive, predictable, or dull ○ reader senses person behind the words • Audience ○ reader feels interaction with writer ○ indicates a strong awareness of audience's needs ○ communicates effectively with audience ○ displays some sense of audience ○ some attention to audience ○ little or no awareness of audience • Language ○ writer effectively adjusts language and tone to task and purpose ○ language is natural and thought-provoking ○ attempts at colorful language often come close to the mark, but may seem overdone or out of place

Table 14 (cont'd)

		○ vivid, precise, and engaging language that is appropriate to the genre ○ writer uses language that is easy to read ○ writer uses language that is difficult to read
408	Figurative Language: Words, phrases or devices used to represent non-literal connections to objects, events, or ideas	• Metaphor • Simile • Personification • Symbolism • Hyperbole • Onomatopoeia • Imagery
409	Semantic Aspects: Words, phrases, or devices used to enhance the meaning of the text from a literal standpoint	• Word Choice ○ words are accurate and specific ○ uses different beginning words for sentences • Transitions ○ ideas are connected with transitions ○ varied transitions ○ paper is linked with transitions ○ smooth transitions between ideas, sentences, and paragraphs ○ connectives • Vocabulary ○ accurate, precise vocabulary ○ chooses vocabulary precisely ○ control of challenging vocabulary ○ academic words ○ domain-specific vocabulary ○ technical vocabulary • Descriptive words ○ descriptive language ○ rich description • Imagery • Humor • Synonyms • Sensory details
410	Citations and References: Attributions for contributed or borrowed material for writing, including quotations	
411	Multimedia: The integration of various mediums of expression or communication as part of writing , including illustrations, photos, video, sound, and digital archival sources to accomplish communicative aims that could not be accomplished using any single medium	

Table 15 Rubric Coding—Jeffery’s (2009) Coding Taxonomy

Rubric Types	Definition	Examples
Rhetorical	Focusing on the relationship between writer, audience, and purpose across criteria domains, and containing terms framed within the context of appropriateness, effectiveness, and rhetorical purpose	• Successfully addresses and controls the writing task with a strong sense of audience and purpose ○ reader ○ audience ○ purposefully ○ effectively ○ appropriately • The writing is focused and purposeful, and it reflects insight into the writing situation ○ the writing situation ○ the rhetorical context
Genre-mastery	Emphasizing criteria specific to the genre students are expected to produce by identifying a specific rhetorical purpose, such as to convince an audience to take action or to engage an audience with a story, and varying rubric content to match prompt types, as well as containing terms framed by the specific communicative purpose that characterize the genre	• A persuasive composition states and maintains a position, authoritatively defends that position with precise and relevant evidence, and convincingly addresses the readers concerns, biases, and expectations ○ “logically” and “clearly” with persuasive or argumentative writing • Clarifies and defends or persuades with precise and relevant evidence; clearly defines and frames issues
Formal	Conceptualizing proficiency in terms of text features not specific to any writing context with features not framed by any particular considerations, such as the author’s thinking or creativity, and with characteristics that might be applicable to a variety of writing contexts, as well as defining good writing in relatively broad terms by focusing on features such as coherence, development and organization	• Is well organized and coherently developed; clearly explains or illustrates key ideas; demonstrate syntactic variety
Cognitive	Targeting thinking processes such as reasoning and critical thinking across domains, and explicitly valuing clarity of ideas, logical sequencing, and other features that implicate students’ cognitions	• A typical essay effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking ○ Explicit emphasis on “critical thinking”
Expressive	Emphasizing writing as a product of the author’s processes, especially creativity, and conceptualizing “good writing” as an expression of the author’s uniqueness, individuality, sincerity, and apparent commitment to the task, as well as containing terms framed by an overarching concern with personality and perspective	• Approach the topic from an unusual perspective, use his/her unique experiences or view of the world as a basis for writing, or make interesting connections between ideas ○ Interesting connection between ideas

Table 16 Seven-Genre Coding Scheme for Prompts—Adapted from Jeffery (2009) and Troia & Olinghouse (2010)

Genre Categories	Characteristics
(P) Persuasive	- Directed students to convince or persuade an audience - Identified a local audience as target for persuasion - Often specified a form for persuasion (e.g. letter, newspaper article, speech) - Specified a relatively concrete issue with clear implications (e.g. attendance policy) - Called for one-sided perspective (did not invite consideration of multiple perspectives) - Key terms: “convince”, “persuade”, “agree or disagree”, “opinion”
(A) Argumentative	- Directed students to argue a position on an issue - Did not identify a specific audience - Did not specify form - Addressed relatively abstract philosophical issue without clear implications - Called for consideration of multiple perspectives - Key terms: “position”, “point of view”
(N) Narrative	- Directed students to tell real or imagined stories - Sometimes directed students to connect stories to themes (e.g. provided quotation) - Did not identify a context (e.g. audience) for writing - Might direct the student to engage the reader - Used words like “event”, “experience” or “a time” to evoke memories - Key terms: “tell”, “describe”, “story”, “narrative”, “imagination”
(E) Explanatory	- Directed students to explain why something is so or what is so - Might present arguable propositions as inarguable (e.g. importance of homework) - Do not explicitly identify a proposition as arguable - But may allow for choice (e.g. explain qualities are important in a sport) - Might include language consistent with argument or persuasion (e.g. “support”) - Typically asked students to address relatively abstract concepts - Typically do not identify a target audience - Key terms: “explain”, “what”, “why”
(I) Informative	Directed students to explain a process or report on concrete, factual information

Table 16 (cont'd)

	• Similar to Explanatory in except for object of explanation (relatively concrete) • Typically do not identify a target audience • Key terms: “explain”, “how”, “procedure”
(AN) Analytic	• Directed students to analyze pieces of literature • Did not identify a target audience • May provide pieces of literature for analysis • Included discipline-specific language and • Referred the work’s author or speaker • Key terms: “describe”, “show”, “author”, “elements”
(D) Descriptive	• Direct students to give details/attributes about an object or event • Key terms: “describe”, “description”, “descriptive text”

Table 17 Standards Genre Coding—Troia and Olinghouse’s (2010) Coding Taxonomy Modified to Accommodate Jeffery’s (2009) Genre Coding Taxonomy

300s Writing Purposes: The variety of communicative intentions that can be accomplished through many different genres . Guiding Question: Is this something that relates to why the writer is writing and does not appear in the actual text?		
Indicator	Definition	Examples
301	Draw: Producing a picture or diagram for the purpose of communicating	illustration, picture, diagram, drawing
302	Record: Copying text or taking notes on information	note taking, copy
303	Remind: Bringing attention to something for the purpose of recall	reminder, list
304	Respond: Responding to a stimulus , such as a question, prompt, or text, through writing	response, personal response, on-demand writing, text as stimulus for writing something new, response to literature
305	Direct: Giving directions, commands, or procedures	how-to, procedure, instructions, manual, technical text
306	Request: Asking for information or action	request, solicitation
307	Entertain/Narrate: Giving an account, either fictional or factual that often provides amusement and enjoyment	narrative, personal narrative, story, memoir, recount, biography, autobiography, fiction, fantasy, fable, folktale, myth, legend, adventure, mystery, tall tale, fairytale, drama, short story
308	Poetic: Evoking imagination or emotion through intentional manipulation of form and language	poetry, free verse, haiku, lyric, ballad, rhyme, sonnet, couplet, cinquain, limerick, dactyl, ode
309	Express: Conveying thoughts, feelings, or beliefs for personal reasons	journal writing, diary writing
310	Exchange: Conveying thoughts, feelings, or beliefs for social reasons	email, blog, letter, editorial
311	Persuade: Convincing an identified audience to act on a specific issue	persuasive essay,
312	Exposit/Explain: Explaining, clarifying, or expounding on a topic ; this may be done generally or in depth through elaboration	explanation, essay, exposition
313	Inform: Giving facts about a subject which may or may not be integrated	informational piece, article, report
314	Describe: Giving details/attributes about an object or event	description, descriptive text
315	Analyze: Systematically and intentionally examining something through details and structure	critique, literary criticism
316	Synthesize: Combining various things into one coherent, novel whole	synthesis, lab report,
317	Summarize: Using a brief statement or paraphrase to give the main points	summary, synopsis, paraphrase
318	Reflect: Thinking deeply and carefully about something after the fact, often using written text to learn	reflections, reflective writing, writing-to-learn

Table 17 (cont'd)

319	Evaluate: Examining the match between others' writing intent and form using criteria	book review
320	Research: Using systematic investigation to obtain information/knowledge for a piece of writing	experiments
321	Functional: Completing forms, applications , and other fill-in types of documents	checks, resumes
322	Argue: Supporting a position on an abstract proposition	opinion piece, argument, position piece

Appendix C Table 18 State Direct Writing Assessments

State	Assessment Year Range	Grades Assessed	Prompts		Rubrics		
			How many Direct/On-Demand Test responses were there?	What genre(s) were the Direct/On-Demand Test ?	What year was the Direct/On-Demand Test gathered from?	What kinds of Scoring Rubrics are used?	What year were the Scoring Rubrics gathered from?
Alabama	2002-2010	G7	1	Narrative, Descriptive, Expository, Persuasive	2004	Holistic and Analytic	2009
Arizona	2005-2010	G7	1	Informational	2005	Analytic	2003
	2005-2010	G8	1	Narrative	2005	Analytic	2003
Arkansas	2004-2006	G7	2	No set genre	2005, 2006, 2007	Analytic	2006, 2007

Table 18 (cont'd)

	2004-2006	G8	2	No set genre	2005, 2006, 2007	Analytic	2006, 2007
California	2002-2008	G7	1	Randomly chosen (Response to literature, persuasive, summary, narrative)	2002	Holistic	2002
Florida	2001-2009	G8	1	Expository or persuasive	2007	Holistic	2007
Idaho	2003-2008	G7	1	Expository	2006	4 point holistic scale	2006
Illinois	2006 fall- 2010	G8	2	Narrative and persuasive	2010	Analytic. There were two rubrics (one for narrative, and one for persuasive).	2010
Indiana	2001-2009	G7	2	Narrative, response to literature, persuasive	2001-2006	Holistic rubrics for Writing Applications and Language Conventions for grades 3-8. The response to literature also had a Reading Comprehension rubric in addition to the WA and LC	2003, 2005,2006

Table 18 (cont'd)

	2001-2009	G8	2	Narrative, response to literature, persuasive	2001-2006	Holistic rubrics for Writing Applications and Language Conventions for grades 3-8. The response to literature also had a Reading Comprehension rubric in addition to the WA and LC	2003, 2005,2006
Kansas	1998-2007	G8	1	Expository	2004	6 traits analytic	unknown
Kentucky	2006-2009	G8	1	Informative, narrative, persuasive	2006, 2007,2008	Analytic	2006-2009
	2001-2005	G7	1	Persuasive	2004	Holistic	2001-2005
Louisiana	2006 spring-2011	G7	1	Expository or narrative	2007, 2008	Two rubrics during this time period: one measured the dimension of composing and the other measured style/audience awareness. Each dimension was worth 4 points, for a possible total of 8 points.	2006-2011

Table 18 (cont'd)

	1999-2011 Spring (LEAP)	G8	1	Narrative or expository	2003, 2006	Always the same rubrics: one measuring composing; another measuring style/audience awareness and a third measuring the conventions of sentence formation, usage, mechanics, and spelling (each dimension worth one point for a total of 4 points).	2001-2006
Maine	Spring 2001 - Spring 2007	G8	1	Rotates between Narrative and Persuasive	2002, 2004	Analytic	2004
Massachusetts	fall 2001- 2010	G7	1	Personal narrative and expository	2007	Analytic (Development, Conventions)	2007
Michigan	2003 winter-2005 winter	G7	1	Writing from experience and knowledge	2003 winter, 2004 winter, 2005 winter	Holistic six-point rubric	2003 winter, 2004 winter, 2005 winter

Table 18 (cont'd)

	2005 Fall- 2007 Spring	G7 & G8	1	Writing from experience and knowledge	2005 fall, 2006 fall	Holistic six-point rubric	2005 fall, 2006 fall
Missouri	Spring 2006- Spring 2010	G7	1	Exposition	2006	Holistic	2006
Nevada	2001-2007	G8	1	Narrative	2007	Holistic and analytic for voice, organization, ideas, and conventions	2007
New York	Spring 2006 - Spring 2007	G7	2 long, 6 short	Not specified	2006	Holistic	2006
	Spring 2006 - Spring 2007	G8	2 long, 6 short	Not specified	2006	Holistic	2006
North Carolina	2003-2008	G7	1	Argumentative	2006	Holistic 4 Point Rubrics for content and 2 point rubrics for conventions	2006
Oklahoma	2006-2010	G8	1	Vary (narrative, expository, persuasive)	2010	Analytic 5 traits	2010
Rhode Island	2005-2010	G8	3 short, 1 long	No set genre (persuasive, response- to-text, informational)	Fall 2006	Short = 4 pt Holistic Long = 6 pt Holistic	Fall 2006
Tennessee	2002-2007	G8	1	Expository/informative	2004	Holistic 6 Point Rubrics	2002-2007
Texas	2003-2010	G7	1	Unspecified (student's can respond however they like)	2009	Holistic	2009

Table 18 (cont'd)

Vermont	2005-2010	G8	3 short, 1 long	No set genre (persuasive, response to text, informational)	2006	short = 4 pt Holistic long = 6 pt Holistic	2006
Virginia	2006-present	G8	1	Not specified	2011	Analytic - 3 Domains	2006
Washington	Fall 1998-Spring 2007	G7	2	Narrative & Expository	2011	Holistic - 2 Domains	2009, 2010
West Virginia	2005-present	G7	1	Randomly chosen (descriptive, persuasive, informative, narrative)	2005	Analytic	2005
Wisconsin	2003-present	G8	1	Not specified	2007	Holistic	2012

BIBLIOGRAPHY

BIBLIOGRAPHY

- American Educational Research Association/American Psychological Association/National Council of Measurement in Education. (2011). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Ball, A. F. (1999). Evaluating the writing of culturally and linguistically diverse students: The case of the African American vernacular English speaker. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing* (pp.225-248). Urbana, IL: National Council of Teachers of English.
- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63, 69-93.
- Bawarshi, A.S., & Reiff, M.J (2010). *Genre: An introduction to history, theory, Research, and Pedagogy*. Reference Guides to Rhetoric and Composition. Fort Collins, CO: WAC Clearinghouse.
- Beck, S. & Jefery, J. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing*, 12(1), 60-79.
- Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication*. Hillsdale, New Jersey: Erlbaum.
- Brunning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist*, 35, 25-37.
- Carroll, W. M. (1997). Results of third-grade students in a reform curriculum on the Illinois state mathematics test. *Journal for Research in Mathematics Education*, 28(2), 237-242.
- Chen, E., Niemi, D., Wang, J., Wang, H., & Mirocha, J. (2007). *Examining the generalizability of direct writing assessment tasks*. CSE Technical Report 718. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Chesky, J. & Hiebert, E. H. (1987). The effects of prior knowledge and audience on high school students' writing. *Journal of Educational Research*, 80, 304-313.
- Chiste, K. B., & O'Shea, J. (1988). Patterns of question selection and writing performance of ESL students. *TESOL Quarterly*, 22, 681-684.
- Cohen, M. & Riel, M. (1989). The effect of distant audiences on students' writing. *American Educational Research Journal*, 26(2), 143-159.

- Conley, M. W. (2005). *Connecting standards and assessment through literacy*. Boston, MA: Pearson.
- Crowhurst, M. (1988). Research review: Patterns of development in writing persuasive/argumentative discourse. Retrieved from ERIC database. (ED299598)
- Dean, D. (1999). Current-traditional rhetoric: Its past, and what content analysis of texts and tests shows about its present (Doctoral dissertation, Seattle Pacific University).
- Dean, D. (2008). *Genre theory: Teaching, writing, and being*. Urbana: National Council of Teachers of English.
- De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, 94(4), 687-698.
- Devitt, A. (1993). Generalizing about genre: New conceptions of an old concept. *College Composition and Communication*, 44, 573-586.
- Devitt, A. (2009). Teaching critical genre awareness. (Bazerman, C., Bonini, A., & Figueriredo D., Eds.). *Genre in a Changing World*. 337-351. Fort Collins, CO: WAC Clearinghouse and Parlor Press.
- Devitt, A., Reiff, M., & Bawarshi, A. (2004). *Scenes of writing: Strategies for composing with genres*. New York: Pearson/Longman, 2004.
- Donovan, C., & Smolkin, L. (2006). Children's understanding of genre and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 131-143). New York: Guilford.
- Dryer, D. B. (2008). Taking up space: On genre systems as geographies of the possible. *JAC*, 28.3-4:503-534.
- Faigley, L., & Witte, S. P. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, 32(2), 2-11.
- Ferretti, R., Andrews-Weckerly, S., & Lewis, W. (2007). Improving the argumentative writing of students with learning disabilities: Descriptive and normative considerations. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 23(3), 267-285.
- Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28(2), 414-420.
- Flower, L. S., & Hayes, J. R. (1981). Plans that guide the composing process. In C. H. Friderksen & J. F. Dominic (Eds.), *Writing: The nature, development, and teaching of written communication* (pp. 39-58). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Gabrielson, S., Gordon, B., & Englehard, G. (1995). The effects of task choice on the quality of writing obtained in a statewide assessment. *Applied Measurement in Education*, 8(4), 273-290.
- Gearhart, M. & Herman, J.L. (2010). Portfolio assessment: Whose work is it? Issues in the use of classroom assignments for accountability. *Educational Assessment*, 5(1), 41-55.
- Gilliam, R., & Johnston, J. (1992). Spoken and written language relationships in language-learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research*, 35, 1303-1315.
- Glasswell, K., Parr, J., & Aikman, M. (2001). *Development of the asTTle writing assessment rubrics for scoring extended writing tasks* (Technical Report 6). Auckland, New Zealand: Project asTTle, University of Auckland.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.
- Gomez, R., Parker, R., Lara-Alecio, R., & Gomez, L. (1996). Process versus product writing with limited English proficient students. *The Bilingual Research Journal*, 20(2), 209-233.
- Graham, S., Berninger, V.W., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology*, 32, 516-536.
- Graham, S. & Harris, K. (2005). Improving the writing performance of young struggling writers-Theoretical and programmatic research from the center on accelerating student learning. *Journal of Special Education*, 39(1), 19-33.
- Graham, S., McKeown, D., Kiuahara, S. A., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in elementary grades. *Journal of Educational Psychology*, 104(4), 879-896.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.
- Hayes, J. R. (1996). A new model of cognition and affect in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing (pp. 1-27)*. Hillsdale, NJ: Erlbaum.
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Ivanic, R. (2004). Discourses of writing and learning to write. *Language and Education*, 18(3), 220-245.
- Jeffery, J. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, 14, 3-24.

- Jennings, M. Fox, J., Graves, B., & Shohamy, E. (1999). The test takers' choice: An investigation of the effect of topic on language-test performance. *Language Testing*, 16(4), 426-456.
- Jonassen, D. H., Tressmer, M., & Hannum, W. H. (1999). *Task analysis methods for instructional design*. Mahwah, NJ: Lawrence Erlbaum.
- Kanaris, A. (1999). Gendered journeys: Children's writing and the construction of gender. *Language and Education*, 13(4), 254-268.
- Lee, J. Grigg, W.S., & Donahue, P. L. (2007). *The nation's report card: Reading 2007* (No. NCEES 2007496). Washington, DC: US Department of Education.
- Linn, R., Baker, E., & Betebenner, D. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16.
- Lubienski, S. T., & Lubienski, C. (2006). School sector and academic achievement: A multilevel analysis of NAEP Mathematics Data. *American Educational Research Journal*, 43(4), 651-698.
- Moss, P. (1994). Validity in high stakes writing assessment: Problems and possibilities. *Assessing Writing*, 1(1). 109-128.
- National Assessment Governing Board. (2007). *Writing framework and specifications for the 2007 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.
- National Assessment Governing Board. (2010). *Writing framework for the 2011 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.
- National Commission on Writing for America's Families, Schools, and College. (2003, April). *The neglected R: The need for a writing revolution*. New York, NY: College Entrance Examination Board. Retrieved from www.writingcommission.org/pro_downloads/writingcom/neglectedr.pdf
- National Commission on Writing for America's Families, Schools, and College. (2003, April). *Writing: A ticket to work...or a ticket out. A survey of business leaders*. New York, NY: College Entrance Examination Board. Retrieved from www.writingcommission.org/pro_downloads/writingcom/writing-ticket-to-work.pdf
- Newcomer, P. L., & Barehaum, E. M. (1991). The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. *Journal of Learning Disabilities*, 24, 578-593.
- Olinghouse, N., Santangelo, T., & Wilson, J. (2012). Examining the validity of single-occasion,

- single-genre, holistically scored writing assessments. In E. V. Steendam, M. Tillema, G. Rijlaarsdam, & H. V. D. Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices* (pp. 55-82). New York: Guilford.
- Pasquarelli, S. L. (2006). *Teaching writing genres across the curriculum: Strategies for middle school*. Charlotte, NC: IAP-Information Age Publishing, Inc.
- Polio, C. & Glew, M. (1996). ESL writing assessment prompts: How students choose. *Journal of Second Language Writing*, 5(1), 35-49.
- Powers, D. E., & Fowles, M. E. (1998). *Test takers' judgments about GRE writing test prompts*. ETS Research Report 98-36. Princeton, NJ: Educational Testing Service.
- Powers, D. E., Fowles, M. E., Farnum, M., & Gerritz, K. (1992). *Giving a choice of topics on a test of basic writing skills: Does it make any difference?* ETS Research Report No. 92-19. Princeton, NJ: Educational Testing Service.
- Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54-66). New York: Guilford.
- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *Software for three-level analysis. Users' guide for v.2*. London: Institute of Education.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Thousand Oaks, CA: Sage.
- Reiff, M. J. & Bawarshi, A. (2011). Tracing discursive resources: How students use prior genre knowledge to negotiate new writing contexts in first-year composition. *Written Communication*, 28, 3: 312-337.
- Redd-Boyd, T. M. & Slater, W. H. (1989). The effects of audience specification on undergraduates' attitudes, strategies, and writing. *Research in the Teaching of English*, 23(1), 77-108.
- Resta, S., & Eliot, J. (1994). Written expression in boys with attention deficit disorders. *Perceptual and Motor Skills*, 79, 1131-1138.
- Rogers, L., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology*, 100, 879-906.
- Rubin, D. B. (1987). *Multiple imputations for nonresponse in surveys*. New York: John Wiley and Sons.
- Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007*. U. S. Department of Education, Institute of Education Sciences. Washington, DC: National Center for Education Statistics.

- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27, 657-676.
- Stecher, B. M., Barron, S. L., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards based assessment on classroom practices: Results of the 1996-1997 RAND survey of Kentucky teachers of mathematics and writing* (CRESST Tech. Rep. No. 482). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Troia, G. A., & Olinghouse, N. (2010-2014). K-12 Writing Alignment Project. IES funded.
- Troia, G. A., Shankland, R. K., & Wolbers, K. A. (2012). Motivation research in writing: Theoretical and empirical considerations. *Reading and Writing Quarterly*, 28, 5-28.
- US Department of Education. (2004). Charting the course: States decide major provisions under No Child Left Behind. Retrieved from <http://www.ecs.org/html/Document.asp?chouseid=4982>.
- U.S. Department of Education, National Center for Education Statistics. (2010). Teachers' Use of Educational Technology in U.S. Public Schools: 2009. *National Center for Education Statistics*. Retrieved April 2014, from <http://nces.ed.gov/pubs2010/2010040.pdf>
- Zabala, D., Minnici, A., McMurrer, J., & Briggs, L. (2008). *State high school exit exams: Moving toward end-of-course exams*. Washington, DC: Center on Educational Policy.
- Zimmerman, B. J., & Risemberg, R. (1997). Become a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22, 73-101.