

©2011

Robert J. Finnegan

ALL RIGHTS RESERVED

WRITING EXPECTATIONS BEYOND HIGH SCHOOL: A STUDY OF THE  
ALIGNMENT OF THE NEW JERSEY HIGH SCHOOL PROFICIENCY  
ASSESSMENT IN WRITING AND COLLEGE-LEVEL EXPECTATIONS

By

ROBERT J. FINNEGAN

A Dissertation submitted to the  
Graduate School-New Brunswick  
Rutgers, The State University of New Jersey  
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Education

written under the direction of

Erica Boling, PhD

and approved by

---

---

---

---

New Brunswick, New Jersey

January, 2011

## ABSTRACT OF THE DISSERTATION

Writing expectations beyond high school: A study of the alignment of the New Jersey

High School Proficiency Assessment in writing and college-level expectations

by ROBERT J. FINNEGAN

Dissertation Director:

Erica Boling, PhD

Despite recent policy initiatives to ensure high school accountability through state-mandated testing, New Jersey high school graduates may not be prepared for the challenges of college-level writing because the state's high school assessment is not aligned with college-level expectations (Brown & Conley, 2007; Conley, 2003). An ever-growing library of reports acknowledge that many states in addition to New Jersey have not aligned their high school tests in writing with college expectations (Achieve, 2007; Silva, 2008); however, other surveys suggest there is alignment between high school writing curriculum and higher education admission tests (Le, 2002; Milewski, Johnsen, Glazer, & Kubota, 2005). This study, therefore, was designed to describe and understand alignment between writing skills needed for success on the high school assessment in comparison to the writing skills that will be demanded of students in college.

This study used the judgments of a group of eight writing instructors to collect in-depth information on how the New Jersey High School Proficiency Assessment (HSPA) writing scoring rubric aligns with college-level standards for success, as defined in the Knowledge and Skills for University Success (KSUS). In addition, instructors also examined 20 sample HSPA writing responses and provided judgments about whether the

essays demonstrated college readiness. Individual interviews were held with all eight participants, and their judgments were coded to further understand what characteristics of HSPA responses demonstrated college preparedness. From performing these analyses, this study's results corroborate with previous studies on the HSPA (Conley, 2003). The HSPA provides some indication of how students can organize writing in test situations, but the HSPA does not align as well with the type of critical thinking and reading-writing synthesis work that will be required at the college level. These results suggest that policymakers should consider revisions to the HSPA writing test such that it would require students to demonstrate skills in using evidence to sustain cogent arguments.

## Acknowledgement

I would like to thank the members of my committee, William Firestone, Alisa Belzer, and Peter Afflerbach for all the helpful insights during the proposal and dissertation process. I would like to extend my most sincere gratitude and appreciation for my chair and advisor, Erica Boling, who took me on as a student despite not having previous coursework experience together. Without Erica's guidance, this study would have never come to fruition, and I am grateful for her diligent support. Most of all, I would like to acknowledge Jennifer Geoghan, my wife, friend, and partner, who read this manuscript in many forms, and encouraged, nurtured, and motivated my work during this entire process.

## Table of Contents

Abstract	ii
List of Tables	vii
List of Illustrations	viii
CHAPTER 1. INTRODUCTION	1
Statement of the Problem	5
Guiding Questions	8
Nature of the Study	9
Significance of the Study	10
CHAPTER 2. REVIEW OF THE LITERATURE	13
Importance of Alignment	14
Alignment Studies	24
Alignment of College Expectations and Writing Assessments	35
Summary	47
CHAPTER 3. METHODOLOGY	49
Overview of Guiding Questions, Assessments, and Participants	49
Overview of Guiding Questions	50
Context of the Assessments	52
Participants	57
Alignment of Scoring Criteria to KSUS Standards	60
Alignment of Student Responses and College Expectations	70
CHAPTER 4. RESULTS	85
Alignment of Test Success Criteria to College Success Standards	85
Alignment of Sample Responses to College Expectations	102
Differences in Judgments across Educational Contexts	113
Can the HSPA Indicate Preparedness for College-Level Writing?	121
Integrating Alignment of Scoring Criteria and Sample Responses	140
CHAPTER 5. DISCUSSION	146
Summary	146

Implications	154
Considerations for Future Research	156
Validity and Limitations	159
REFERENCES	163
APPENDIX A. PARTICIPANT INSTRUCTIONS	171
APPENDIX B. DEPTH OF KNOWLEDGE LEVELS FOR WRITING	172
APPENDIX C. KSUS OBJECTIVES RATING SHEET	173
APPENDIX D. HSPA SCORING RUBRIC RATING SHEET	176
APPENDIX E. CLEP SCORING RUBRIC RATING SHEET	178
APPENDIX F. HSPA SCORING RUBRIC	179
APPENDIX G. CLEP SCORING RUBRIC	181
APPENDIX H. WRITING PROMPT AND FIVE-PARAGRAPH ESSAY SAMPLE	183
APPENDIX I. ADDITIONAL TABLES AND FIGURES	184
CURRICULUM VITAE	190

## List of Tables

Table 1. Relationship in Study Design to Aspects of Validity	27
Table 2. Overview of Research Questions, Procedures, and Data	54
Table 3. KSUS Objectives Matched to HSPA Rubric Element A.1 by Number of Participants	64
Table 4. Linking Research Questions, Themes, Codes Families, and Sample Quotations	81
Table 5. Summary of Alignment between HSPA and KSUS Using Agreements of 50% or More	88
Table 6. Selected Matches between KSUS Objectives and HSPA Rubric	93
Table 7. Summary of Alignment between CLEP and KSUS with Matches of 50% or More	96
Table 8. Selected Matches between KSUS Objectives and CLEP Rubric	97
Table 9. Comparison of HSPA and CLEP DOK Ratings Across All Participants	101
Table 10. Summary of Characteristics of Essays Related to Judgments of College Readiness	109
Table 10. Summary of Characteristics of Essays Related to Judgments of College Readiness, continued	110
Table 11. Percentage of Essays Judged to Show College Readiness by Teaching Context and HSPA Score	114



## List of Illustrations

Figure 1. Relationship between defining college expectations and code families	80
Figure 2. Comparison of DOK ratings between KSUS and HSPA matches in standard E	92
Figure 3. Number of college ready judgments by HSPA score	104
Figure 4. College-ready judgments for persuasive essays by HSPA score	106
Figure 5. College-ready judgments for speculative essays by HSPA score	106
Figure 6. Quadrant showing how overall features of essays related to college readiness	111

## CHAPTER 1. INTRODUCTION

The fourth season of the critically acclaimed series *The Wire* portrayed a teacher's frustration over state assessment demands as well as changes in district policy regarding state test preparations. Despite being a mathematics instructor, the teacher in the show was forced by his administrators to prepare students for the state writing test by going over a decontextualized writing assignment. In one scene, all teachers were told to focus solely on preparing students to respond to the writing test in a structured and predictable fashion with the hope that the school's test scores would increase. In a following scene, however, the teacher eventually abandoned the idea of teaching to the test, moving his instructional focus back to teaching math with hands-on group projects. In telling his supervisor about not wanting to teach to the test, the teacher questioned the validity of the test and its results if the current educational response encouraged a drill-based approach to learning.

While fictional, this scene encapsulates many of the prominent sentiments in the field of education, and these sentiments have prompted the design of this study. Like the character in *The Wire*, I too stood before a classroom of students in a West Baltimore middle school and was bewildered by the need to navigate troubled home situations, inadequate materials, low student motivation, and increasing assessment demands that didn't seem to account for the contextualized nature of learning for my students. As I moved into a career in educational assessment on a national-level test, I began to see how important assessment design and validity were to maintaining an appropriate balance between learning goals and assessment goals. It became clear to me that assessments

need to align with and reflect instruction, not drive instruction into shallow drills for test preparation.

The current study is an attempt to speak to one aspect of the issues that frustrated me as a teacher and test developer, and which are presented dramatically in *The Wire*: What does a test (its scores and its tasks) mean for the teachers it is intended to help inform and judge? In particular, the current study seeks to understand to what degree the New Jersey High School Proficiency Assessment (HSPA) in writing can inform teachers, parents, and policymakers about student preparedness for college. In the following pages, I will describe the problem, research questions, methods, and results of my analyses. However, it seemed appropriate to first acknowledge that I am interested in examining the role of assessments in educational contexts because I believe the widespread perceived tensions between teaching and testing can have a potentially positive or negative effect on learning. The positive or negative potential is governed by a complex combination of factors. It is my hope that my examination of the HSPA writing assessment will add valuable information on how to better use and understanding tests and in doing so add to the current dialogue on the purpose of assessment in education.

### *Importance of the concept of signaling*

Throughout my time as a graduate student, I have investigated the tensions between testing and teaching. My early pilot study research involved interviewing teachers about their perceptions of tests, and I found that literacy instructors did not believe that the HSPA reflected their pedagogy (Finnegan, 2005). When moving on to plan a dissertation project, I came across the concept of signaling used by Kirst and Braco (2004). Kirst and Braco developed their definition of signaling for a book on policy

initiatives intended to bridge the divide between secondary and postsecondary education. They “view admissions and placement standards and institutional arrangements as policies that communicate signals, meaning, and expected behavior to students and secondary schools” (p. 19). This concept seemed like a useful one for my continued studies, as I was always interested in the signals that tests send to teachers and meaning teachers make of tests.

After considering several study ideas, it became clear that some work should be done to understand how the HSPA fits into the larger educational dialogue on college preparedness. Similar to Kirst and Braco’s (2004) policy studies, I was interested in whether the HSPA provided “clear, consistent, and reinforced signals [that] enhance[d] the college knowledge of prospective students in secondary schools.” (p. 19). This study was designed to understand the potential signal the HSPA sends in terms of its ability to indicate readiness for college. In particular, I was interested in the HSPA writing assessment. It should be noted that the developers of the HSPA writing assessment did not seek to measure college readiness. Instead the explicit purpose of the test was to measure language arts abilities that a student should have acquired in order to graduate from high school.

Despite the HSPA being originally intended to measure whether a student has sufficiently mastered aspects of the high school curriculum, an examination of the signal sent by the assessment about college readiness provides an important link to how proficiency is defined across the educational system. I have served as a writing instructor at a public university in the state, and experienced how first-year students struggled to understand and acclimate their writing to the college standard. Even though high school

students must pass the HSPA to graduate, students I taught struggled with the mandatory expository writing course. In many cases, those same students may have earned high scores on the language arts component of the HSPA, so they had already determined that an outside assessment validated their abilities in writing. Regardless of the HSPA's intended purpose, the secondary consumers of it – parents and students – have created unintended links in its message.

In our knowledge-based economy, a college diploma is becoming a prerequisite for career success (Hoffman, Vargas, Venezia, & Miller, 2007), and rigorous assessment systems have been recommended as a possible way to make educational opportunities more equitable and transparent (Resnick & Resnick, 1992). Therefore, I began to question exactly what specific characteristics of the HSPA signaled about a student's potential across the educational landscape. Was the HSPA signaling to students, parents, and teachers what it means to be proficient in writing, thereby enabling a clear and smooth transition to postsecondary writing?

It is important to note that the concept of signaling that is used throughout this study is meant to speak to the ability of the assessment to provide clear and consistent meaning about student performance. The term signaling is not meant to inform how students are sorted into remedial or credit-bearing courses. Instead, the term is used throughout as a way to frame a decision about how well the high school test mirrors the types of writing expected at the college level. It also should be noted that I did expect some instances in which the HSPA would not align with college expectations, because of the very fact that it wasn't intended to measure college preparedness; however, after performing the analyses described below, the HSPA does in fact measure college

readiness on many levels but fails to measure readiness on every possible level. For the students who presume to be on a college track of high school, their HSPA results may cause confusion because they interpret high school's purpose as preparation for college. I will return to those details later in this dissertation.

### Statement of the Problem

Because the most effective assessment systems are ones that achieve alignment between appropriately defined educational goals and fair measurements of those goals, recent developments in policy and research have sought ways to link high school assessments to college readiness (Callan, Finney, Kirst, Usdan, & Venezia, 2006). Writing ability is of particular importance to college readiness. According to the National Survey of Student Engagement (2008), the importance of writing in postsecondary study is virtually undeniable, and most faculty agree that clear articulation of understanding, creativity, and problem-solving is necessary across the university curriculum.

Despite recent policy initiatives to ensure high school accountability through state-mandated testing, it is unclear whether students in New Jersey are prepared for the challenges of college-level writing because the state's high school assessment is not aligned with college-level expectations (Brown & Conley, 2007; Conley, 2003). An ever-growing library of surveys and reports acknowledge that many states in addition to New Jersey have not aligned their high school tests in writing with college expectations (Achieve, 2007; Silva, 2008); however, other surveys suggest there is alignment between high school writing curriculum and higher education admission tests (Le, 2002; Milewski, Johnsen, Glazer, & Kubota, 2005).

Because of the mixed messages in the field, some work should be done to describe how the HSPA writing assessment aligns with expectations for college-level writers. Several researchers have developed methods for examining how tests align to standards (Porter, 2006). In a series of projects supported by the Council of Chief State School Officers (CCSSO), Norman Webb developed a definition of alignment and methods to analyze alignment of state tests to standards. According to Webb (1997), “alignment is the degree to which expectations and assessment are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expect to know and do” (p. 3). It is important to note that Webb’s conception of alignment seeks to understand the consistency of signals between the test and student learning. Similarly, this study questions whether the HSPA consistently signals the skills that students will need after high school as they proceed in postsecondary education.

Previous studies (Brown & Conley 2007; Conley, 2003) have examined the alignment of the HSPA with the Knowledge and Skills for University Success (KSUS). Conley (2003) found that the New Jersey writing assessment should be examined more closely “to determine if [it has] the potential to provide useful information to students, high schools, or postsecondary institutions regarding the ability of students to succeed in college” (p. 11). While these previous analyses of the HSPA indicate additional research is needed, I considered how these previous studies were performed to see if additional angles of alignment could be pursued. It should be noted that Conley (2003) and Brown and Conley (2007) did not examine two critical aspects of alignment. First, neither study examined how the scoring criteria for the HSPA related to the scoring criteria of a college-level placement exam. For instance, the College Level Examination Program

(CLEP) is accepted by colleges for credit or used for placement and is similar to the HSPA in that it contains a direct writing assessment section. Previous studies have not examined whether the HSPA and a college-level exam (such as CLEP) have similar criteria for success, and whether they address similar or different aspects of the KSUS standards.

The second and possibly more important omission in previous research was that no study has examined student responses as a way to better understand how the HSPA relates to college expectations. In fact, I could not identify other studies that examined how sample responses from any other state writing assessment related to college expectations. An examination of the alignment of student samples to college expectations provides information on the most concrete level. While previous studies have examined the writing prompts and scoring criteria, they failed to examine and explicitly describe the value of the writing produced in the test situation. This study, therefore, was designed to capture explicit information about the HSPA student writing samples and how those responses related to college expectations.

A critical gap exists in the research literature on alignment for performance-based writing assessments because previous studies have used methodologies originally designed for analyzing alignment in multiple-choice based tests in science, mathematics, and reading. Writing assessments, on the other hand, are different from mathematics and science assessments because they rely upon relatively few test items or tasks to serve as evidence of proficiency. In particular, previous studies have not performed an in-depth and comparative analysis of the stated performance requirements found in the scoring rubrics of a high school-level assessment and a college-level assessment. Further,



previous studies have not simultaneously collected judgments from the same expert panel on both the alignment of the scoring criteria and the alignment of evidence of proficiency demonstrated in sample responses.

By replicating aspects of previous studies by Brown and Conley (2007) and Conley (2003), the present study can directly address gaps in the literature by performing an alignment study of the content of the HSPA's scoring criteria against the scoring criteria for a college-level test in writing. At the same time, the present study explores new angles for studying alignment by asking participants to provide a judgment of whether performances on the HSPA are adequate indicators of student readiness to enter credit-bearing college courses in writing.

### Guiding Questions

The following questions and sub-questions provide the structure for this inquiry:

1. How do the New Jersey High School Proficiency Assessment (HSPA) writing scoring criteria align with college expectations?
  - a. How do the HSPA writing scoring criteria align with published consensus objectives of university success (i.e., the KSUS)?
  - b. How do scoring criteria of a college-level writing test align with published consensus objectives of university success?
  - c. What comparisons can be made between the HSPA and the college-level test in terms of alignment to consensus university success objectives?
2. How do benchmark sample responses from the HSPA align with college expectations?

- a. How do scored sample responses from the HSPA align with the judgments of readiness for college provided by writing instructors from high school, 2-year postsecondary, and 4-year postsecondary educational contexts?
- b. What do interviews with instructors from different educational contexts suggest about similarities and/or differences in expectations of college-level writing?
- c. What do interviews with instructors suggest about the ability for the HSPA samples to indicate preparedness for college?
- d. What comparisons can be made between alignment as described through an analysis of college-readiness judgments of sample HSPA essays and alignment as described through a content analysis of scoring criteria?

#### Nature of the Study

Eight instructors with experience in writing assessment were recruited to accomplish two main tasks. In the first task, participants examined the alignment of the scoring criteria used in the HSPA and a college-level writing test to the KSUS standards. The KSUS provide a common language for analyzing alignment of the disparate test scoring criteria. The purpose of including the college-level test is to establish if test criteria at the college level could provide comparative details about whether the HSPA addresses college readiness.

This study uses a modified version of the content alignment methodology established by Webb (1997). The Webb methodology was originally developed to examine large sets of items against state standards; however, this study modified that procedure by focusing on scoring criteria rather than on sets of items. Using the Webb

methodology to examine the scoring criteria provides a common structure to studies extant in the literature and also provides comprehensive data analysis criteria for examining alignment.

The second task involved participants providing judgments about the college-readiness demonstrated in actual student responses from the HSPA writing test. Smith (1992, 1993) established an empirical basis for the efficacy of using college instructors' expert ratings of essays as placement tools. Since the time of Smith's research, many colleges in the United States have used "expert rating" sessions as the focus of their placement protocol (Hout, 2002a). This study used a similar expert rating process to ascertain whether the participants found evidence of college readiness in the sample responses. Interviews were performed to gather in-depth information on how participants arrived at their judgments of college readiness. For the purpose of gathering judgments from participants, a set of 20 sample responses was selected from the *2008 HSPA Criterion-Based Holistic Scoring: A Writing Handbook*. By virtue of the student samples being selected as benchmarks by the state, they serve as exemplars of the type of performances expected at various levels of the scoring criteria. Through this explicit analysis of benchmark sample responses, this study collected data about whether the student performances on the HSPA writing assessment related to college expectations.

### Significance of the Study

This study's methodological developments from previous research fill critical gaps in the current research literature by:

- Comparing HSPA scoring criteria to a college-level writing test's scoring criteria, and

- Including a systematic review of student responses to further examine the evidence of alignment between performances on the HSPA to college expectations.

By modifying the widely-used Webb alignment analysis methods to match with writing assessment concerns, the current study investigates ways in which the scoring criteria from various measures of writing proficiency signal preparedness for college. In addition, a more descriptive dimension is developed by asking instructors to evaluate the relevance of benchmark performances on the state test to college expectations. The current study adds valuable information to the research literature about whether the high school writing assessment can meaningfully signal preparedness for college by combining the Webb (1997) method for judging alignment with an interview-based analysis of student responses.

This study is also important because the HSPA writing assessment is similar to many other state and national assessments, such as the National Assessment of Educational Progress (NAEP). Like NAEP, the HSPA is a direct writing assessment, meaning that the test asks students to write first-draft essays in timed situations as opposed to answering multiple-choice questions about grammatical conventions or writing-related skills. The assessment, therefore, seeks to understand how students write rather than simply how they understand grammatical conventions in the abstract. As Hout (2002a; 2002b) and Yancey (1999) have described, writing assessment involving first-draft essays are preferable in terms of face validity in comparison to multiple-choice assessments of writing. While direct assessments are useful, I wanted to create explicit descriptions of ways in which the HSPA's performance-based writing section succeeds or

fails at signaling college readiness. By doing so, I hoped to provide some information to parents, students and teachers that will lead to potential solutions for easing the transition between writing for high school and writing for college.

## CHAPTER 2. REVIEW OF THE LITERATURE

In this review of the literature, I discuss three areas. First, I discuss the importance of performing alignment studies by presenting studies that show the effects of testing on teacher beliefs, pedagogy, and professional development opportunities. In the second section of this literature review, I describe previous studies that have been done to address alignment between assessments and standards. In the final section this literature review, I discuss alignment studies that involved writing assessments, and in doing so I reveal how teacher judgment provides the necessary groundwork for this study.

Throughout this literature review, my intentions are to establish context, both in regard to theory and research, for the current study. In particular, the intention of the first section of the literature review is to provide a broad perspective of the consequences of testing and how previous studies have understood signals constructed by assessments. Through my analysis of the literature, I have found some indications that when teachers are engaged in assessment on a professional level, the negative impact of testing on teacher beliefs, pedagogy, and professional development is mitigated (if not turned toward a positive outcome).

In the second section of this review, a core definition of alignment is provided, forming the basis for methods of exploring alignment between testing and expectations. In the third section of this review, the goal is to provide a bridge between the first and second sections. That is, I will connect how writing assessments have been studied in particular and provide a rationale for using methods for studying alignment that involve instructors' knowledge and judgments. In total, the three sections of this literature review respond to three distinct questions, providing a conceptual literature-based approach for

the current study. The first section answers why a study of alignment is important considering the context of accountability and assessment-based reform efforts, while the second section answers how alignment has been studied previously and how alignment is defined for this study. Section three exposes how particular issues in the field of writing assessment inform the design of this study.

### Importance of Alignment

An enormous amount of time, effort, and money are spent on creating assessments intended to measure the educational achievement of students in the United States. The study of alignment between testing and college expectations is important because policies, including high school testing policies, send signals about college readiness to students and schools (Kirst & Braco, 2004). If some alignment between testing and ultimate college-level expectations is a goal, then some investigation should be done first that clarifies whether tests align with teacher beliefs, pedagogy, and professional development at the secondary level.

Resnick and Resnick (1992) described the potential of the standards and accountability movement to provide well-conceived performance-based assessments that encourage deeper learning. However, the research evidence is mixed on whether tests (as currently designed and administered) can accurately measure the complex nature of learning (Moss, 2007). This is especially true in language arts and writing assessments where context, audience, motivation, and social languages need to be understood and engaged in the attempt to make meaning (Bakhtin, 1986; Hillocks, 2002a).

Before discussing other alignment studies, it is important to acknowledge a few studies that place emphasis on testing and the potential consequences of testing. In

particular, I will discuss how testing can have an effect on teachers' beliefs, pedagogy, and professional development. From the many possible studies on these topics, I will mention a few here to point out why an examination of alignment is important to a healthy dialogue about assessment. In addition, these studies show why engaging teachers' professional judgment and knowledge are critical aspects of aligning assessments and learning.

*Alignment is Important Because Tests Affect Teacher Beliefs*

A variety of studies have shown that teachers navigate significant theoretical differences when comparing the mode of learning and assessment in classrooms to the mode of assessment as administered on state-mandated tests. For instance, a pair of studies (Assaf, 2005; Hoffman, Assaf & Paris, 2001) performed by researchers in Texas shows the predominant beliefs about state testing. Hoffman et al. (2001) conducted several quantitative analyses of survey responses from teachers in Texas to show that teachers had generally negative attitudes about the Texas Assessment of Academic Skills (TAAS). In particular, teachers felt that TAAS did not accurately measure achievement for minority students, limited English proficient students, or general student learning.

Assaf's (2005) study of a reading specialist in Texas reported that over-time the signals testing sent changed her teaching philosophy.

[She] abandoned her personal and professional philosophies gleaned from years of experience for a testing curriculum. Her reading instruction changed drastically, from rich and authentic discussions about books to a quiet, often subdued atmosphere of silent reading and mastery of low-level test skills isolated from real reading. (p. 164)

Assaf's qualitative case study originally sought to learn more about authentic reading instruction, but after carefully selecting her participant/informant, she found that this



professional had changed her pedagogy to adapt to the new framework of accountability in education. By working more toward the goals of the test, Assaf observed that this reading specialist covered materials without the authentic depth that contemporary movements in professional development of literacy instruction have encouraged (see for example, Harwayne & Calkins, 1991; Heath, 1983).

Similar findings have been explored by Agee (2004) and Brown (2004). Agee found that a teacher with progressive ideals and a commitment to teaching multicultural literature changed to a more skill-based instruction at the high school grades where high-stakes tests would be administered. Brown's regression analysis of survey data found that teachers believed assessments were relevant to their work, but that state-mandated tests should not be used for student accountability. Brown stated that this research implies "that the development of assessment policy should include identification of an appropriate response to teachers' conceptions of assessment" (p. 315). It should be noted that these studies show that tests send important signals to teachers regarding how they formulate their teaching goals, which is also found in Flores and Clarke (2003).

These studies illustrate a small but representative sample of the literature in the field on the negative impact of tests on teacher beliefs. Therefore, it is critical to study the relationship between tests and larger educational expectations (e.g., high school to college expectations). If teachers' professional beliefs and test constructs are at odds, standards-based reforms have a reduced potential to encourage consensus about good instruction and college-readiness. Most importantly, each of these studies found that teachers did not perceive mandated assessments as consistent signals of their values or

educational goals. However, not all studies found that testing created only negative situations in regards to pedagogy, and this is discussed in the following section.

*Alignment is Important Because Testing Affects Pedagogy*

While the previous section shows that teachers' conceive of mandated assessments as running contrary to their own practices in educating students, this section will examine literature on the impact of mandated assessments on pedagogy. Some studies claim that mandated assessments are creating positive changes in practice; other studies find that teachers are preparing students in "symbolic" ways to take the test, or are teaching necessary testing skills in parallel but not connected to their chosen pedagogy.

Firestone and Mayrowetz (2000) contend that external pressure of assessment has a variety of outcomes; most notably, while content may change, high stakes or the perception of stakes may not change the way subjects are taught or that testing may simply create a "symbolic" response that appears like change. More specifically, Firestone et al. (2000) found that the Maryland School Performance Assessment Program (MSPAP) created observable changes by encouraging teachers to construct extended problem solving activities for students in mathematics; however, these activities, as observed by the researchers, were often flawed in that the activities were "often overly coached or provided structured directions that the open-ended nature of the tasks were diminished" (p. 736). Nevertheless, the overall thrust of this paper suggests that teachers' intentions were to prepare students for skills they needed to be successful – not on assessments, but in academic pursuits; in fact, one teacher said "his job was to 'get them ready for algebra. That's my personal opinion. You know, not to get them ready to pass the MEAs' " (p. 728).

Firestone, Monfils, Camilli, Schorr, Hicks, and Mayrowetz (2002) showed that test preparation activities are often associated with inquiry learning and long-term reforms in practice at the elementary level in New Jersey. By using a mixed-method analysis that combined a large-scale survey with classroom observations, this study provided a good example of taking a balanced approach to answering difficult and often controversial questions about testing. Firestone et al. (2002) reported that

Teachers' responses to the ESPA [the state test] are encouraging them to explore approaches to teaching mathematics and science that are more intellectually challenging by making the connection between abstract symbols and more concrete representations and by asking children to explain what they are doing and to justify the procedures used. (p. 1506)

They based this finding largely on the ways that teachers were using rubrics and asking students to prepare for content covered on the test, which was a similarly supported finding in Pomplun (1997).

In general, these findings were more positive than most about the type of reform encouraged by the test; however, this might be due to the mathematics and science focus in Firestone et al. (2002). Research reported earlier focused on literacy, which may lend itself less readily to common practices with rubrics and standardization because literacy is more sociologically situated than other subject areas (Vygotsky, 1978).

Rex and Nelson (2004) presented profiles of two high school English teachers that show how literacy teachers approach test preparation differently than those examined in Firestone et al. (2002). While voicing some displeasure with the idea of preparing students for the state test, the teachers examined in Rex and Nelson did allow for test preparation to become part of their curriculum; however, in doing so they subjugated test preparation to another track on their curriculum. Both teachers maintained authentic and

goal-oriented approaches toward literacy development, including poets' theatres and writing workshops that examined the works of the classics and other contemporary media. When preparing for the state test, both teachers dropped some of these more "authentic" literacy methods and taught five-paragraph essay construction in a more explicit manner.

Stone and Lane (2003) showed that when teachers adapted surface-level features of instruction to conform to the state test (rather than focus on broadening the content they covered to match that required by the test) that students scored lower on average. In other words, teaching to the mere form of the test appears to create additional problems for teachers and students in schools that are trying to address the test without considering how the content of the test can be picked up and embedded or considered as part of their larger pedagogical identities. While Rex and Nelson's (2004) teachers set test preparation off from their regular instruction, their pedagogy was rich and varied to begin with. Stone and Lane's finding, however, could be used as support for aligning assessment signals with the established pedagogical identities and best practices in instruction.

Another study by Johnson, Smagorinsky, Thompson, and Fry (2003) found that a teacher was encouraged to teach writing in a generic five-paragraph theme during her first year of teaching even though she hadn't considered or taught the five-paragraph theme in her student teaching or graduate training. Hillocks (2002a) argued that teaching the five-paragraph theme does not appropriately prepare students for authentic writing situations and represents a sort of thinking that doesn't provide opportunities for critical thinking. Likewise, the researcher/informant in Johnson et al. (2003) didn't value teaching the five-paragraph theme, but felt pressure to adopt this teaching practice.

Again, from this research we can see that mandated assessments have caused a teacher to alter their pedagogical orientation away from factors that are known to be important to college-readiness (such as critical thinking) due to signals from the test that contradict their conceptions of best practices.

In this section, I've presented studies that provide mixed results on how teachers' pedagogy is impacted by mandated large-scale assessment contexts. While teachers feel that assessment can play a role in supporting accountability on the school level or to direct student learning, several studies indicated that changes in pedagogy often occur more frequently on a surface level. Teachers are including different content in their classes to address test content requirements, and teachers are adapting some methods (especially using rubrics) to help students prepare for tests. These studies show that assessments are not exclusively constraining teacher professionalism; instead, teachers often prepared students for the test in addition to maintaining their more complex and critical-thinking oriented curriculum. For this reason, it is important to continue to study how tests align with practices and expectations in the field. In the next section, I will cover studies that show how assessments can interact with teacher professional development.

#### *Alignment is Important Because Tests can Affect Professional Development*

In several of the studies previously mentioned, teachers can be generally said to believe in assessment when it is used to improve student learning (Brown, 2004; Hoffman et al., 2001). In this next section, I will review several studies that examine teachers' beliefs about assessment when engaged in professional development activities that involve active agency in the assessment. In this way, teachers are given a

professional voice and a stake in the assessment. Therefore, supporting the value of teachers' professional judgments might be used as a fulcrum toward making consistent signals between assessment and learning systems. The value of teachers' professional judgments directly relates to the design of this study in that judgments are used as valuable data of alignment criteria, especially when looking at the utility of sample responses as indicators of college readiness.

While providing a review of the literature in the field of writing pedagogy and links to assessment, Higgins, Miller, and Wegmann (2006) provided suggestions for teachers to use "best practices" in teaching writing that simultaneously address requirements of most large-scale writing tests. The best practices that they reviewed from the literature, and recommend for classroom use, have authentic learning in mind—learning that is meant to address real skills for students that align with current, progressive, and constructivist perspectives in the field. For instance, Higgins et al. (2006) called for using writing process instruction that "encompasses the human act of composing and the human gesture of response, thus preparing students to meet writing requirements for any purposes they may encounter throughout their lives" (p. 313). When assessments and test preparation serve the goal of learning, there is a better mesh between teacher beliefs and practices, which, in turn, increases the likelihood of producing consistent signals across learning and assessment.

Gambell and Hunter (2004) investigated whether teachers' involvement as scorers of student responses on a large scale assessment "promoted or eroded the professionalism of teachers in a climate of educational reform" (p. 699). By interviewing teachers that were involved in a scoring project in Canada, Gambell and Hunter found that their

informants “used the national scoring session to broaden their own evaluative frame of reference. Being able to adapt the national scoring experience to classroom use was a tangible benefit for these teachers.” (p. 716). According to this study, all four informants gained corroborating evidence for their own assessment practice or improved their classroom assessment practices.

However, it is important to note that these improvements in their professional development were seen by these teachers as being on a “functional” level (i.e., a practical level of instructing students to become successful on the particular assessment tasks as opposed to more critical or empowering professional development activities). Nevertheless, as with Higgins et al. (2006), we see here that when teachers’ professional identities are engaged (instead of simply attempting to program teachers to rigidly follow prescribed curriculum), the potential for positive outcomes increases.

Dekker and Feijis (2005) found that a professional development program focused on formative assessment practices changed teacher perceptions of assessment and how they used assessments. In particular, teachers worked into their curriculum aspects related to a pyramid model of mathematics instruction—with the bottom of the pyramid representing students’ abilities to replicate specific skills, the middle portion of the pyramid represents students demonstrating the ability to integrate mathematical tools into problem solving, and the top of the pyramid representing the higher level ability of generalizing mathematical thinking and reasoning. By using this pyramid for formative assessment, teachers were able to structure learning based on student performance of classroom-based assessments. Because the program mirrors a large-scale assessment, it implies that these students would be better-suited to performance on a large scale

assessment (however, that was not the purpose of this study). Nevertheless, teachers' beliefs about assessment were quickly altered as part of this study, showing that engaging teachers' in professional development involving assessment may lead to both better assessment performance and lower levels of anxiety about testing.

Borko (1997) reported on a multi-year staff development project that was designed to help teachers incorporate classroom-based performance assessments into their reading and mathematics instructional programs. The research team did not intend on merely giving teachers set practices, but instead sought out ways to help teachers develop assessments that met their instructional goals. Borko's research focused on stimulating teachers as professionals within assessment activities instead of handing down edicts. The study focused on three schools in Colorado at the third grade level and found that by the end of the first year teachers indicated that they knew more about what their students understood and could do in reading and mathematics than they had in previous years. However, as part of the study design, the researchers obtained a waiver from the states mandated test while they participated in the professional development research activities. Therefore, Borko can be used to show that teachers gained confidence and ability in using sound assessment procedures for the purpose of guiding learning; however, this study does not show how the professional development activities linked back to the state's large scale assessment.

The studies presented in this section show that interacting on a professional level with assessment practices changed teachers' perception of assessment positively and may prove to create positive influences for student performance on large-scale assessments. In addition, these studies show that increasing engagement between teacher professionalism



and assessment has the potential to minimize potential anxiety and confrontation between the goals of assessment and learning.

Taken together, the research presented provides a strong indicator that studies on the alignment between testing and teaching are important. Because testing can have an effect on teacher beliefs, pedagogy, and professional development, alignment studies should continue to be performed in order to find out multiple ways in which tests interact with teaching and learning. As Kirst and Usdan (2007) pointed out, there is a long history of changes in the importance of aligning secondary and postsecondary education, and performing studies that make signals of success clearer will provide students, teachers, administrators, and the public with a more comprehensive view of what it means to be prepared for college. In the next section of this literature review, I will discuss ways that alignment between tests, standards, and teaching has been studied previously. In doing so, I will build toward the rationale for my study of alignment.

### Alignment Studies

Messick (1989) claimed that judging validity of a test “requires evaluation of the intended or unintended social consequences of test interpretation and use” (p. 84). This concept of validity looks at consequences of tests in terms of potential effects a test has on teaching and learning (Firestone et al, 2002; Hoffman et al, 2001; Johnson et al 2003; Rex & Nelson, 2003). The preceding section of this literature review examined those sorts of consequential effects of tests in order to support the idea that studying alignment is useful and necessary. However, to further define what is intended by alignment studies, it is important to consider other aspects of Messick’s unified concept of validity. In this section of the literature review, I will discuss studies that have analyzed alignment of

tests to learning goals, but first, as a way to introduce the concept of alignment, I will provide a short description of how alignment studies relate to broader concepts of test validation.

### *Test Validity and a Definition of Alignment*

While there is not enough room in this chapter to fully describe test validation theory, a brief discussion of aspects of test validity will help situate a definition of alignment. A generally accepted model of test validity was developed by Messick (1989) and has been referred to as a “unified” concept of construct validity in which the test is examined for validity in multiple ways. Miller & Linn (2000) provide a succinct overview of Messick’s concept by describing six aspects of construct validity:

1. Content, which examines the relationship between a test’s content and the construct it is intended to measure,
2. Substantive, which examines theories and process models used as a basis for performances on tasks,
3. Structural, which examines how the internal structures of test (such as the scoring and data analysis) relate to the construct of interest,
4. Generalizability, which examines whether results can be replicated across a variety of context purposes,
5. External, which examine the relationship of test result to variables that are external to the test, and
6. Consequential, which (as mentioned above) examine the intended or unintended consequences of using tests.

The concept of alignment, as Miller and Linn (2000) acknowledged, is less than clear in that a “specific meaning of ‘alignment’ and how it should be determined are seldom discussed [in the literature]. The concern here is issues of assessment alignment and the constructs that are invoked by content standards” (p. 368). Therefore, alignment is typically related to a more direct analysis of the content of the assessment and how it relates to the intended purpose of the test. In the context of this study, alignment can be defined by the analysis of the HSPA scoring rubric against consensus university success standards found in the KSUS. In addition, this study uses the aspect of substantive validity to consider how performances (i.e., sample responses) on the assessment align with expected processes that typically indicate readiness for college-level writing.

While Miller and Linn (2000) situated alignment in content validity, I contend that a comprehensive view of alignment may involve any of the aspects of validity mentioned above. In this current study, I look at alignment on the content and substantive levels, and aspects of this study’s design and results also speak to structural, external and consequential aspects of validity. In Table 1, I have marked the relationship between this study and the aspects of validity as either direct or tangential. By direct, I mean that this study’s design directly examines this aspect of the assessment (and therefore its validity). By tangential, I mean that this study will speak to issues involved in this aspect of validity but the design does not directly address these aspects.

Table 1

*Relationship in Study Design to Aspects of Validity*

Aspect of validity	Aspect of current study	Relationship
Content	Analysis of scoring rubric	Direct: Through analyzing scoring rubrics, alignment of content to construct is described.
Substantive	Analysis of sample responses	Direct: Through interview protocol on features of writing sample, this study will describe processes related to successful performances on the writing tasks.
Structural	Analysis of scoring rubric	Tangential: Through analyzing the scoring rubrics, this study investigates internal structures of the writing assessment.
Generalizability	N/A	N/A
External	Analysis of rubric and sample response	Tangential: Through sampling a panel from across educational contexts, and by comparing the HSPA to CLEP, this study speaks to external validity of the assessment.
Consequential	Analysis of sample responses	Tangential: Through an analysis of whether HSPA has the potential to inform college placement, potential consequences are considered.

NOTE: This study does not relate to a definition of generalizability based on the fact that this study is not intended to gather empirical information related to results.

While I argue that it is important to extend the concept of alignment studies to address issues beyond test content, I do maintain that connections to previous research provide a strong frame for the methods used in this study. It is important, therefore, to consider the literature on alignment that has defined the field thus far. Through this discussion of the previous literature, I focus on aspects of alignment studies that speak to both the content and substantive aspects of validity. The literature review below will support the methods I've selected to perform this study, and will inform the final section of the literature on writing assessment alignment issues.

*Webb Alignment Studies*

In a series of projects supported by the Council of Chief State School Officers (CCSSO), Norman Webb developed a definition of alignment and methods to analyze alignment of state tests to standards. According to Webb (1997), “alignment is the degree to which expectations and assessment are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expect to know and do” (p. 3). It is important to note that Webb’s conception of alignment seeks to understand the consistency of signals between the test and student learning. Webb goes on to outline a concept of alignment, and provides five foci for alignment analyses:

1. Content focus,
2. Articulation across grades and ages,
3. Equity and fairness,
4. Pedagogical implications, and
5. Systems applicability.

Webb’s content focus has become the most popular method for studying alignment between assessment and standards. The content focus is likely popular for several reasons.

First, Webb (1997) designed a way to quantify aspects of the content so that alignment conclusions could be quantified across a set of criteria. Traditional content validity work relies on the judgments of experts, and Webb’s criteria for analyzing content alignment allowed those judgments to be articulated across aspects of alignment. As part of this methodology, Webb brings together a panel of experts in a content area (usually teachers and administrators familiar with the test and/or the standards), and they

analyze the expectations or standards-document as well as the test questions for a variety of criteria. In particular, Webb developed six criteria for looking at content alignment:

1. Categorical concurrence, or the concept that the categories of content correspond in the assessment and expectations,
2. Depth of knowledge, or the level of cognitive complexity was the same in the assessment and the expectations,
3. Range of knowledge, or the criteria that the assessment and expectations cover a similar range within a topic or category,
4. Structure of knowledge comparability, wherein the underlying concept of knowledge is in agreement across the expectations and the assessment,
5. Balance of representation, or that a similar emphasis is given to different topics, instruction activities, and tasks, and
6. Dispositional consonance, where the assessment should embody similar attitudes and beliefs about the subject that are articulated in the expectations or standards.

A second likely reason for the popularity of the Webb's (1997) content focus may be found in the easily understood depth of knowledge criteria. The depth of knowledge levels allow for broad complexity definitions to be applied across expectations and assessment tasks. As part of the Webb methodology, a panel of experts provides judgments on the complexity represented in the standards and provides complexity judgments on related assessment tasks. Webb (1999) created depth of knowledge levels so that panelists could provide ratings of complexity on a scale from 1 (lowest complexity) to 4 (highest complexity). The depth of knowledge definitions used in this study can be found in Appendix B.

The depth of knowledge definitions are different for each subject area. The writing depth of knowledge definitions provide a hierarchy of skills—from level 1 demanding demonstration of mechanical skills associated with punctuation and grammar to level 4 requiring higher-level thinking and analysis. The depth of knowledge ratings provide a taxonomy of higher and lower levels skills that will allow comparison across the disparate instruments and standards. Arguably, any alignment judgment is subjective; however, using the depth of knowledge ratings provides a framework for making such judgments and subsequently a way to analyze and describe potential differences between the testing instruments and/or standards. The depth of knowledge levels provide face validity because various stakeholders can identify with a hierarchical ranking of complexity.

The content focus of analyzing alignment of expectations and assessment tasks has become popular for many state policymakers because it provides a comprehensive view of how assessments address various aspects of a set of standards. Webb (1999) and Wixson, Fisk, Dutro and McDaniel (2002) implemented the methodology described above to analyze multiple state assessment systems in math, science, and the language arts. Both studies provided useful information that was generally accepted by the research community. For example, Wixson et al. (2002) found that there was a great deal of variability across states in the degree to which their tests addressed their content standards, and some state tests neglected to include the more challenging and critical material that the standards required. Wixson et al. demonstrated how the Webb criteria can be meaningfully applied to language arts tests, allowing for conclusions to be drawn about how a state's test relates to the complexity demanded in their own standards.

The Webb methodology has been used by state policymakers and researchers to analyze over 20 state assessment programs, and Conley (2003) and Brown and Conley (2007) used the Webb methodology to analyze state assessment and college expectations. In addition, the National Assessment Governing Board (2009) recently selected the Webb content alignment methodology to perform a set of studies on whether the mathematics and reading assessments of National Assessment of Educational Progress and the SAT and ACT provided similar information about high school student readiness for college, work, and the military.

Because of the popularity of this method in the literature and across various studies for state assessment alignment, I have selected this method for the current study. By providing a comprehensive view of how expectations relate to an assessment, the Webb method will provide a means for this study to measure how scoring criteria used on the HSPA aligns with college expectations. I will describe the Webb methodology again in further detail in chapter 3. However, in order to provide a complete view of the field, I will detail several other alignment methodologies.

#### *Achieve Alignment Studies*

As with the Webb methodology, the Achieve method of analyzing alignment brings together educators and experts to provide judgments on test questions in relation to the written standards or expectations in the content area of interest (Resnick, Rothman, Slattery & Vranek, 2003). Also similar to the Webb methodology, the Achieve alignment protocol includes the following dimensions:

1. Content centrality, or the match between content of the test question and the related standard,



2. Performance centrality, or the degree to which the type of performance (cognitive demand) in the test item matches the related standard,
3. Challenge, or the criterion that a set of items requires students to understand challenging subject area knowledge or skills, and
4. Balance and range, or the degree to which the test covers the full range in the standards and does so with a similar emphasis.

According to Resnick et al. (2003), who performed a multi-state study of alignment using the Achieve methodology, the process of judging alignment isn't exactly straightforward and requires the use of judgment on the part of the evaluator (or referred to herein as raters) involved in the study. Raters in the Achieve methodology use criteria similar to that used when scoring performance-based assessment tasks such as open-ended questions. The rubric to make alignment decisions allowed for four types of matches: (a) a strong match, (b) a match that was not specific across a standard and test item, (c) a partial match between a test item and a standard, and (d) inconsistent. In addition to analyzing individual test items for content and performance match, the Achieve methodology calls for reviewers to make a judgment of the test as whole in relation to challenge and balance and range. Overall, these categories of analyzing alignment are fairly similar to Webb's alignment method.

Achieve has done alignment analyses for 14 states, but the reports produced by Achieve are confidential and only limited published research on this method could be found. Nevertheless, it is recognized as a well-known and widely-used methodology, but it was not selected for use in this study because it hadn't been used to evaluate expectations and assessments created in disparate context, as Conley (2003) and Conley

and Brown (2007) studies did with the Webb methodology. In addition, the Webb methodology provides additional measures of alignment that are useful for this current study. Since the current study bolsters the Webb alignment method with an analysis of sample responses, it was decided to use the more quantitative focus of the Webb methodology in order to explore alignment from specifically different perspectives.

### *Survey of Enacted Curriculum*

While the Webb (1999) and Achieve methodologies for analyzing alignment are similar, a method for analyzing alignment developed by Blank, Porter and Smithson (2001) provided a different way of analyzing test content against instructional content. This method relies on teacher reports of the areas of standards that are “enacted” in their curriculum, and in this way provides some indication of how the standards, the test, and the curriculum all relate to each other. Blank et al. (2001), working with the support of CCSSO, asked 600 teachers from 20 schools in six states to complete a survey about the content of their eighth-grade mathematics curriculum. By performing a content analysis of the six state assessments in mathematics, Blank et al. developed a quantitative measure whereby instructional and testing content correlated; in doing so, normal correlation statistics were created (0 for no correlation 1.0 for perfect correlation). Their findings indicated fairly low alignment between tested and taught mathematics content (approximately .23); however, the samples were drawn from a convenience sample and results may have differed if the sample were based on probabilities across the state’s population.

While the results are of interest, the most important aspect of Blank et al. (2001) and other studies by Porter and Smithson (2001) is how alignment can be studied by

combining teacher surveys and content analyses. While surveys have obvious weaknesses, including the possibilities of low response rates and inaccurate reporting of instructional practices, Porter (2002) supported the use of surveys in studying alignment, citing that a prior study (Porter, Kirst, Osthoff, Smithson, & Schneider, 1993) had indicated fairly high rates of correlation between teacher reports on what is taught. Porter (2002) did admit, however, that surveys do not perform adequately as indicators of the quality of instruction. Nevertheless, when studying alignment, achieving a valid estimate of what is taught in comparison to what is tested is the primary concern, and this method can potentially provide that information. That being said, this method of studying alignment was not selected for use in this study largely based on practical constraints of time, budget, and sampling.

The study of enacted curriculum requires teachers to complete the survey multiple times during the year and at the end of the year, resulting in a multi-year study. To provide instructors with incentives would be costly with a large sample of teachers responding to the survey, and collecting data and coding data from a large sample would also be costly in time or budget (if additional research assistants were used to speed up the coding process). Lastly, sampling would require a complex, stratified random sample of teachers from across the state to be a valid representation of the population of interest. Therefore, this study best lends itself to the expert model approach found in the Webb (1999) and Achieve alignment studies.

In conclusion, this section of the literature review provided an overview of three of the most important methodologies for studying alignment between standards, assessments, and teaching. While there are certainly more studies than those mentioned

here, this section provides the basis for studying alignment and how the literature is applicable to the current study. These studies, however, examined math, science, and (to a lesser extent) language arts assessments. In the next section of the literature review, I will discuss additional alignment studies that further explain the framework used in the current study. In particular, I will touch on studies related to college expectations and writing assessments.

### Alignment of College Expectations and Writing Assessments

In this final section of the literature review, I will further describe how I used the literature in the field to inform the design of this study. In particular, this section will clarify particular challenges for aligning writing assessment across the high school and college domains, and it will cover how using instructors' judgments is situated in the literature. Recent studies have provided unclear or potential contradictory evidence that large-scale writing assessments align with the expectations of writing proficiency. In this section, I will briefly consider some of the recent research in the field that presents such contradictions. I will touch on these key research studies as a way to frame the need to perform the alignment analyses that have been conducted in this study.

#### *Extending Alignment to Analyzing College Expectations*

Conley (2003) and Brown and Conley (2007) performed content analyses based on the Webb alignment methodology to compare how the Knowledge and Skills for University Success (KSUS) were addressed by various state assessment programs in math, reading, and writing. The KSUS were developed through a consensus process involving university faculty from across the nation, representing 20 research institutions from the Association of American Universities (Conley, 2003). The KSUS provides

specific content knowledge that students need in order to be successful in college entry-level core academic subjects (such as writing). The Webb alignment methodology, as mentioned above, compared a state's grade-level content standards to a state's test. Conley and Brown & Conley used that same methodology but replaced the state high school standards with the KSUS, thereby providing an indication of whether assessments at the state level in high school could provide useful information about readiness for college.

Brown and Conley (2007) determined that half of the KSUS objectives were addressed by state writing assessments, fewer states had assessments that aligned well with critical thinking aspects of the KSUS objectives. Nevertheless, only half of KSUS objectives for writing were found to be addressed by state writing assessments, and this was thought to be a high percentage. Conley (2003) found that the New Jersey writing assessment should be examined more closely "to determine if [it has] the potential to provide useful information to students, high schools, or postsecondary institutions regarding the ability of students to succeed in college" (p. 11). It is clearly important to describe which aspects of the KSUS objectives are not addressed by specific assessment tasks, and in the case of the current study it is important to look at the HSPA for this indication of alignment.

While Conley's findings point out that there are gaps between state high school tests and the criteria for university success, this gap has also been well-established elsewhere in the literature (Hout 2002a; Yancey, 1999). Moving students into college without the need for remediation is a universal goal (as it is reported to be in the media and through policy reports, cf. NSSE, 2008; Strong American School, 2008), and further

analyses of how the state requirements for high school writing proficiency need to be examined in terms of the writing proficiency needed to enter credit-earning courses. Several other studies have looked at alignment between high school and college-level assessments in a variety of ways.

A study that evaluated alignment between curricula and the SAT writing test claimed that the SAT writing test does reflect the curricula of a large national sample of instructors at both the high school and college level (Milewski et al., 2005). The SAT alignment study asked a large sample of teachers to complete a survey, indicating if predetermined instructional factors were included in their curricula (e.g., the frequency with which they covered such topics as “using supporting details and examples” in their writing instruction). After gathering survey data, Milewski et al. aligned the responses of the survey to skills measured by the assessment. Their findings reported that “skills covered in the classroom appear to be closely aligned with the content specifications of the new SAT writing section” (p. 19).

While alignment to the most frequently cited classroom activities was established, Milewski et al. (2005) acknowledged that the SAT writing test could not address some important writing skills because of constraints of a timed testing situation. Skills not covered by the SAT included using prewriting, generating multiple drafts, writing research papers, using peer groups for feedback and revisions, responding to the needs of different audiences, and understanding writing as a process of invention and rethinking. While these survey-type results can provide broad indications of content covered in college and high school, they may fail to fully describe how the test relates to factors of success in college classes. Because this research suggested alignment but failed to

examine student performances on the assessment against instructor expectations, the current study was designed to gather alignment data both in terms of the test instrument (as shown in the scoring rubric) as well as what levels of writing were credited as showing proficiency (as shown in the sample responses on the HSPA).

It should be acknowledged that the Milewski et al., (2005) study was sponsored and published by the College Board, and similar findings that claim alignment between high school curriculum and admission tests were found by a study sponsored by ACT, which essentially shares the market for undergraduate admissions testing with the College Board (ACT, 2007). ACT also performed a survey and found that the results “support the importance of all six major aspects of writing measured” in the test they develop (p. 11).

While both of these studies show alignment, it is important to examine the results critically. I do not intend to claim that these studies are invalid, but it may be possible that claims made in these reports are framed in terms of the lens in which the research was designed—namely, the surveys were intending to examine elements of the test in comparison to broad perceptions of what instructors believed to be important. The first part of this proposed study will look to measure test content relevance to factors related to college success in terms of the placement test, and doing so in terms of nationally recognized standards of university success. By looking at the assessment with a lens or set of criteria developed independently of the tests, a different perspective on alignment may emerge.

Another study by Roth, Crans, Carter, Ariet and Resnick (2001) found that the Florida tenth grade test was highly-correlated with the passing rates of a postsecondary

placement test. The author claims “the finding that tenth grade scores on a standardized test (the GTAT) is the strongest predictor of success in passing the Reading and Writing portions of the CPT holds out promise that such a test can serve to detect college unpreparedness early, alerting students, parents, and counselors that closer monitoring of course selection and grades is needed” (p. 80). This finding informs the current study because it quantifies that some state tests may be useful predictors of college success, but further content relevance analyses should be done in order to more closely describe the nature of the underlying processes that are being assessed before making claims on quantifiable correlations. Correlations can be helpful, but also can also lead toward unfounded truths (Messick, 1989).

Several large-scale assessments and instructional policy programs intended to align high school and college are currently in use or in development. One of the pioneer large-scale projects meant to align K-16 educational systems is the California State University Early Assessment Program, which began in 2005 and brought together 40 teachers from the Long Beach Unified School District and faculty from California State University Long Beach to develop a program to prepare high school students for the demands of college (Knudson, Zitzer-Comfort, Quirk, & Alexander, 2008). The project team developed an assessment regime within an expository reading and writing course to address students who were at risk of failing the CSU placement exam. The program was found to have statistically significant positive results on the passing rate of the CSU placement test; students who experienced the curriculum developed by the project showed higher scores than those in similar schools without the program (Knudson et al., 2008).



While the program for offering high school students the chance to align their skills with college expectation had a positive impact on success in the case of Long Beach, the task of aligning curriculum in high schools with college expectations is not without controversy. Knudson et al. (2008) reported that teachers were concerned about how they might navigate between the instructional demands of the state graduation standards and the CSU college preparation curriculum. For this reason, it is important to look at the graduation requirements as shown in state high school tests and examine whether those requirements match with college-level expectations. These findings from the CSU-based project are similar to those being studied by staff from the University of Oregon and the South Carolina Department of Education.

While survey-based research and empirical studies comparing placement test scores to programs or other tests suggest that performance on the state's large-scale assessment might correlate to curricula or performance expected of university-level students, there is a significant amount of literature that contradicts this general finding. In particular, Hillocks (2002a, 2008) has claimed that large-scale writing assessments fail to address the critical-thinking aspects of writing (a finding that is supported by Conley, 2003 and Brown and Conley, 2007). Berlin (1996) categorized rhetorical modes of writing and claimed that post-modern writing composition curricula should be based around his concept of social-epistemic rhetoric. In engaging on a critical level the construction of social phenomenon, analyzing ways of thinking, and pursuing the act of making meaning as a social construct, the social-epistemic moves beyond other rhetorical structures where students are encouraged to perform discrete tasks within predefined genres such as narrative or expository writing. The narrowly defined rhetoric of most

classroom instruction, called the current-traditional rhetoric (CTR), forms the basis of most writing assessments (Hillocks, 2002a) and therefore, writing assessments as currently conceived have been challenged on the basis of content validity by such theorists.

Hillocks (2008) reported results of a meta-analysis of quasi-experimental writing research studies, and his findings indicate that practices that engage students in writing, reading examples of writing, or working in self-directed small writing groups showed larger gain than groups involved in decontextualized writing activities (such as learning grammatical patterns in isolation) or generic writing formulae (like CTR exercises). Because Messick (1989) identified various threats that may indicate that judgments based on the scores may not be supportable, most importantly “construct under-representation” and “construct-irrelevant variance,” it is important to examine what kinds of writing practices are part of the state and placement assessments to see if the measures appropriately engage students in writing tasks. The construct may not cover the entire domain, so some judgment will always be made about whether the content and underlying theories of the test (aka the construct) actually measure what is thought to be important. Current thinking in writing assessment holds that some aspects of writing assessments may under-represent the domain or provide construct-irrelevant tasks (Agee, 2004; Hillocks, 2002a).

Therefore, Hillocks (2008) work suggested that additional work should be done to see what aspects of the assessment tasks align with established success criteria and best practices in the field of writing. Keeping in mind that the goal is to improve student writing ability, it should be acknowledged that critical thinking is a key aspect to

articulating a position. Interestingly, social aspects of writing that Hillocks claimed to be associated with greater gains in writing proficiency are similar to aspects of writing that Milewski et al (2005) acknowledge to be outside the scope of the SAT (and by extension other similarly designed writing assessments).

While many have claimed portfolio assessments hold the potential of assessing the critical type of writing that Berlin espouses (Elbow, 1996; Murphy & Yancey, 2008), the research provides mixed indications of whether this claim can be supported. Several studies have discovered some benefits in writing assessment from using ePortfolio systems. Acker and Halasek (2008) found that ePortfolios created a useful environment where college and high school faculty can support the transition of high school writers to college-level expectations.

Acker and Halasek's (2008) findings also revealed that while high school and college writing instructors do not respond to starkly different aspects of student writing, they did find that the two types of instructors emphasize different qualities in student writing—high school instructors emphasized thesis and example relationship (a hallmark of CTR), while college instructors emphasized responses that recognized the various purposes, audiences, and genres needed in the various disciplines. However, Callahan (1997) found that students were often forced to produce essays simply for inclusion in the portfolio to satisfy state standards and that such portfolio assessments didn't necessarily engage critical reflection as part of the writing process (Yancey, 1999).

Because of the work done that argues that writing assessments may not address critical factors in pedagogy, further studies of content relevance of the instruments themselves should be done. Also, because studies on using portfolios for writing

assessment have produced a variety of results, additional work should be done to see how samples of writing on the current assessments align with expectations of first-year college success. One hypothesis is that the HSPA will not represent the underlying theories of writing espoused by faculty, but this hypothesis should be tested, especially in light of reports previously cited herein that claim there is alignment between tests and curricula (Milewski et al., 2005) and the predictive ability of state tests on scores for placement tests (Roth et al., 2001).

### *Writing Assessment Alignment and the Importance of Teacher Judgment*

As discussed in the previous section, there are some potential discrepancies in the field of writing between what ought to be assessed and what ought to be taught. Because of these discrepancies in the field, the current study combines a traditional alignment methodology (Webb, 1997) with an examination of student samples that seeks to elicit the judgment of teachers about the usefulness of the test as an indicator of college readiness. In this final section of the literature review, I will discuss literature that supports the use of teacher judgment as a way to examine alignment of the HSPA against college expectations.

In his essay “Toward a New Discourse of Assessment for the College Writing Classroom,” Hout (2002b) created a theme, similar to the one I’ve been working around in this literature review, that writing assessment and teaching have not shared a common discourse. The lack of connection between writing assessment and teaching can be seen in the contradictions of curriculum alignment presented in the previous section as well as the first section of this literature review that cites studies on the potentially negative effects of testing on the teaching of writing. As Hout pointed out, an examination of

writing assessment is important “[s]ince grades and assessment signify what we value in instruction, connecting how and what we value to what we are attempting to teach seem crucial” (p. 166).

Brown and Conley (2007) used the concept of signaling theory to support the content alignment analyses they performed between state tests and college expectations. Indeed, Hout’s (2002b) concept of assessment signifying instructional values and Brown and Conley’s use of signaling theory (from Kirst & Braco, 2004) share the common interest in providing clarity to students, instructors, policymakers, and the public about how assessments relate to expectations. This study is constructed around this same common interest, and looks at the HSPA to understand ways in which the test may be useful in providing information on college readiness.

Because of the lack of a unifying discourse around writing assessment and teaching, this study included an analysis of teacher judgments of sample HSPA responses. Conley, Aspengren and Stout (2006) used teacher judgments of content that was most important, habits of mind they wished to develop, and effective instructional practices to perform an in-depth analysis describing entry-level college courses. Taking an inventory of the skills and dispositions that are required of high school students moving into college is an important first step, and a logical extension of a skills inventory involves ensuring that assessments of student preparedness address those skills and dispositions.

Placement tests are a typical part of assigning first-year college students to the most appropriate course for their learning needs, and a substantial market has developed around administering and scoring placement tests such as the College Board’s

Accuplacer writing test. Placement tests in writing are usually first-draft writing exercises based on a prompt or short reading passage and administered prior to attending their first semester of classes, which is similar to the HSPA in that the assessment exercise is outside the learning context of a classroom and does not allow for reflection and revision. For many colleges, this placement essay is written and scored by testing companies, but many colleges, including Rutgers, The State University of New Jersey, grade their own placement essays with the use of faculty (personal communication, 7/29/2009). Therefore, using instructor judgment of the usefulness of HSPA as a measure of college readiness may provide information that previous alignment studies mentioned above failed to acknowledge.

In fact, using faculty to make placement judgments has been shown to be effective and correlated well with later student success (Smith, 1993). In addition, Smith's (1992; 1993) research shows that the courses that teachers most recently taught have a high correlation to the types of placement judgments they make. While mostly concerned with ensuring fair and reliable placement decisions, Smith's (1992) findings show that teaching context greatly informs how instructors develop expert rating models based on the courses they last taught. Smith pooled together several rating experiments with instructors who most recently taught one out of three levels of writing courses. The course levels A-C represented gradual progress in writing ability. Collectively, instructors of particular writing courses consistently accepted or rejected entrance into their level of course, showing that instructors have a shared and nuanced understanding, which may or may not be completely articulated, about what types of writing indicate readiness of particular writing courses. Smith (1993) performed think-aloud protocols with instructors

while they rated essays for placement, and found that instructors made specific comments about the writer as a student in addition to making comments about the essay. This finding was echoed by Conley et al. (2006) in that the “frame of mind” exhibited by the student or desired of them is equally as important to perceptions of success and readiness as are more hard skills (e.g., grammar, spelling). For these reasons, this study looks to see how instructor judgments of HSPA samples may or may not exhibit frames of mind and hard skills associated with college readiness. Because teaching contexts were important indicators of placement decisions, this study will gather a panel of instructors from across educational contexts (high school, community college, and university settings) to examine HSPA responses and describe any differences and similarities in judgments of the usefulness of the HSPA as a measure of college readiness.

As Hout (2002a) acknowledged, Smith’s research has become a common practice for many universities; however, it has not been exploited for the purpose of aligning high school assessment and college expectations. Therefore, the current study uses the concept of teacher judgment as a key factor of alignment. As cited in the first section of this literature review, there are several ways that assessment can impact teaching and learning, and when teacher’s professional judgments are engaged a positive relationship between teaching and assessment has been shown to occur. By asking writing instructors to provide a judgment about whether HSPA sample essays exhibit qualities of college readiness, a more comprehensive picture of alignment can be described. In the following chapter, I will discuss the methods in more detail, but it should be noted that performing a follow-up interview with members of the panel (the same panel that is performing the alignment study of the scoring rubrics using the Webb methodology) provides additional

information about aligning writing assessment and teaching values. Most importantly, by reviewing the literature on traditional alignment studies and contradictions in the field of writing assessment, it is clear that a mixed method alignment analysis is necessary in order to more completely answer the question that Conley (2003) posited. That is, can the HSPA provide useful information to students, parents, and college faculty about the ability of high school graduates to succeed in college? By combining methods, this study can distinguish how alignment of the instrument by traditional content alignment methodologies relates to alignment decisions based on teacher judgment, with follow-up interviews providing descriptive details that other studies in the literature have not achieved.

### Summary

In this literature review, I supported the need for this alignment study by discussing a sample of studies that discuss the potential consequences of testing. Consequences of testing provide a strong rationale for doing this current work because unexamined consequences will lead to confusion and tension in the field (as evidenced by the research cited). I provided a rationale for studying alignment using the Webb methodology by discussing the most important and widely used alignment methodologies. By reviewing literature about aligning college expectations and assessments, it is clear that additional research can be done to untangle some of the discrepancies in the field. In particular, the current study uses a mixed methodology to address these discrepancies in the field by combining traditional alignment methodologies with an examination of how teachers judge the relevance of sample responses on the HSPA. This study, therefore, provides a more comprehensive analysis



of the alignment between the NJ HSPA and college-level expectations. In the next chapter, I discuss the methods for this study in more detail.

### CHAPTER 3. METHODOLOGY

In this chapter, I cover the methods used for collecting and analyzing data related to the alignment of the High School Proficiency Assessment (HSPA) writing assessment to college expectations as defined by the Knowledge and Skills for University Success (KSUS) and writing instructors. This chapter will cover information in the following order:

1. An overview of guiding questions, assessments analyzed, and participants,
2. Data collection and analysis methods related to the examination of HSPA and College Level Examination Program (CLEP) scoring criteria, and
3. Data collection and analysis methods related to the examination of the HSPA student responses

In particular, I will provide explanations of how data were collected and analyzed so that results could be created that indicate how the HSPA writing assessment signals college readiness.

#### Overview of Guiding Questions, Assessments, and Participants

As discussed in the previous chapter, an analysis of the consequences of assessment reveals connections between assessment validity, alignment, and teacher beliefs and pedagogy, especially when attempting to signal preparedness for college across educational contexts (Kirst & Braco, 2004). This study combines two methods to analyze whether the HSPA sends consistent signals of college readiness.

First, this study examines whether the HSPA test instrument provides a consistent college-ready signal through analyzing the writing scoring criteria. To perform this analysis, I used a modified version of the Webb (1997) alignment methodology which

examines “the degree to which expectations and assessment are in agreement” (p. 3). Following previous research by Conley (2003) and Brown and Conley (2007), I used the Knowledge and Skills for University Success (KSUS) as the set of standards with which to align the HSPA scoring criteria. To provide a comparative element, I also analyzed the alignment of the College Level Examination Program (CLEP) writing scoring criteria to the KSUS. By analyzing both scoring criteria, I am able to compare the ways in which the HSPA and CLEP align to the KSUS, which serves as a common definition of college-level expectations.

Second, this study analyzes how actual test performances signal college readiness by interviewing instructors from across educational contexts about their judgments of a sample of HSPA writing responses. As Patton (1990) noted, “the purpose of interviewing, then, is to allow us to enter into the other person’s perspective. Qualitative interviewing begins with the assumption that the perspective of others is meaningful, knowable, and able to be made explicit” (p. 278). Through interviewing high school, community college, and university writing instructors, I create an explicit description of how HSPA samples relate to college expectations.

### Overview of Guiding Questions

The procedures for this study were selected to examine two aspects of the HSPA in order to answer two broad questions. First, *how do the New Jersey HSPA writing scoring criteria align with college expectations?* This first question is answered by performing a content alignment study considering how the scoring rubric addresses aspects of the KSUS. Within this first question several detailed questions were used to

frame the analysis of the relationship between the HSPA and college expectations. In particular, the following questions were pursued.

- a. How do the HSPA writing scoring criteria align with published consensus objectives of university success (i.e., the KSUS)?
- b. How do scoring criteria of a college-level writing test align with published consensus objectives of university success?
- c. What comparisons can be made between the HSPA and the college-level test in terms of alignment to consensus objectives of university success?

The second broad question that this study examines is the relationship of test performances to college expectations: *How do benchmark sample responses from the HSPA align with college expectations?* This question is answered with qualitative procedures based on interviews with study participants. Within this second question, the following more detailed questions are answered through an analysis of the interview data with writing instructors:

- a. How do scored sample responses from the HSPA align with the judgments of readiness for college provided by writing instructors from high school, 2-year postsecondary, and 4-year postsecondary educational contexts?
- b. What do interviews with instructors from across different educational context suggest about similarities and/or differences in expectations of college-level writing?
- c. What do interviews with instructors suggest about the ability for the HSPA samples to indicate preparedness for college?

- d. What comparisons can be made between alignment as described through an analysis of college-readiness judgments of sample HSPA essays and alignment as described through a content analysis of scoring criteria?

Each of the questions posed above is answered by a procedure used in the study.

The first broad question is answered through an analysis of the HSPA scoring rubric and the CLEP scoring rubric in comparison to the KSUS standards and objectives. As described below in more detail, the Webb (1997) alignment methodology was employed to examine the content alignment of the scoring criteria from both assessments against the college-level standards. By doing so, the data collected can shape an analysis of whether the test instruments provide a common signal of success in terms of how the direct writing assessment measure is scored. The second broad question is answered through analyzing interviews with instructors about their judgments of the college readiness exhibited in sample responses. The procedures used to examine the second broad question take advantage of the expert reader model (Smith, 1993) and qualitative data analyses used as part of interview-based studies (Creswell, 1998; Marshall & Rossman, 1999). Table 2 provides an overview of how the research questions used in this study are answered by aspects of the procedures employed.

#### Context of the Assessments

According to the New Jersey Department of Education website, the HSPA “is used to determine student achievement in reading, writing, and mathematics as specified in the New Jersey Core Curriculum Content Standards.” The 2008 administration of the HSPA involved two types of writing tasks. The first writing task, commonly referred to as the picture prompt, was categorized as a speculative writing task in which a student

looks at a picture and writes a narrative based on it. Students were allowed thirty minutes to complete a draft narrative based on a picture of two people leaning against the back of a car (see the prompt in Appendix H). The second writing task was categorized as a persuasive task in which students respond to a brief textual prompt on an issue or problem. The persuasive prompt in 2008 asked students to consider whether a school board should adopt a student-run honor council to help with school discipline.

Both the speculative and persuasive tasks are scored using a 6-point scale (1 representing the least success and 6 representing the greatest success) using the state-developed criterion-based holistic rubric (NJ DOE, 2008a). The scoring criteria focus “on four features: content/organization, usage, sentence construction, and mechanics. For any given sample, these criteria serve as an indicator of how well the writer communicates an intended message to a given audience” (NJ DOE, 2008a, p. 1). (See Appendix F for the full scoring rubric.) HSPA student responses in writing are scored by two independent readers. The picture prompt reports to students the average of the two independent scores on the guided holistic scale of 1-6, while the persuasive prompt reports the sum of the two raters on a 2-12 scale.

The HSPA may be called a high-stakes test because it is part of the graduation requirement process. Students must score at the Proficient level or higher, and if they fail to do so they must go through an alternative track process which may or may not grant the student a waiver for graduation. In 2008, 70.6% of all students scored at the Proficient level and 12.2% scored at the Advanced Proficient level in language arts (NJ DOE, 2008b). Unfortunately, writing and reading are not scaled separately, so performance results are not available for writing alone. In addition, information could not be collected

from the NJ DOE on the percentage distribution of all students across the HSPA score scale despite communicating with DOE representatives via email.

Table 2

*Overview of Research Questions, Procedures, and Data*

Research Question	Procedures		Data
	Content Alignment	Interview	
1.a. HSPA writing scoring criteria alignment with KSUS	X		Quantitative data based on Webb criteria for alignment
1.b. College-level writing rubric alignment with KSUS	X		Quantitative data based on Webb criteria for alignment
1.c. Comparisons between HSPA and the college-level test in terms of alignment to KSUS	X		Comparison of criteria for alignment
2.a. HSPA samples' alignment with readiness for college		X	Instructor ratings of preparedness on sample essays
2.b. Interviews' suggestions on HSPA samples indicating preparedness for college		X	Interview data on aspects of samples that do or do not meet expectations
2.c. Interviews' suggestions about similarities and/or differences in expectations across educational contexts		X	Interview data on whether expectations in relation to sample essays is different across participants
2.d. Comparison of alignment between based sample essays and as described through analysis of scoring criteria	X	X	Holistic view of data gathered from both procedures

### *Purpose of Using CLEP as a Comparison*

While the focus of this study is to examine how the HSPA relates to college standards – the KSUS – I also examined how a college-level test relates to these same college standards. By looking at a college-level test, I was able to build comparative evidence about the related requirements in the HSPA scoring criteria and those of a college-level assessment of writing. For the purposes of this study, I selected the College Level Examination Program (CLEP) English composition test, which is published by the College Board. According to the *CLEP Official Study Guide: English Composition* (College Board, 2009), “Most people who take CLEP exams do so to show that they have already learned the key material taught in a college course” (p. 1). Students who take a CLEP test can apply to their college for credit. “Many colleges grant six semester hours (or the equivalent) of credit toward satisfying a liberal arts or distribution requirement in English” for the CLEP English composition test (College Board, 2009, p. 2). Therefore, by comparing the alignment of the scoring rubric used in the CLEP and HSPA to the common KSUS standards, I have captured descriptive details to probe how the high school assessment relates – or fails to relate – to college standards.

Using the CLEP exam for the purpose of alignment provides several advantages. First, CLEP is scored using a similar human scoring process used on the HSPA, whereas some off-the-shelf placement tests (e.g., Accuplacer) use an electronic scoring algorithm to score writing samples. Second, the CLEP writing prompts are somewhat similar to those used on the HSPA’s persuasive task. The CLEP and HSPA prompts ask students to consider a broad issue and argue for a point of view. Because the prompts are similar, the analysis of the scoring rubrics can isolate differences between the two tests. In addition,



the CLEP is accepted for college credit at Rutgers and other public universities in the state of New Jersey.

Through an examination of how the CLEP criteria and the HSPA criteria align against the KSUS, the hypothesis that the assessments are measuring evidence of proficiency in similar or different ways can be tested and described. Measuring how the CLEP relates to the KSUS and thereby comparing CLEP to the HSPA provided a systematic way of judging whether different expectations are embodied in the scoring criteria, which in turn revealed key differences between expectations to graduate high school and expectations to enter credit-bearing college courses or even place out of some required college courses. It is also important to note that this analysis takes account of only the tests requirements in the scoring rubric, and does not consider the multiple-choice questions that are part of the assessment. In this case, this analysis shouldn't be seen as an alignment analysis comparing the CLEP test as a whole to the KSUS or HSPA. This analysis considers only the scoring rubric because the HSPA focuses the measurement of writing proficiency on the performance-based writing sample, and the purpose of this study is to examine how that performance-based task relates to a similar performance-based task at the college level.

#### *Document Samples Used for Judging Alignment*

For the purpose of analyzing the alignment of scoring criteria, I used the HSPA and the CLEP writing composition tests' scoring criteria (See Appendices F and G). All materials have been collected from publically available resources. The HSPA scoring rubric was acquired by ordering the *March 2008 Criterion-Based Holistic Scoring: A Writing Handbook* from the New Jersey Department of Education. The CLEP scoring

criteria were purchased and downloaded from the College Board website. The HSPA sample responses were contained within the 2008 holistic scoring handbook from the Department of Education.

The sample student responses used to judge alignment are benchmark examples selected by the New Jersey Department of Education as exemplars of the six score levels according to the scoring criteria. Interview discussions about the exemplar papers focused on specific qualities of the writing shown in the sample responses and, therefore, provide data on how the benchmark papers align with expectations of college-level writing. The sample of 20 essays represented a stratified sample from across the performance distribution as indicated in the six levels of the scoring criteria.

Because this study is interested in preparedness for college, more samples at the higher end of the scale were examined. Participants did not see the state's assigned score prior to providing their own rating. The stratification for this study contained the following numbers of samples at each score level on the HSPA scoring criteria: Level 1: 2 samples; Level 2: 2 samples; Level 3: 4 samples; Level 4: 4 samples; Level 5: 4 samples; Level 6: 4 samples. When reading each sample, the participants had access to the scoring criteria and the writing prompt. The state's assigned score was revealed to participants only after the participant provided their judgment and after the essay's qualities were discussed during the interview session.

### Participants

The eight participants who took part in this study were instructors at the university (4-year postsecondary), community college (2-year postsecondary), and high school levels. The participants from the study present a balance across the educational

spectrum concerned with bridging high school success and college expectations. Because all participants are volunteers, the sample selection was one of convenience; however, there were some calculated recruiting attempts made in order to ensure that participants were reliable sources of information and were experts in their field. All participants had experience teaching writing at the high school, community college, or university level. Also, all participants recruited had experience with holistic scoring for placement essays or with preparing students to take the HSPA.

In recruiting volunteers, I made contact with the director of the National Writing Project (NWP) at Rutgers, and a posting calling for volunteers was made to the group's website. In addition, I contacted faculty members at the community college and university level through calling and emailing writing department contacts as listed on their websites. In total, I wrote to or called six postsecondary institutions. The NWP call for volunteers yielded half of my participants. Messages written to two New Jersey community colleges and two 4-year colleges yielded the other half of my participants.

As part of recruitment, instructors were informed that their identities and the particular school in which they teach would be kept confidential. Participants were told that they would be asked to participate in two meetings (one coding training/exercise and one interview), and that they could withdraw from participation at any point in the study. In order to make participation as easy as possible for volunteers, meetings were offered over the telephone or at a location convenient to the participant. Participants were informed that regardless of meeting locations, audio recordings would be taken as part of the study. Participants were encouraged to participate based on two factors: participation in the study could provide useful data on alignment between high school and

postsecondary writing; and participants were given a nominal (\$50) gratuity for offering their time.

As Webb (1999) suggested, an alignment panel should have at least 4 members who are experts in the content area being analyzed. From the recruitment attempts described above, I received signed consent forms from eight volunteer participants. Three participants taught at the high school level, were NWP members, worked with students on passing the HSPA, and had attended state- or district-sponsored training about the HSPA scoring procedures. One of these high school instructors was also a part-time writing instructor at the college level. Three participants taught at the community college level, and had experience scoring holistic writing essays for placement and college entrance tests. Two participants taught at four-year universities, and also had experience with scoring essays for placement and college entrance tests.

After making contact with and receiving consent forms from participants, I sent a packet of materials to each volunteer and met with them over the phone to describe the tasks they were to perform (see Appendix A for directions sent to participants). Upon participant completion of the coding of the scoring rubrics and rating of the sample responses (the processes for both tasks are described below), the researcher scheduled a meeting with each participant individually for a follow-up interview on their judgments of the sample responses. Each individual interview session took approximately 45 minutes and covered aspects of each sample that led to their judgments about readiness for college-level writing courses.

### Alignment of Scoring Criteria to KSUS Standards

As described above, the data collection methods for this study occurred in two parts because this study analyzes alignment from two perspectives—namely, the alignment of scoring criteria and the alignment of sample responses to college expectations. In the first data collection step, participants provided ratings on scoring criteria used as part of the HSPA and CLEP writing assessments. This step employed a modified version of the Webb methodology described in chapter 2.

#### *Scoring Criteria Alignment Data Collection and Rating Process*

After collecting signed consent forms, I sent a packet of materials to each participant. Each packet contained instructions, KSUS objectives, HSPA and CLEP scoring rubrics, depth of knowledge definitions, HSPA writing prompts, and the set of 20 HSPA sample essays. After participants received their packet, we scheduled a time to talk on the phone so that I could explain the contents of the packet and train participants on the alignment rating process. Based on the Webb (1997; 1999) methodology, the alignment rating process occurred in three steps. I have described each step in the paragraphs below because the steps are sequential. Each participant followed these procedures first for the HSPA scoring criteria alignment analysis and then repeated this process to rate the CLEP scoring criteria.

First, participants provided a depth of knowledge (DOK) rating to each KSUS standard and objective (Standards are broader goals, and objectives are subcategories within a standard; e.g., “Successful students apply basic grammar conventions in an effort to write clearly” is KSUS writing standard A, whereas “use subject-verb agreement and verb tense consistently and correctly” is objective A.2.). In total, there are six KSUS

standards in writing and 30 objectives that each received a DOK rating between 1 and 4. The writing depth of knowledge definitions provide a hierarchy of skills—from level 1 demanding demonstration of mechanical skills associated with punctuation and grammar to level 4 requiring higher-level thinking and analysis. The depth of knowledge ratings provide a taxonomy of higher and lower levels skills that will allow comparison across the disparate instruments and standards. The KSUS standards and DOK definitions used in this study can be found in Appendix C and B, respectively.

The second step involved participants providing a DOK rating to each aspect of the HSPA (and subsequently the CLEP) scoring rubric. In their packet, each participant had the aspects of the HSPA and CLEP rubrics represented in a table on a coding form. See the form in Appendix D and E, respectively. The HSPA rubric was divided into 15 aspects and the CLEP rubric was divided into six aspects. The third and final step of the scoring rubric alignment process involved participants matching elements of the scoring rubric to one or more KSUS standard or objective.

During our phone call, I trained participants on how to code elements of the standards using the DOK definitions. As Webb (1997) has done previously, discussion with participants occurred on how to apply the definitions for the particular purpose of rating the standards. The training consisted of showing participants example standards and providing ratings for those standards, discussing elements of making the judgment based on the depth of knowledge definitions. Participants then individually rated each of the standards and objectives.

For instance, during the call with participants, I mentioned the first KSUS objective: A.1. *Identify and use correctly and consistently parts of speech, including*

*nouns, pronouns, verbs, adverbs, conjunctions, prepositions, adjectives and interjections.*

I then discussed the possible depth of knowledge rating for this standard, indicating that I thought this was at DOK level 1 because it was related to word-level mechanics. We discussed why this DOK rating was appropriate, and talked about applying the DOK levels in general. In particular, I mentioned that the DOK levels are focused on describing types of writing tasks; however, our goal was to use these levels as a guide to indicate a hierarchy of skills demanded in the scoring criteria.

After we gained consensus on the DOK level of the first KSUS objective, I then discussed the possible links to the HSPA rubric. I provided an example of linking KSUS A.1 to HSPA rubric element B.3. *Pronouns usage/agreement*. We then discussed the appropriateness of this link. In particular, our conversation focused on the fact that both the scoring rubric and the objective mentioned correct usage of pronouns. After our training phone conversation, each participant performed the rest of the rating process on their own. Participants were told to contact me if they had questions or problems during the coding process, but no participants needed additional assistance after the training session. After they completed the task and after our interview, participants returned the coding sheets (and the entire packet of materials) to me. The analysis processes discussed in the next section confirm that participants generally understood the rating procedure, but some differences in opinions and DOK ratings were noticed.

### *Analysis of Alignment Ratings*

This study used the Webb (1997; 1999) methodology to analyze how participants' judgments of the depth of knowledge and matches between the HSPA and CLEP scoring rubrics aligned with college expectations as stated in the KSUS. By doing so, a

description has been constructed of how the scoring criteria, which are the stated requirements for success on the test(s), present a consistent signal with stated success standards from the KSUS. In the following section, I will first discuss how the Webb methodology was employed to manage rater agreement. I will then explain how I analyzed participants' judgments with Webb's (1999) alignment criteria to determine how the HSPA or CLEP rubrics did or did not align with the KSUS.

*Examining agreement.* Webb, Herman, & Webb (2007) showed the importance of accounting for agreement across raters when making alignment conclusions. The alignment coding sheets provided by all eight participants were analyzed for agreement. As Webb et al (2007) recommended, I examined matches between the HSPA/CLEP rubrics and the KSUS in two ways. In the first run through, the matches included all ratings regardless of rater agreement. The second run examined matches when at least half of the participants agreed on the match between the HSPA rubric and the KSUS objective.

Because each rater can (and was encouraged to) match any rubric element to multiple KSUS objectives, there was some variation in ratings of matches between an element of the rubric and the objective to which it applied. For some rubric elements (such as those that addressed spelling or mechanics), there was a direct match in the KSUS standards. For those straightforward matches, all eight participants made a match between the rubric and corresponding KSUS objective.



Table 3

*KSUS Objectives Matched to HSPA Rubric Element A.1 by Number of Participants*

KSUS Objective	Number of participants
D3. use a variety of sentence structures appropriately in writing, including compound, complex, compound-complex, parallel, repetitive and analogous sentence structures.	1
D4. present ideas to achieve overall coherence and logical flow in writing and use appropriate techniques such as transitions and repetition to maximize cohesion.	1
D5. use words correctly; use words that mean what the writer intends to say; and use a varied vocabulary.	4
D6. demonstrate development of a controlled yet unique style and voice in writing where appropriate.	1
E1. know the difference between a topic and a thesis.	2
E2. articulate a position through a thesis statement and advance it using evidence, examples and counterarguments that are relevant to the audience or issue at hand.	5
E3. use a variety of methods to develop arguments, including compare-contrast reasoning; logical arguments (inductive/deductive); and alternation between general and specific (e.g., connections between public knowledge and personal observation and experience).	1
E4. write to persuade the reader by anticipating and addressing counterarguments, by using rhetorical devices and by developing an accurate and expressive style of communication that moves beyond mechanics to add flair and elegance to writing.	4
E5. use a variety of strategies to adapt writing for different audiences and purposes, such as including appropriate content and using appropriate language, style, tone and structure.	4
E8. use appropriate strategies to write expository essays that employ supporting evidence; use information from primary and secondary sources; incorporate charts, graphs, tables and illustrations where appropriate; anticipate and address readers' biases and expectations; and explain technical terms and notations.	1

However, not all matches were so straightforward. For example, there was no perfect match between the HSPA rubric aspect A.1 (*communicates intended message to intended audience*) and the KSUS standards. Therefore, participants indicated various

different objectives that linked to this aspect of the scoring criteria. In total, there were 10 different KSUS standards matched to HSPA rubric aspect A.1 by at least one participant. However, four KSUS objectives were linked to this element of the scoring rubric by at least half (four) of the participants. The table below shows the KSUS matches that were made for HSPA rubric element A.1 and the number of participants that indicated the match.

This table shows that participants varied in their indications of matches, especially when there wasn't a direct correlation between an element of the scoring criteria and the KSUS objectives. Some participants indicated five or six matches for each rubric element, while others indicated only one or two matches. It should be noted that this table shows the rubric element with the greatest variation, and that not all rubric elements were judged as matching to so many different objectives. However, I provided this example to show that consensus could be achieved even when there were divergent opinions about matches. Webb et al. (2007) has suggested the importance of relying on only matches made by half of the participants, and this study supports making alignment judgments when such consensus is achieved. In the following chapter, the presentation of results will focus on matches when at least half of the participants agreed to the match between the scoring rubric and the KSUS objective.

All the HSPA and CLEP rubric elements were matched to one or more of the KSUS objectives by four or more participants, and 10 of 15 HSPA elements had agreed matches by at least seven participants. One benefit of using the Webb (1999) methodology is that the focus on alignment can be maintained despite some aberrant matches assigned by participants. It is important to note that the goal of this study (and

others using the Webb methodology) is not to maintain or measure participant reliability. Instead, the goal here is to determine whether there is a reasonable consensus that elements of the HSPA scoring criteria (or CLEP criteria) align to aspects of the KSUS objectives. Regardless of discrepancies in matches between KSUS objectives and HSPA rubric elements, it is clear that HSPA rubric element A.1 does correspond to college expectations as stated in the KSUS standards. In other words, some of the KSUS objectives are being measured by this element of the HSPA rubric. Complete results of this analysis are discussed in the next chapter. First, however, it is important to more fully explain the analysis procedures used to examine how the HSPA and CLEP rubrics align with the KSUS.

#### *Analysis of Alignment with Webb Criteria*

After analyzing agreement, I summarized the alignment decisions using the methods developed by Webb (1997) and later used by Brown and Conley (2007). In particular, the summary of alignment ratings analyzed alignment using four criteria: (a) categorical concurrence, (b) depth of knowledge consistency, (c) range of knowledge, and (d) balance of representation. By using these criteria, specific facets of alignment between the tests and the KSUS can be made explicit. Also, using these criteria provided an indication of alignment that has been accepted by many in the assessment field (Porter, 2006). Below are descriptions of how these criteria for alignment were used for this study.

*Categorical convergence.* The purpose of analyzing categorical convergence is to determine whether the assessment addresses all the standards. If entire standards are addressed by the assessment, then some aspect of the expectations for student success

would not be captured by the assessment. Webb (1997) maintained that an assessment has achieved categorical convergence if six or more test items (or questions) relate to a given standard. However, because this study does not examine a corpus of items, I used a modified version of this criterion developed by Wixson et al. (2002) in which at least one element of the scoring criteria addresses a standard for which it was determined to provide coverage. In chapter 4, I provide descriptive details about whether each KSUS standard achieved categorical convergence.

*Depth of knowledge consistency.* The purpose of examining the depth of knowledge consistency is to see whether the assessment and the standards address the domain (writing) with a similar level of complexity or depth. Webb (1997) defined the depth of knowledge consistency as the percentage of matches between an assessment item and a standards statement wherein the assessment item had a depth-of-knowledge rating equivalent to or higher than the standard's rating. "That is for the state assessment to be aligned with respect to depth of knowledge, at least half of the assessment items should be at or above the cognitive complexity level of the corresponding" objective (Brown & Conley, 2007, p. 143). This study modifies this benchmark in that we are examining aspects of the writing scoring rubric against standards. Therefore, if at least half of the matching elements of the scoring rubric have the same or higher ratings of depth of knowledge to a corresponding standard/objective, the standard and scoring criteria will be deemed consistent in terms of depth of knowledge.

For example, the objective D.2 is *Use paragraph structure in writing as manifested by the ability to construct coherent paragraphs and arrange paragraphs in logical order*. Six of eight participants rated this objective as a 3 on the depth of

knowledge scale, and so a 3 was recorded as the consensus depth of knowledge rating for this objective. The HSPA scoring rubric element A.3. *Logical progression of ideas* was indicated as matching objective D.2 by seven participants and was also rated as a 3 on the depth of knowledge scale. In other words, the depth of knowledge rating was consistent between the standard and the matching element of the HSPA rubric. If over 50% of the scoring criteria matches have the same or higher depth of knowledge ratings than those assigned to the standard, then the scoring criteria will be determined to have a similar depth of knowledge.

*Range of knowledge.* The purpose of examining range of knowledge is to ascertain whether the assessment and the standards have a similar breadth of representation of the domain of interest (e.g., writing proficiency). In the case of this study, I gauged the percentage of objectives within a standard that were addressed by an element of the scoring rubric. This study considered the writing standard to be aligned in range of knowledge when at least half of the objectives within a standard are addressed by at least one element of the scoring criteria.

*Balance of representation.* The purpose of examining balance of representation is to see how the emphasis in the standards overall relate to the scoring criteria matches. In comparison to the range of knowledge criterion, balance of representation examines how the proportion of objectives addressing each standard compares to the proportion of elements of the scoring rubric addressing each standard. Webb (1999) created a complex formula for dealing with assessments with large items pools; however, because this study looks at relatively fewer aspects of the scoring rubric in comparison to the standard, a simplified formula was applied. I summed the total number of hits and created a

percentage of the total hits for each standard. I then subtracted the percentage of hits for each standard from the percentage of total objectives in each standard. This simplified formula allows for a comparison between the percentage of hits in a standard and the percentage of objectives within a standard.

For instance, standard E has nine objectives and there are 30 objectives in total across all KSUS standards. Represented as a percentage, standard E accounts for 30% of all the KSUS objectives. After collecting ratings from all participants, there were 244 matches between the scoring rubric and the KSUS objectives. Of course, this is the sum of matches as any single aspect of the rubric matched to various objectives. Within standard E, there were 61 matches, which represented 25% of the total number of matches. The difference between the proportion of objectives in standard E and the proportion of hits addressing standard E is 5%. For the purpose of measuring alignment, I represent this difference in percentages as an index of .95. Webb recommended a benchmark of .70 for balance of representation so that there is a sufficiently high degree of overlap between the test and the standards. For this study, balance of representation also used a benchmark of .70; however, no standard fell below .78.

In total, the alignment ratings and analyses described above provided a way to measure the differences between consensus university success standards (the KSUS) and the criteria for successful writing on the HSPA and CLEP tests. In addition, using these criteria allowed for comparisons to be made across the HSPA and CLEP tests to see if the assessments are using systematically different criteria to measure writing proficiency. However, it is also important to examine the assessment beyond the level of the test instrument. Many validity researchers suggest, and writing assessment experts agree, that

it is important to understand how actual student performances on an assessment relate to expert judgments of student proficiencies needed for participation in credit-bearing college-level courses (Hout, 2002a; Messick, 1989; Smith, 1992, 1993).

### Alignment of Student Responses and College Expectations

The second step of data collection involved collecting judgments from participants on the college readiness indicated by 20 sample HSPA writing responses. This second step involved interviewing each participant for approximately 45 minutes to record their judgments of each sample HSPA essay as well as collect their thoughts on the overall ability of the HSPA to signal college readiness.

While performing a comprehensive analysis of the alignment between the HSPA rubric and college expectations is clearly important, a more naturalistic impression of alignment between the state assessment and college expectations can be acquired through directly gathering instructors' impressions about the usefulness of the HSPA to indicate readiness for college. As shown in the second chapter, the research literature shows that testing can have an effect on teachers' beliefs, pedagogy and professional development. When teachers' professional judgments are engaged in the assessment process, however, teachers perceive assessment in a more positive light. Therefore, this study asked instructors who examined the HSPA scoring rubric to also examine a set of benchmark student responses. The analysis of these responses involved not only compiling instructors' judgments about whether the essay indicates college readiness, but also involved interviews with each participant in order to more fully understand how instructors constructed their understanding of acceptable writing.

Interviewing participants about the benchmark samples provided a critical set of data for this descriptive study. The purpose of this study, as stated above, is to understand if the HSPA provides a consistent signal of success in terms of writing proficiencies needed beyond high school. In-depth interviewing with writing instructors from a variety of teaching contexts provided data to better describe the underlying social processes, beliefs, and structures that build toward a definition of what it means to be ready for college (Marshall & Rossman, 1999). As Smith (1992) has shown, teacher perceptions of essays are highly dependent on their most recent teaching experience; therefore, performing in-depth interviews with teachers across contexts can provide detailed information about how these various instructors formulate their expectations. In addition, performing interviews adds a different dimension to this study that cannot be captured by merely examining the scoring criteria and the writing prompts, as other alignment studies have done previously (Brown & Conley, 2007; Le, 2002).

As Patton (1990) noted, “the purpose of interviewing, then, is to allow us to enter into the other person’s perspective. Qualitative interviewing begins with the assumption that the perspective of others is meaningful, knowable, and able to be made explicit” (p. 278). This study’s design allowed me to create explicit descriptions of expectations based on instructors’ expert readings of sample responses. The interviews cover specific aspects of each essay that indicated or failed to indicate readiness for college. Also, the protocol asked instructors to provide an overall opinion of the potential success of the HSPA to indicate college readiness in writing.

*Process for Judging Benchmark Essays*



To understand the relationship of student assessment performances to college-level expectations, the eight participants recruited for this study were asked to provide a judgment for 20 essays written by eleventh grade students as part of the HSPA in 2008. These samples have been collected from the public records available from the New Jersey Department of Education and have no identifying school or student information associated with them. The responses were selected by the authors of the *March 2008 Criterion-Based Holistic Scoring: A Writing Handbook* as benchmark exemplars of one of the six levels of the scoring criteria.

Participants read each of the 20 samples selected for this study and provided one of the following dichotomous judgments: (a) exhibits evidence of preparedness for college-level writing; or, (b) does not exhibit evidence of preparedness for college-level writing. The expert rating process does not rely upon a set of primary trait descriptions or holistic scoring criteria; however, expert ratings have been shown to correlate as well as or better than other essay scoring methods to success in first-year writing courses (Smith, 1993). After rating the sample essays, the researcher met with each participant to perform a retrospective interview on how each essay revealed evidence of student readiness for college. The interview protocol asked the participants to consider qualities of each essay that do or do not meet expectations. While all judgment is subjective, the protocol was designed to elicit as much descriptive detail as possible about how those judgments are made.

### *Interview Protocol*

After rating all responses, the following questions were used to prompt discussion on each sample essay. These core questions were repeated for each sample essay, but as

interviews went on, some flexibility in asking all questions was exercised when participants' discussions sometimes covered material without prompting. The set of questions below was developed to echo, but not exactly repeat, elements found in the KSUS and concepts of writing common in the literature (Berlin, 1996). For instance, the focus on grammar, mechanics, and organization are apparent in the KSUS, while Berlin's (1996) idea of epistemic rhetoric would encourage a focus on critical thinking and creativity.

1. Let's take a look at the first student response. Would you say that this essay indicates that this student is ready for college-level writing?
  - a. If yes, what do you see in the essay that makes you think so?
  - b. If no, what types of evidence would you expect the essay to show that would allow you to make that judgment?
2. Regarding writing characteristics shown in this sample, how is it similar to or different from essays you would expect of a college-level writer?
  - a. How is the grammar, diction, and mechanics similar or different?
  - b. How are the organizational and/or transitional elements similar or different?
  - c. How are the creative or critical thinking elements similar or different?
  - d. How are the aspects related to creating a writer's voice similar or different?
3. Describe how this essay demonstrates (or fails to demonstrate) an understanding of the genre of academic writing that you would expect of a student entering credit-bearing courses in writing.

4. What other qualities in the essay demonstrate readiness (or lack of readiness) for college-level courses?

Based on this set of questions, I collected a set of descriptive attributes about each sample essay that describe whether it does or does not indicate readiness for college-level writing. At the end of the interview, I asked some follow-up questions that more broadly address the alignment of the HSPA to college expectations (see questions below). This set of questions was drafted to respond to the research literature that shows teachers have a negative impression of large-scale writing assessment and its potential for indicating college readiness (Hillocks, 2002a). The main goal in asking for their overall opinion was to ascertain whether they felt the test provided a useful signal of college readiness and what aspects of the assessment failed to measure their expectations for college-level writing.

1. Describe how the essays on the HSPA relate to the expected skills of college-level writers.
2. In what way is the HSPA similar to and different from what you use (or know about) college-level placement tests?
3. What do you see as a strong aspect of way the HSPA measures writing proficiency?
4. What do you see as a weak aspect of the way the HSPA measures writing proficiency?
5. What would be one change that you would want to see in the high school assessment of writing to help ensure students are prepared for college?

#### *Interview Data Management*

All interviews were audio recorded and transcribed verbatim and all filler words, repetitions, and mispronunciations were reproduced as presented in the audio file.

Interview transcripts were shared with participants as a way to ensure quality through member checking of what transpired during the interview. I used ATLAS.ti 6.0 to store all transcript text and capture codes. Several runs through the transcripts were made to determine the differences and similarities of judgments across raters and subsequently to describe evidence in the essay that led the rater to make his or her judgment. The analysis of interview data, therefore, consisted of a series of groupings of judgments of which essays indicated readiness as well as who made those judgments.

The interview data management was critical to answering questions about how sample HSPA responses may or may not provide a consistent signal of success for students. First, each interview transcript was saved as a separate primary document. By having each transcript as a separate document, disaggregating judgments across educational contexts was easily performed. When looking through codes and themes (as is described below), I simply had to select a primary document to see how a particular instructor viewed a specific essay. In addition, because Atlas.ti allows for queries to be constructed that will search for codes and co-occurrences across primary documents, I was able to see how factors related to college readiness connected to the entire data set.

#### *Coding Interview Data*

I made several runs through the interview transcripts in order to develop a description of how sample essays aligned to expectations of college readiness. First, I used microanalysis to analyze interview data in small chunks (Strauss & Corbin, 1998). Interview transcripts were first grouped by essay. The grouping of chunks by essay was a

natural outcome of the protocol because I asked all eight participants to read and judge the same 20 essays. Using Atlas.ti, I blocked out discussions that related to each essay. I also coded that block for each participant identifying their judgments about whether an essay demonstrated college readiness or not. After breaking down each interview transcript into discussions about each of the 20 essays, I further grouped each of those discussions as addressing elements of the protocol. In particular, I grouped segments of each transcript that were related to participants' impressions of whether the organization, grammar, ideas, and other elements of the writing indicated college readiness. In this way, all aspects of the interview transcript were labeled with a code that relates to a particular research question which was also related to a particular interview question. I then coded these groups so that I could search for patterns within each of these larger groups. I used deductive and inductive codes to understand what participants said when making a judgment about the college-readiness of a particular essay.

*Deductive codes.* When I read the interview transcripts, I did one read through that focused on using deductive coding. In particular, the KSUS provided a useful deductive coding framework. Primarily, the KSUS were a useful deductive framework because they represent a comprehensive view of writing that was developed through a national consensus process. In addition, using the KSUS standards for coding also allows for links to be made across the scoring rubric analysis and the interview-based sample essay analysis. I read through the transcripts looking for evidence of all the KSUS objectives. The following bulleted list provides the codes and related KSUS objectives that were found in the interview transcripts.

- Grammar: relates to KSUS A. Successful students apply basic grammar conventions in an effort to write clearly.
- Punctuation: relates to KSUS B. Successful students know conventions of punctuation and capitalization.
- Evidence of planning: relates to KSUS D.1. Know and use several prewriting strategies, including developing a focus; determining the purpose; planning a sequence of ideas; using structured overviews; and creating outlines.
- Logical flow: relates to KSUS D.2. Use paragraph structure in writing as manifested by the ability to construct coherent paragraphs and arrange paragraphs in logical order.
- Paragraphing: relates to KSUS D.2. Use paragraph structure in writing as manifested by the ability to construct coherent paragraphs and arrange paragraphs in logical order.
- Clear focus: relates to KSUS D.4. Present ideas to achieve overall coherence and logical flow in writing and use appropriate techniques such as transitions and repetition to maximize cohesion.
- Transitions: relates to KSUS D.4. Present ideas to achieve overall coherence and logical flow in writing and use appropriate techniques such as transitions and repetition to maximize cohesion.
- Diction: relates to KSUS D.5, Use words correctly; use words that mean what the writer intends to say; and use a varied vocabulary.
- Personal voice: relates to KSUS D.6. Demonstrate development of a controlled yet unique style and voice in writing where appropriate.

- Importance of evidence: relates to KSUS E.2. Articulate a position through a thesis statement and advance it using evidence, examples and counterarguments that are relevant to the audience or issue at hand.
- Creating an argument: relates to KSUS E.3. Use a variety of methods to develop arguments, including compare-contrast reasoning; logical arguments (inductive/deductive); and alternation between general and specific (e.g., connections between public knowledge and personal observation and experience).
- Understanding of audience: relates to KSUS E.5. Use a variety of strategies to adapt writing for different audiences and purposes, such as including appropriate content and using appropriate language, style, tone and structure.
- Fiction strategies: relates to KSUS E.9. Use strategies to write fictional, autobiographical, and biographical narratives that include a well-developed point of view and literary elements; present events in logical sequence; convey a unifying theme or tone; use concrete and sensory language and pace action.
- Importance of revision: relates to KSUS standard F. Successful students both use and prioritize a variety of strategies to revise and edit written work to achieve maximum improvement in the time available.

*Inductive codes.* While I read the transcripts looking for specific aspects related to the KSUS, I also read through the transcripts and sought patterns in the discussions. This reading for patterns allowed for the development of inductive codes. For instance, a set of

codes about solutions for making writing assessments signal college preparedness emerged from the interview transcripts. These codes were as follows

- Computer portfolio solution
- Synthesizing multiple texts
- Test lacks transparency

In addition, the following codes were found through seeing patterns in participants' talk about their judgments of the sample essays.

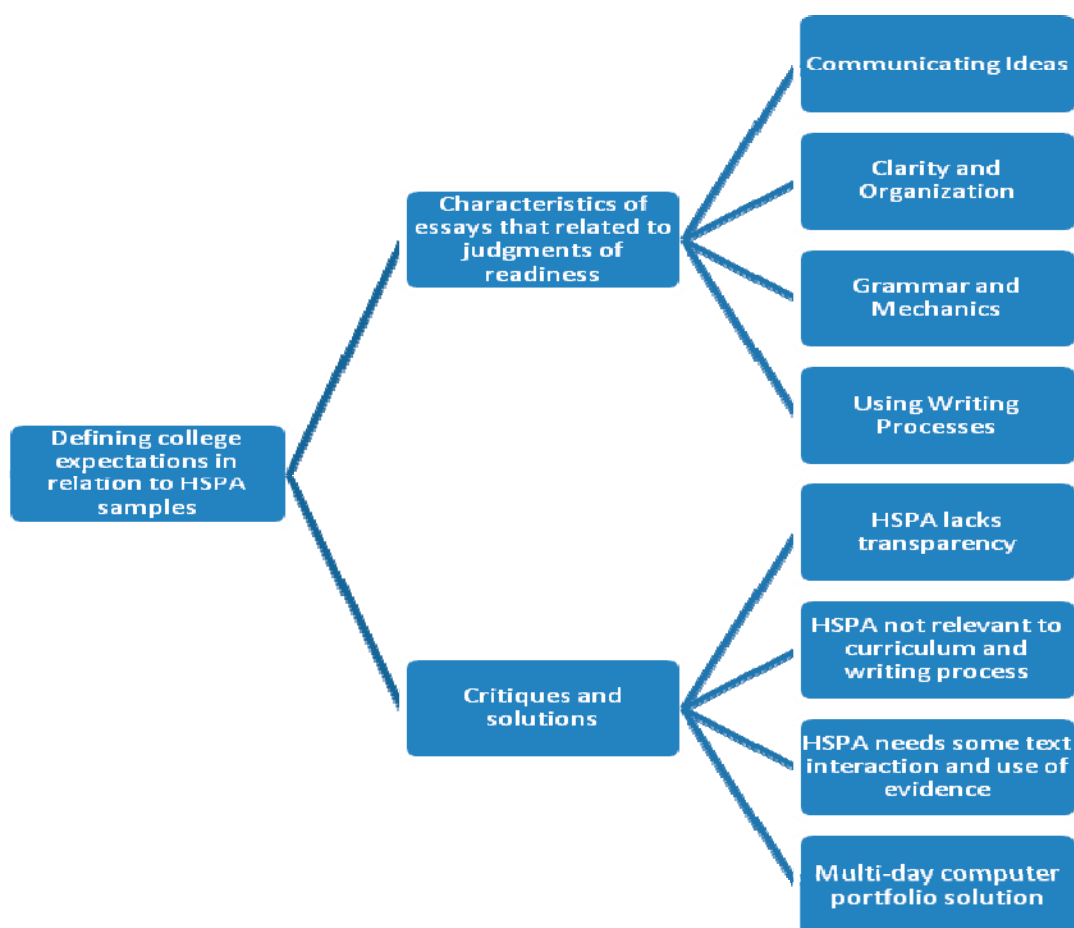
- 5 paragraph essay
- Lack of development
- Ideas more important than errors
- Creative approach
- Lacking creativity
- Critical thinking
- Language issues, ESL
- Predictable structure
- Creativity over clarity

While these codes were not formulated prior to coding the transcripts, they all relate to aspects of the literature. For instance, the computer portfolio solution has been discussed by Acker and Halasek (2008) and synthesizing multiple texts can be related to Berlin's (1986) idea that college students to be critical consumers of multiple types of texts. In addition, issues related to 5-paragraph essays relate to Hillocks (2002a) critique that standardized writing tests encourage essays that lack creativity and critical thinking.

*Connecting Codes, Developing Themes*



After coding the transcript with deductive and inductive codes, I employed qualitative analytic procedures to reduce and group the interview transcript data (Marshall & Rossman, 1999). I went through all of the codes to develop code families. By finding patterns of codes or strands of codes within which certain chunks of text fell, I developed families of codes, using Stake's (1995) concept of categorical aggregation. Beyond building interpretations within categorical families, using categorical aggregation allowed me to make interpretations across the data set. In particular, patterns were sought about qualities in each essay that indicated readiness (or failed to indicate readiness).



*Figure 1.* Relationship between defining college expectations and code families.

Each code family can be related back to a specific research question linked to the analysis of sample HSPA essays. The following table provides an overview of how each research question was related to a theme, code family, and a sample quotation of a code from that code family.

Table 4

*Linking Research Questions, Themes, Codes Families, and Sample Quotations*

Research Question	Theme(s)	Code Families	Sample Quotation
How do scored sample responses from the HSPA align with the judgments of college readiness?	Characteristics of essays	<ul style="list-style-type: none"> <li>Communicating ideas</li> <li>Clarity and organization</li> <li>Grammar and mechanics</li> <li>Writing processes</li> </ul>	<p>If the ideas can be there in a timed writing environment, then that to me suggests college readiness, regardless of structure.</p> <p>Family: Communicating ideas Code: ideas more important than errors (Interview 2, line 107)</p>
What do interviews suggest about the ability for the HSPA samples to indicate preparedness for college?	Critiques and solutions	<ul style="list-style-type: none"> <li>Communicating ideas</li> <li>Clarity and organization</li> <li>HSPA lacks transparency</li> <li>HSPA irrelevant to curriculum</li> </ul>	<p>I mean we tell them all the time don't worry about what you really think. Make up your statistics. Make up your reasons. It just has to have your development. It doesn't have to be true. But, um, to me that doesn't prepare a kid to... how do I then take my own argument and sustain my own argument in college.</p> <p>Family: HSPA irrelevant to curriculum. Code: Test separate from curriculum (interview 8, line 111)</p>
What do interviews with participants from different contexts suggest about similarities and/or differences in expectations?	Characteristics of essays, Critiques and solutions	<ul style="list-style-type: none"> <li>Clarity and organization</li> <li>Communicating ideas</li> <li>HSPA irrelevant to curriculum</li> </ul>	<p>It is written in a pedestrian style ... it doesn't reflect critical thinking... it seems very high school.</p> <p>Family: Clarity and organization Code: Predictable structure (interview 7, line 95)</p>

I performed the categorical aggregation by searching the transcript for co-occurring codes by sample essays, indicating all codes that were mentioned by more than one instructor. By querying combinations of codes in Atlas.ti, I found that quotations related to two of the KSUS standards (standard D: clarity and coherence, and standard E: communicating ideas) were the most frequently cited when participants were making judgments about readiness demonstrated in the essays. Therefore, I created code families to represent these two KSUS standards. The code families related to *communicating ideas* and *clarity and coherence* were associated with over 300 quotations and therefore were noticeably important to instructors' judgments.

I also created code families for two other factors related to essay features: *grammar and mechanics* and *using writing processes*. Grammar and mechanics relates to three of the KSUS standards on grammar, punctuation and capitalization, and spelling, while the writing process family embodies features of writing that indicated the student engaged in elements of the writing process (e.g., planning or revision). Grammar and mechanics and writing processes were not associated with as many quotations as were clarity and coherence or communicating ideas; however, these two code families were nevertheless associated with judgments of essays.

From the 47 codes used to group and examine interview transcripts, I created four code families that provide a structure for examining whether HSPA essays signal college readiness. I also created four additional code families related to critiques and solutions that instructors posed about the alignment of HSPA writing assessment to college

expectations. Figure 1 provides an overview of the code families that emerged from examining the interview transcripts.

### *Synthesizing the Alignment of Scoring Criteria and Sample Responses*

As a way to synthesize the results of the scoring criteria alignment and the sample response alignment, I examined data across both sources and determined if any corroborating or contradictory patterns emerged. For instance, the KSUS standard E is *Successful students use writing to communicate ideas, concepts, emotions and descriptions to the reader*. From performing the scoring guide alignment analysis, I found that this standard was not represented with a consistent depth of knowledge in the HSPA scoring rubric. In other words, the HSPA scoring rubric did address students' abilities to communicate ideas, but did so with a lower level of complexity. This finding is discussed in detail in the results section; however, this particular finding was especially useful when considering how the scoring rubric analysis was related to the sample responses analysis.

In going through the transcripts, I found, as mentioned above, that participants often mentioned the proficiency with which essays appeared to communicate ideas. When the ideas in an essay did not represent critical thinking or lacked support, participants judged those essays as not indicating college readiness. In the final section of the results chapter, I will cover how results from both data collection sources corroborate or provide different evidence of the HSPA's ability to signal college readiness.

Because the KSUS were used as part of the scoring rubric analysis and were used as codes during the analysis of interview transcripts, I was able to create patterns between

the two data sets. In addition, I examined data from both parts of the study with the major themes and code families that I developed as part of the interview transcript analysis.

Looking across data sets helped me to develop a more complete picture of alignment and bolstered the validity of results from these disparate techniques used to gauge alignment. As Conley (2003) suggested, the HSPA should be examined more closely “to determine if [it has] the potential to provide useful information to students, high schools, or postsecondary institutions regarding the ability of students to succeed in college” (p. 11). By performing a comparative alignment study of scoring criteria of HSPA and CLEP, and by examining alignment of HSPA student responses to expert judgments of college readiness, a more comprehensive picture of whether the HSPA can provide useful information emerged.

## CHAPTER 4. RESULTS

In this chapter, I present the results of this mixed-methods study of the alignment between New Jersey High School Proficiency Assessment (HSPA) and college expectations in two main parts. In the first part, the results of the alignment analysis of the HSPA writing assessment scoring criteria to college success standards will be presented. The second main part of this chapter then summarizes the results of the interview-based study of the alignment between HSPA sample responses and college expectations.

### Alignment of Test Success Criteria to College Success Standards

This first section will cover the analyses that were performed involving the alignment of the HSPA scoring criteria to the Knowledge and Skills for University Success (KSUS) standards and objectives. I will then discuss how the College Level Examination Program (CLEP) assessment aligns to the same college-level standards and objectives. Finally, comparisons will be made between the alignment analyses across both tests in order to develop a richer picture of how the HSPA relates to college standards. It is important to keep in mind that the focus of this research is on whether the HSPA has the ability to represent college readiness. The CLEP is being used as a comparative element so that additional details about possible similarities and differences can be made explicit.

#### *How HSPA Scoring Criteria Relate to the KSUS*

By looking at the scoring rubric of the HSPA in relation to the KSUS, a picture emerges about the link between the requirements for success on the test and the consensus skills and knowledge needed for success in college. I used eight individual

ratings from instructors recruited for this study and analyzed the match between the HSPA and the KSUS objectives using the four criteria developed by Webb (1997) and described in chapter 3. While the HSPA was designed to relate to the New Jersey High School standards, the majority of KSUS objectives were also addressed by the HSPA writing assessment; however, not all standards were addressed, mostly due to limitations in the current test design that prevent the measurement of students' revision and editing skills.

Table 5 provides an overview of the how each of the major standards within the KSUS relates to the Webb criteria for alignment. In this table, I provide whether the criteria meet the Webb standard for alignment, and I provide details on what each match means. Note that a match between the rubric and the standard means that some aspect of the HSPA scoring rubric was coded as addressing an objective within the standard. As stated earlier, the Webb (1997) criteria are subject to judgment and do not represent an absolute benchmark. Policymakers and educators should consider these data and determine whether the degree of matches noted in these analyses meet with their expectations for student proficiency. Briefly, the criteria are as follows:

- Categorical convergence: Answers the question of whether the scoring rubric covers the skills and knowledge addressed by the KSUS.
- Depth of knowledge (DOK) consistency: Answers the question of whether the scoring rubric addresses the matching aspects of the KSUS with a similar level of complexity.

- Range of knowledge: Answers the question of whether the scoring rubric addresses a similar range of aspects of writing proficiency as is presented in the KSUS.
- Balance of knowledge: Answers the question of whether the scoring rubric addresses aspects of writing with a similar proportion as is presented in the KSUS.

While most objectives were matched to an aspect of the rubric, it is clear that revision and editing skills were not addressed by the HSPA. Of course, this content alignment analysis would not be necessary to make that determination because it is a well-known constraint of the current design of one-draft performance-based writing assessments. The HSPA is administered in timed periods of 60 minutes or 30 minutes for the persuasive and narrative/speculative prompts, respectively, and the scoring rubric does not ask raters to examine whether any evidence of revision appears in the students essay. Indeed, this constraint of the current model of performance-based writing assessment is acknowledged in the literature and has been frequently cited as one potential weakness of the validity of first-draft-based writing assessments (Milewski et al, 2005; Hillocks, 2002a). I will discuss the issue of revision in relation to the importance of the process as well as instructor judgment in the next section, but at this point it is important to note that the HSPA clearly has no way to measure one category of the KSUS standards.

As stated in chapter 3, I performed the alignment analyses in two ways: one counted all ratings regardless of agreement and another took only matches that were agreed upon by more than half of the participants. Table 4 summarizes results on matches



between the rubric and the standards when at least four (or 50%) of the raters agreed on the match between an element of the HSPA rubric and a KSUS objective. In some cases, participants provided many matches while other cases yielded fewer matches.

Table 5

*Summary of Alignment between HSPA and KSUS Using Agreements of 50% or More*

KSUS Standard	Categorical Convergence	Depth of Knowledge Consistency	Range of Knowledge	Balance of Knowledge
A. grammar conventions	Yes. Matches = 9	Yes. 100% of matches were at or above the DOK level of the objective.	Yes. 100% of the objectives were matched to the rubric.	Yes. Index of .83 for proportion of matches to objectives.
B. punctuation and capitalization	Yes. Matches = 4	Weak. 50% of matches were at or above the DOK level of the objective.	Yes. 100% of the objectives were matched to the rubric.	Yes. Index of .97 for proportion of matches to objectives.
C. conventions of spelling	Yes. Matches = 2	No. 0% of matches were at or above the DOK level of the objective.	Yes. 67% of the objectives were matched to the rubric.	Yes. Index of .97 for proportion of matches to objectives.
D. write clearly and coherently	Yes. Matches = 8	Yes. 62% of matches were at or above the DOK level of the objective.	Yes. 71% of the objectives were matched to the rubric	Yes. Index of .94 for proportion of matches to objectives.
E. communicate ideas, concepts, emotions, and descriptions	Yes. Matches = 8	No. 38% of the matches were at or above the DOK level of the objective.	Yes. 88% of the objectives were matched to the rubric.	Yes. Index of .90 for proportion of matches to objectives.
F. revise and edit	No. Matches = 0	No.	No.	No.

Because there was no focus group session, consensus was more difficult to achieve; however, using the rating of at least half of the participants provides a *de facto* consensus. In fact, there were few deviations in terms of meeting alignment criteria when using only matches with at least 50% agreement. The notable difference, other than the total number of hits being intrinsically lower when only counting those with 50% agreement, was the fact that the depth of knowledge consistency for KSUS standard E (communicating ideas) dropped and therefore it failed to meet the criteria for alignment; when all ratings were accounted for, the depth of knowledge consistency was weak for this standard. The ratings are summarized for all raters, regardless of agreement, in Appendix I.1.

While the HSPA does not address the KSUS standard related to revision, it appears to cover other standards in a fairly comprehensive manner. The HSPA rubric addresses all five additional standards, and in doing so covered a good range and balance of the objectives within a standard. In other words, the HSPA rubric contains specific elements that address each of the standards in the KSUS with the exception of revision. This is why each standard met the categorical convergence criteria. In addition, in all of the standards except revision, the criteria for range of knowledge and balance of knowledge are met; however it should be noted that the range of knowledge in KSUS standards C and D (*spelling* and *using conventions to write clearly*, respectively) were notably lower than in other standards. Nevertheless, the criteria for alignment are met in both cases.

Two standards were not aligned in terms of depth of knowledge, while one other had only a weak relationship in this regard. The HSPA scoring rubric addresses spelling

at a lower level of depth of knowledge because it does not address some particular complexities of spelling the way the KSUS standards do. For instance, the KSUS standards ask that students know the difference between commonly confused terms, show the ability to spell new and unfamiliar words, and use and know the limitations of a word processor's spell checker. The three objectives in the KSUS spelling standard were subsequently rated with a higher DOK level than the comparable element of the HSPA scoring criteria, which more generally mentions spelling as a factor when grading an essay but fails to describe particular elements of spelling in the rubric.

Likewise, the HSPA rubric has only a weak link in terms of depth of knowledge when looking at punctuation and capitalization for a similar lack of specificity. The HSPA rubric generally mentions capitalization as part of the scoring criteria, but does not go into detail, as the KSUS standards do, about issues such as use of ellipses, colons, hyphens, semi-colons, capitalizing proper nouns, and consistently avoiding run-on sentences. The added level of description in the KSUS standards resulted in participants assigning higher DOK ratings than the more general HSPA capitalization and punctuation references.

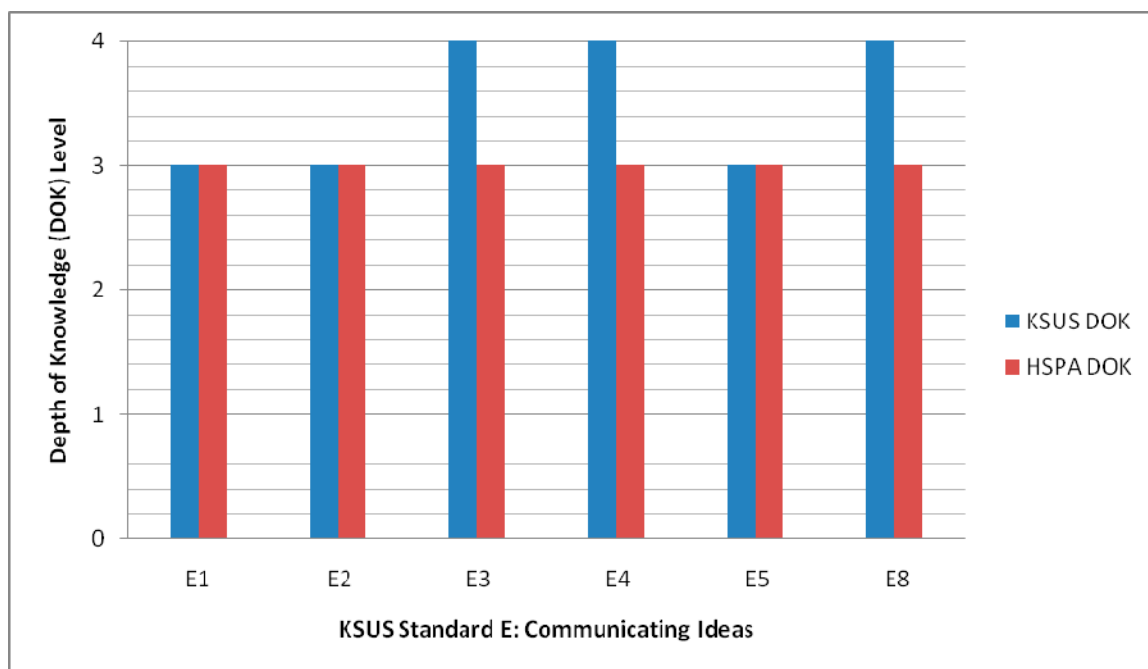
The most important weakness, however, is the one distinguished with regard to the DOK link between the rubric and the KSUS standard E (*successful students use writing to communicate ideas, concepts, emotions, and descriptions to the reader*). The HSPA, by the very nature of it being a performance-based writing assessment, is designed to look more at how a writer can develop ideas within a particular time frame (Hout, 2002a). A performance-based writing assessment may not be as reliable in measuring editing and mechanics skills as a multiple-choice assessment of spelling,

punctuation, and capitalization. The HSPA open-ended writing test focuses more on the nature of the student's abilities in developing ideas within a specific (and perhaps constrained) situation. Therefore, the central focus of a performance-based writing assessment is to understand how a writer behaves in the context of writing an actual piece rather than examining isolated skills such as spelling and punctuation.

Participants' judgments indicated that when connections were made between the HSPA rubric and KSUS objectives rated at the highest DOK level (4) there was an increase in instances of the HSPA representing the skill at a slightly lower DOK level. Below is a figure showing all the DOK ratings and matches for Standard E, where the DOK level of the standards and objectives were predominantly rated 3 or 4. This figure is provided to illustrate that the HSPA rubric is unsuccessful in representing the higher level skills. When a KSUS objective was rated 3, the related HSPA rubric elements could meet the KSUS DOK rating, but when the KSUS objective is rated as a 4 it is frequently matched to HSPA rubric elements with lower DOK ratings. This figure shows only matches for standard E that were agreed upon by at least four raters. As shown below, objectives E6, E7, and E9 were not matched by at least four participants to an element of the HSPA rubric. For the full text of each objective and rubric element, see Appendix C and D, respectively.

While the HSPA rubric fails to represent the higher level DOK ratings in the KSUS standards, it should be noted that the HSPA rubric was never intended to measure college readiness. Therefore, the fact that the rubric is slightly lower on higher level skills such as representing ideas and constructing arguments should be expected. In fact, the HSPA, even though geared to the New Jersey high school standards, does measure many

of the skills described on the KSUS with acceptable levels of depth and range according to the Webb (1999) alignment criteria.



*Figure 2.* Comparison of DOK ratings between KSUS and HSPA matches in standard E.

The HSPA doesn't fully address the development of ideas with the same level of depth of knowledge that the KSUS standards do. In this way, the HSPA fails to signal college readiness. While the assessment is designed to measure student writing with performance tasks, these tasks are not scored with an equivalent level of complexity that are described in the KSUS. The following table provides a few illustrative examples of KSUS objectives within standard E. These examples demonstrate where the HSPA rubric fails to address the KSUS with a similar depth of knowledge. By examining these examples, it is clear that the HSPA rubric does not achieve as high a complexity level because of its lack of specificity. While the KSUS objectives address explicit ways to communicate an argument, the HSPA simply provides a general statement that essays should communicate to an audience or should develop a focus. It should be noted that this

table doesn't include all objectives. These examples are representative of the inconsistencies in terms of depth of knowledge.

Table 6

*Selected Matches between KSUS Objectives and HSPA Rubric*

KSUS Objective in Standard E	HSPA Rubric Element Matched to Standard
E.3 use a variety of methods to develop arguments, including compare-contrast reasoning; logical arguments (inductive/deductive); and alternation between general and specific (e.g., connections between public knowledge and personal observation and experience) DOK = 4	A.2. Relates to topic with a single distinct focus. At the highest level, unified and coherent. Well-developed. DOK = 3
E.4 write to persuade the reader by anticipating and addressing counterarguments, by using rhetorical devices and by developing an accurate and expressive style of communication that moves beyond mechanics to add flair and elegance to writing. DOK = 4	A.1. Communicates intended message to intended audience. DOK = 3
E.8. use appropriate strategies to write expository essays that employ supporting evidence; use information from primary and secondary sources; incorporate charts, graphs, tables and illustrations where appropriate; anticipate and address readers' biases and expectations; and explain technical terms and notations. DOK = 4	A.5 Appropriate details and information: at the highest level, details are effective, vivid, explicit, and/or pertinent. DOK = 3

At least four raters agreed that the HSPA rubric examples and objectives shown above matched, but the rounded average of the DOK levels were lower for the HSPA rubric than for the objectives. Comparing the text in a row, it is clear why raters judged that the depth of knowledge described in the KSUS objectives exceed those in the HSPA scoring rubric. The scoring rubric is clearly looking at a static, single draft sample that

doesn't allow time for research or thoughtful consideration of rhetorical devices. The KSUS objectives in standard E, which is about communicating ideas, suggests a great deal of complexity within the writing process, but the HSPA rubric, perhaps because of the constraints of the testing situation, cannot measure the complexities of the process in the same way.

For instance, KSUS objective E.8 states that college students should be able to use source information to provide supporting evidence for an argument which includes anticipating the audience's bias and other factors. The scoring rubric element A.5 does require that the essay use details that are pertinent, but obviously using primary sources and anticipating the bias of an audience is not possible in the testing situation in the current design. This shows that the testing situation described in the HSPA fails to achieve the highest level of knowledge required by the KSUS. In fact, the absence of the higher level skills on the HSPA is echoed later in the discussion of how teachers described the relationship between the sample responses and college expectations.

#### *How CLEP Scoring Criteria Relate to the KSUS*

In the next section of this chapter, I will address the second research question in this theme: whether the CLEP writing rubric aligns with the KSUS standards. As with the examination of the HSPA rubric, this alignment analysis is undertaken to describe how the requirements for success on the performance-based aspect of the writing test align with consensus standards for success in college. The CLEP rubric was rated by the same study participants who rated the HSPA rubric, and the same alignment criteria were applied in analysis; however, it should be noted that the CLEP rubric describes requirements for success on the test in a more general way and could only be broken

down into six distinct parts. It is also important to note that this analysis takes account of only the tests requirements in the scoring rubric, and does not consider the multiple-choice questions that are part of the CLEP English composition assessment. As mentioned in chapter 3, the scoring rubric is the only aspect considered in this analysis because the goal is to understand how the HSPA writing test relates to a college-level performance-based assessment of writing. Despite there being only six distinct aspects of the rubric, the CLEP aligns with the KSUS standards that are typically the focus of performance-based writing tests (i.e., development, organization, and mechanics). Table 6 provides an overview of alignment to each of the KSUS standards, and includes only those with matches from at least half of the participants.

As shown in Table 7, half of the KSUS standards were not addressed by the CLEP rubric (punctuation and capitalization, spelling, and revision were not addressed). However, as noted previously, the CLEP English Composition test does contain a multiple choice section which covers mechanics and grammar conventions in more detail, though capitalization and spelling are not specifically mentioned in the testing materials for the multiple-choice portion either. Similar to the HSPA, the CLEP does not offer a way for students to revise their written work; however, again, the CLEP test as a whole does have a multiple-choice portion assessing ability to improve faulty sentences. While this might address surface-oriented skills of editing, it does not consider revision in the sense of the KSUS standards. This analysis of the CLEP scoring rubric shows that revision, spelling, and punctuation and capitalization are not measured by the CLEP performance-based writing task.



Table 7

*Summary of Alignment between CLEP and KSUS with Matches of 50% or More*

KSUS Standard	Categorical Convergence	Depth of Knowledge Consistency	Range of Knowledge	Balance of Knowledge
A. grammar conventions	Yes. Matches = 1	Yes. 100% of matches were at or above the DOK level of the objective.	No. 33% of the objectives were matched to the rubric.	Yes. Index of .99 for proportion of matches to objectives.
B. punctuation and capitalization	No. Matches = 0	No.	No.	No.
C. conventions of spelling	No. Matches = 0	No.	No.	No.
D. write clearly and coherently	Yes. Matches = 5	Yes. 100% of matches were at or above the DOK level of the objective.	Yes. 71% of the objectives were matched to the rubric	Yes. Index of .78 for proportion of matches to objectives.
E. communicate ideas, concepts, emotions, and descriptions	Yes. Matches = 5	Yes. 100% of the matches were at or above the DOK level of the objective.	Yes. 55% of the objectives were matched to the rubric.	Yes. Index of .84 for proportion of matches to objectives.
F. revise and edit	No. Matches = 0	No.	No.	No.

When the CLEP rubric addresses a standard it does so with sufficient depth of knowledge, range, and balance, with the exception of the range of coverage in the KSUS standard A which addresses mechanics and grammar. As shown in the table above, the CLEP rubric focuses on KSUS standards D (write clearly) and E (communicate ideas). There were far more hits on both of these standards than any of the others. Also, in terms

of depth of knowledge, the CLEP rubric addressed objectives within both standards D (write clearly) and E (communicate ideas) with a similar depth of knowledge overall.

Table 8

*Selected Matches between KSUS Objectives and CLEP Rubric*

KSUS Objective in Standard E	CLEP Rubric Element Matched to Standard
E.4. write to persuade the reader by anticipating and addressing counterarguments, by using rhetorical devices and by developing an accurate and expressive style of communication that moves beyond mechanics to add flair and elegance to writing. DOK = 4	demonstrates a high degree of competence and sustained control although it may have a few minor errors DOK = 4
E.5. use a variety of strategies to adapt writing for different audiences and purposes, such as including appropriate content and using appropriate language, style, tone and structure. DOK = 3	addresses all elements of the writing task effectively and insightfully DOK = 4
E.8. use appropriate strategies to write expository essays that employ supporting evidence; use information from primary and secondary sources; incorporate charts, graphs, tables and illustrations where appropriate; anticipate and address readers' biases and expectations; and explain technical terms and notations. DOK = 4	develops ideas thoroughly, supporting them with well-chosen reasons, examples, or details DOK = 4

The selected matches above show how the rubric describes the requirements for success at a slightly deeper level of knowledge. Similar to a finding by Le (2002), the inclusion of the requirement that the essay have some degree of insight allowed for raters to judge that the CLEP rubric rose to the 4-level on the DOK scale. Likewise, there were similar mentions in the CLEP rubric about providing thorough and well-chosen supporting details. These types of details, while only briefly alluded to, provided enough

indication for participants to determine that it was at the highest DOK level. Developing scoring requirements that speak to the complexity of the E standard on communicating ideas is difficult, but the CLEP was rated higher in this regard than the HSPA. This shows that CLEP, when it does measure a standard, does so in a way that might provide a clearer signal about college readiness in terms of addressing the complexity communicating ideas.

### *Integrating CLEP and HSPA Alignment*

Looking across the two scoring requirements and their alignment with the KSUS standards reveals some common struggles for writing assessments to address college-level skills and also a few differences between the HSPA and CLEP criteria. The challenge for the scoring rubrics to address the process of revision is largely due to the operational designs of writing assessments as first-draft, prompt-based tasks, and the weaknesses to address some aspects of development and communication of ideas may also reside in the types of tasks students are asked to perform. Hereafter, I will cover the aspects of the KSUS standards that both scoring criteria failed to cover, and then I will cover a few instances in which the HSPA and the CLEP rubric differed in terms of alignment with college-level skills.

The most obvious similarity between the HSPA and CLEP scoring requirements is that they are not intended to judge a student's ability to revise their written work. The assessment does not allow time for the revision process and for that reason the scoring criteria do not address revision. The fact that assessments are not currently designed to address revision does not mean that writing assessments should not or could not be designed to address this important process. Previous studies have shown the importance

of revision to college-level work, and revision is acknowledged by many to be the process that enables writers to achieve higher-level skills (Acker & Halasek, 2007; Yancey, 1999).

In addition to the common failure to address revision, the HSPA and CLEP rubrics were similar in terms of the tendency to avoid focusing on spelling and punctuation/capitalization. The CLEP rubric did not address these skills explicitly, though these skills may be subsumed under the general reference made to grammar and mechanics. The HSPA rubric does mention both skills but in the most general way. On the other hand, the KSUS standards devoted an entire standard to spelling and one objective in another standard to capitalization. While both rubrics addressed spelling minimally or not at all, it may be assumed that essay scorers would likely take into account frequent misspellings or faulty capitalization even though it isn't a major aspect of the scoring criteria. The guided holistic scoring process identifies writers with consistent errors that interfere with the meaning of the piece (Hout, 1994). Again, this tendency of readers will be discussed in more detail later when examining participants' judgments of sample HSPA essays.

The most important difference between the CLEP and HSPA was the depth of knowledge ratings as judged by study participants. Again, the HSPA did not have a strong connection to the KSUS in terms of depth of knowledge consistency, while the CLEP rubric was closer to the KSUS in terms of complexity. This was especially true for the KSUS standard E (communicate ideas). This difference is an important finding because it shows that the HSPA holistic rubric might fail to measure the upper end of the scale appropriately due to its lack of reference for a top level essay to demonstrate some

sort of insight on a topic. This deficiency is significant because the development of ideas has been cited as an important feature and justification for administering performance-based writing assessments (Hout, 2002a; Yancey, 1999).

It should be mentioned that if there were no performance-based writing section at all and if the measurement of writing only occurred with multiple-choice sentence correction items, then the depth of knowledge in terms of communicating ideas would obviously be non-existent or very low. However, even with the advances made in the field with the introduction of performance-based writing assessments, some higher level skills needed for college are not being measured by the current scoring criteria. Namely, the scoring criteria do not measure students' abilities to develop an insightful essay with their arguments supported by evidence.

At this point it is important to note that the Webb alignment methodology, with its four distinct alignment criteria, allowed for me to investigate and identify this weakness in the scoring standards. In fact, the overall depth of knowledge rating for the HSPA rubric, averaged across all raters was 1.8, while the overall average for the CLEP rubric was 3.2. Indeed this difference is somewhat due to the fact that the CLEP rubric does not address the lower level skills of spelling, capitalization, and lower level mechanics in an explicit manner. However, it should be noted that the depth of knowledge ratings for the HSPA items that linked to standards D (write clearly) and E (communicate ideas) were lower than the CLEP ratings for those same standards. See Table 9.

Table 9

*Comparison of HSPA and CLEP DOK Ratings Across All Participants*

KSUS Standard (DOK)	HSPA DOK	CLEP DOK
Overall (2.5)	1.8	3.2
Standard D - write clearly (2.9)	2.5	3.6
Standard E - communicate ideas (3.1)	2.9	3.6

Therefore, while the HSPA and the CLEP do technically meet a variety of criteria regarding the alignment with the KSUS, the CLEP rubric was judged as having a similar depth of knowledge to the standards while the HSPA was lower in this regard. The next section covers how participants judged the HSPA essay samples, echoing the findings about depth of knowledge. By describing these various aspects of alignment of the scoring criteria, specific details about requirements for success on tests have been described in relation to the standards for success in college. With just a few exceptions, namely revision, spelling, and a moderately lower depth of knowledge in the communication of ideas, the HSPA aligns with the KSUS standards in mechanics, clarity, and communication. Similarly, the CLEP aligns with the KSUS standards on clarity, and communication, but does not focus on spelling, capitalization, and mechanics. The CLEP, however, does measure the communication of ideas with a higher level of depth of knowledge than does the HSPA.

From analyzing the scoring rubrics against these specific set of criteria, these similar failings to address elements of standards for college readiness indicate that the common designs for writing assessment should be rethought, and I will return to this idea in the next chapter. In the next chapter, I will address possible avenues for further research and possible design efforts that could be undertaken in order to increase the coverage and depth of current writing assessment designs.

While the analysis of the scoring criteria provides a clear indication of how the tests fail to address college success standards, limiting an analysis of alignment to only these elements of the test instrument would impede understanding and the ability to fully describe what would be needed to change in the current writing assessment paradigm. I designed the second part of this study to probe whether performances on the writing test meet or fail to meet college expectations.

#### Alignment of Sample Responses to College Expectations

While examining alignment through a systematic study of the scoring rubric provides an indication of how the tests success criteria relate to college expectations, examining how instructors' definitions of readiness relate to official scores assigned to HSPA sample responses provides a deeper dimension of how the test signals college preparedness. In the following section of the results, I'll discuss what I learned from performing interviews with eight participants in this study based around their judgments of whether sample HSPA essays indicated college readiness.

This section is divided into four major segments. First, I will display some of the overall data on how the participants judged the 20 sample papers in terms of overall agreement of college readiness. Concurrent with that is a brief examination of the aspects

or characteristics of essays that led participants to make judgments regarding college readiness. Following the discussion of judgments overall, I will move on to discuss the patterns of judgments that emerged based on participants' teaching contexts: high school, 2-year postsecondary, and 4-year postsecondary. The third section will examine how participants described their view of the HSPA in terms of its ability to indicate college preparedness and how the HSPA might be altered in order to align with their curriculum. Embedded in this section is the awareness that all participants felt the test did not adequately cover the full spectrum of skills and abilities that they expect at the college level or within their pedagogy at the high school level, a finding that has been established by other studies (Johnson et al, 2003; Hillocks, 2002a, Yancey, 1999). In the final section of this chapter, I will describe comparisons between the alignment of the scoring rubric to the KSUS standards and the alignment of sample responses to college expectations as uncovered by the interview-based analyses.

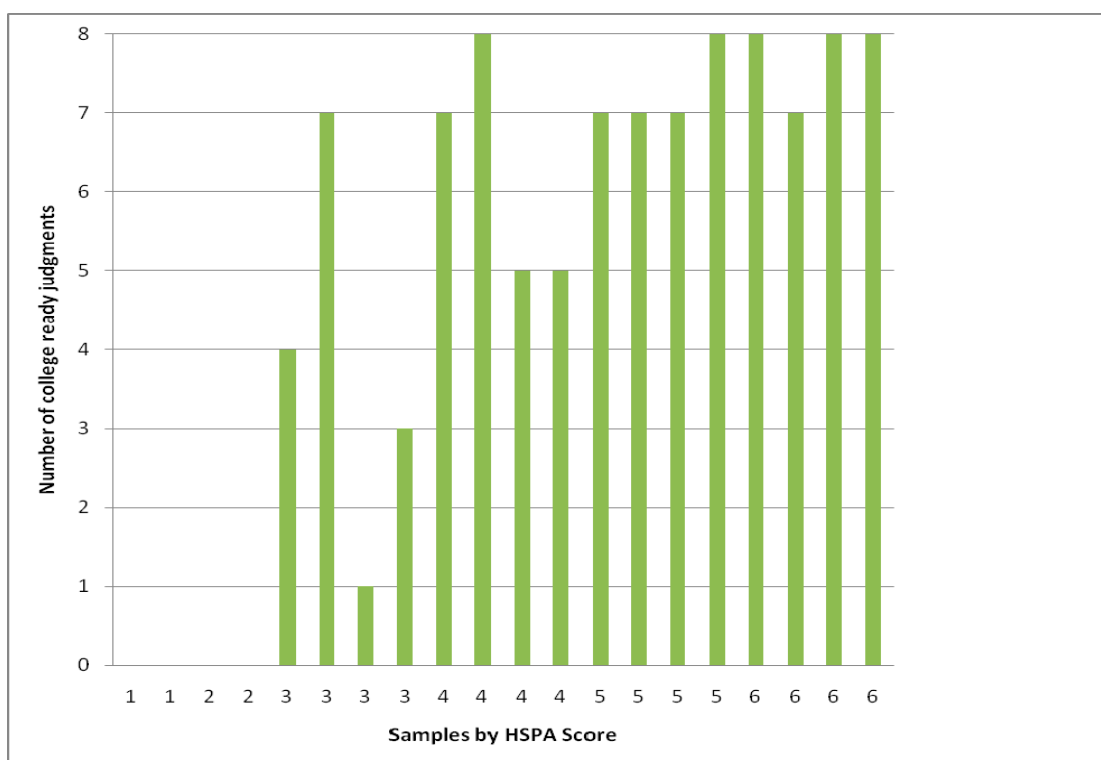
### *Judgments of College Readiness*

Judgments of college readiness were closely aligned to the selection of sample responses. In scoring parlance, the upper half of the scoring range is marked by essays at the 6, 5, and 4 levels, and the set of sample responses selected from the upper half represented 60% of samples used in this study. In other words, there were 20 samples total, and 12 of the samples were benchmark samples from the upper half of the scoring scale. Similarly, approximately 60% of the samples were rated to indicate college readiness by the eight participants. In total, 12 of 20 samples were rated to be ready for college by at least seven of eight participants, or, when examined as a total, 101 of 159 total judgments across all eight raters were thought to indicate college readiness (63.5%).



One sample response could not be rated as demonstrating readiness or not by one participant. While many samples represented borderline examples, all participants eventually arrived at a decision with the exception of this one sample.

Figure 3 provides a summary of participants' judgments about whether each sample demonstrated readiness for college. As shown here, one essay that was scored a 3 was also deemed to indicate college readiness by the majority of the instructors (7 of 8). In addition, participants did not fully agree on the college readiness indicated by two samples that were scored a 4 on the HSPA scale—only five of eight instructors said these two 4-level essays indicated college readiness.

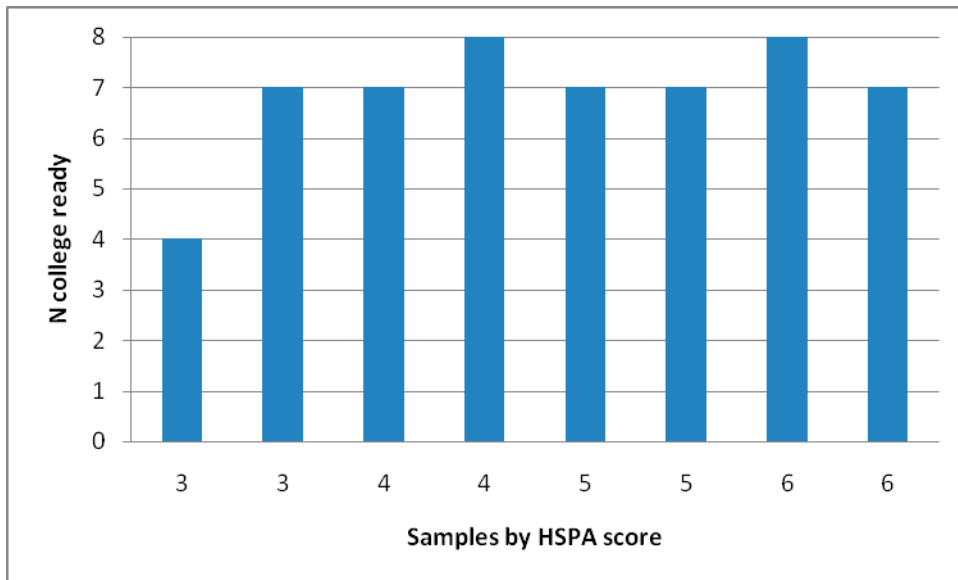


*Figure 3.* Number of college ready judgments by HSPA score.

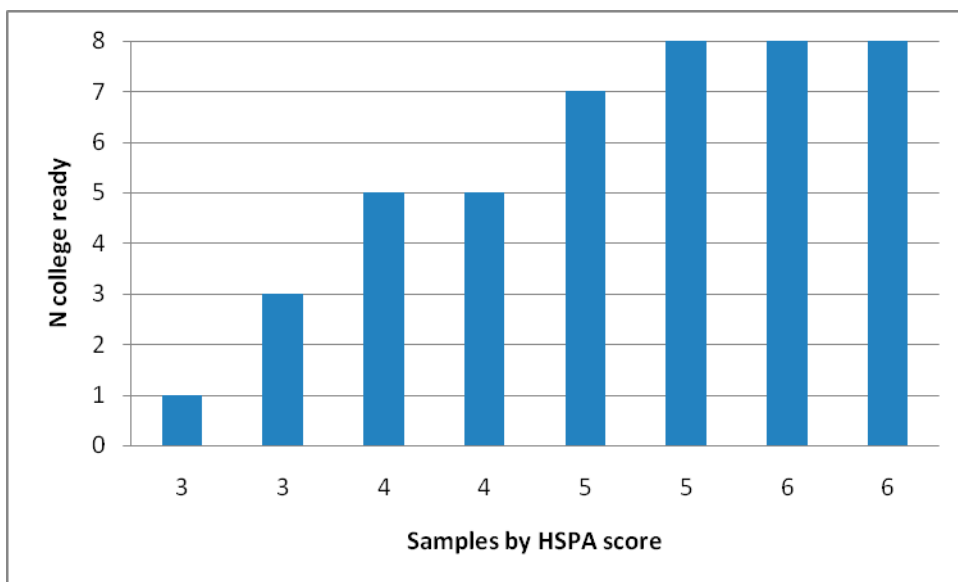
As shown in Figure 3, it is interesting to note that approximately half of the possible ratings for 3-level samples were deemed to show college readiness; this is of interest because 3-level papers represent the top of the lower end of the scoring scale and

are sometimes questionable indicators of grade-level skills. According to the New Jersey Department of Education (2008a), papers scored at the 1-, 2-, or 3-level “tend to lack command of standard written English and are, therefore, unable to convey the task message sufficiently for the purpose and audience intended” (p. 9). However, one 3-level sample in this set was judged to show college-readiness by 7 of 8 participants, and another 3-level sample was judged to show college readiness by half of the participants. On the most general level, this figure shows that sample responses on the HSPA do indicate readiness for college and do so even at lower-middle range (score 3) at times.

Another way to examine the overall descriptive statistics is by type of writing task. The first 10 samples in the set of 20 asked students to write a persuasive letter to a school board that was proposing the creation of a student-led disciplinary committee, while the second 10 samples in the set of 20 asked students to write a speculative or narrative piece of writing based on a picture showing two people leaning against a car. When college readiness is examined by these two types of writing, a slightly different picture emerges. Figures 4 and 5 provide the judgments of participants by score level and prompt type for samples rated at the 3-, 4-, 5-, and 6-levels.



*Figure 4.* College-ready judgments for persuasive essays by HSPA score.



*Figure 5.* College-ready judgments for speculative essays by HSPA score.

While 5- and 6-level essays were seen as showing college readiness by the majority of participants across both types of writing situations, samples rated at the 4- or

3-level were more likely to be rated college-ready in the persuasive genre than in the speculative/narrative genre. There are several potential reasons for this discrepancy that have to do with the testing situation and the process by which the sample responses are scored according to the HSPA rubric.

First of all, the speculative/narrative writing prompt only allows the students 30 minutes in which to write their response, while the persuasive prompt is administered in a 60-minute testing period. With the more limited time frame in mind, some papers are perhaps scored at the 3- and 4-level by HSPA raters because their understanding of the testing conditions take the constraint of the administration into account. In addition, participants in this study mentioned that judging the college-readiness of a speculative/narrative essay was challenging because the nature of those writing tasks were atypical of their expectations of writing at the college-level. Overall, the judgments of college readiness by participants in this study indicate that the HSPA samples at the upper-end of the score scale (5 or 6) consistently meet with expectations (with a few exceptions by some raters).

Likewise, papers at the lower end of the HSPA scale (1 or 2) consistently were seen as lacking evidence of college readiness. Samples in the middle of the score scale (at the 3- or 4-level) provide some mixed messages about the ability of the HSPA to indicate college readiness. Samples at the 4-level were likely to be judged college-ready for the persuasive prompt, and somewhat likely (5 of 8) to be judged college-ready for speculative prompt. Samples at the 3-level were more likely to show college-readiness when responding to the persuasive prompt. It is important to look at how the HSPA score levels related to college readiness judgments because the same holistic scoring rubric is

used across both prompt types. However, examining only the score points assigned as part of the HSPA does not provide the full story of how the HSPA relates to college expectations.

When thinking of how scored responses on the HSPA relate to judgments of college readiness, it is particularly important to examine why participants made their judgments. Because I interviewed each instructor on the nature of the judgments about each essay, I gathered in-depth information on what characteristics of the essays they believed were showing or failing to show college readiness. In order to determine the characteristics that were related to participants' college expectations, I coded interview transcripts in a variety of ways as presented in chapter 3. The resulting themes evolved from categorical aggregation of instructors' judgments and their statements about why they made those judgments.

Table 10 provides a summary of over eight hours of recorded talk about the participants' judgments of the sample essays. While it is impossible to provide a summary that captures each participant's discrete opinion, the participants' judgments tended to converge around particular issues in the essays. In the far right column, I provide codes that at least two participants mentioned when discussing their decisions of that particular sample. This table of characteristics provides an overall picture of what was important in constructing a judgment about how each of these essays related to college expectations.

It is important to note that the codes in Table 10 apply to both positive and negative judgments of the essay being examined. For instance, the code "diction" could refer to both strong and weak diction depending on the judgment of the essay. However,

the point of grouping co-occurring codes is that patterns across positive and negative judgments for generic categories were revealed. From constructing this summary, I found that communicating ideas with supporting evidence, providing adequate structure or organization, and using language appropriately were all associated with papers that were judged to indicate college readiness. Conversely, lower level papers tended to have little or no control over language and lacked development. In these lower level papers, the writer's ideas were shrouded by organizational or grammatical issues.

Table 10

*Summary of Characteristics of Essays Related to Judgments of College Readiness*

Essay	HSPA score	% college ready	Codes agreed by at least two participants
1	5	88% (7 of 8)	5 paragraph, focus, paragraphing, diction, importance of evidence, clarity, communicate, lack of development, lacking creativity, personal voice, predictable structure, transitions
2	4	88% (7 of 8)	Focus, paragraphing, ideas more important than errors, importance of evidence, importance of revision, clarity, communicate, lack of development, lacking creativity, understanding of audience
3	2	0%	5 paragraph, diction, importance of evidence, clarity, grammar, lack of development, lacking creativity,
4	6	100% (8 of 8)	Logical flow, critical thinking, diction, ideas more important than errors, importance of evidence, clarity, communicate, grammar, punctuation, transitions, understanding of audience
5	3	50% (4 of 8)	5 paragraph, paragraphing, clarity, communicate, grammar, lack of development, lacking creativity, predictable structure, transitions
6	1	0%	Grammar, language problems
7	5	88% (7 of 8)	Focus, creative approach, critical thinking, diction, ideas more important than errors, importance of evidence, clarity, communicate, grammar, personal voice
8	3	88% (7 of 8)	Focus, creative approach, importance of evidence, clarity, communicate, grammar, transitions
9	4	100%	Focus, creative approach, importance of evidence, clarity, communicate, transitions

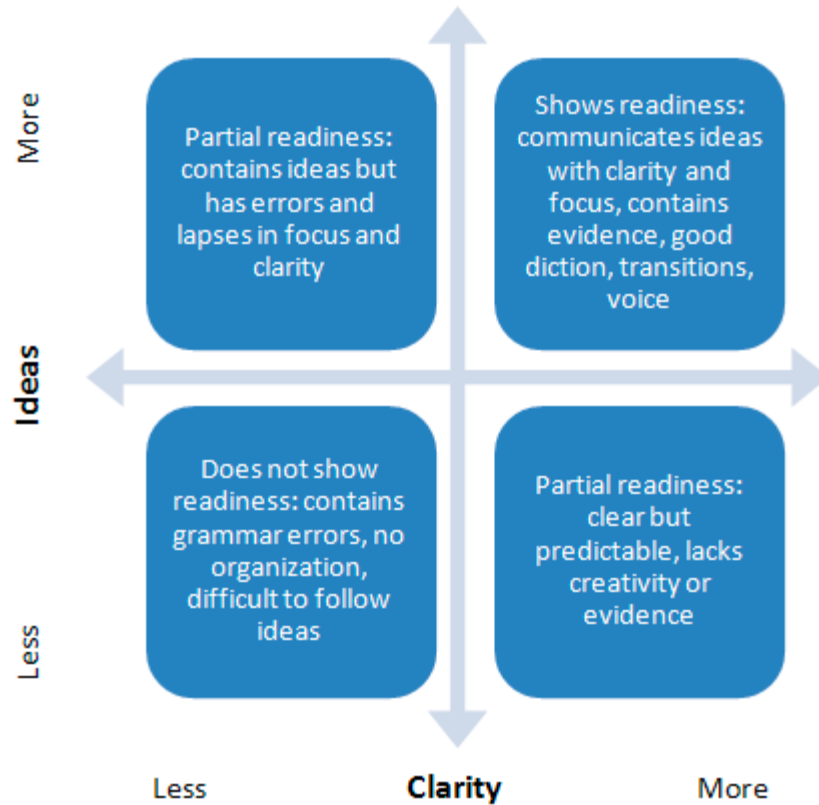
Table 10

*Summary of Characteristics of Essays Related to Judgments of College Readiness,  
continued*

Essay	HSPA score	% college ready	Codes agreed by at least two participants
10	6	88% (7 of 8)	Importance of evidence, clarity, communicate, lack of understanding of genre, personal voice, understanding of audience
11	5	88% (7 of 8)	Focus, creative approach, importance of evidence, clarity, communicate
12	4	62% (5 of 8)	Clarity, communicate, grammar, lack of development
13	6	100% (8 of 8)	Clarity, communicate,
14	2	0%	Lack of development
15	3	13% (1 of 8)	Focus, ideas more important than errors, clarity, grammar, punctuation
16	4	71% (5 of 7)	Focus, creative approach, ideas more important than errors, importance of revision, clarity, communicate, grammar
17	5	100%	Logical flow, diction, importance of evidence, clarity, communicate,
18	1	0%	Grammar, language problems
19	3	38% (3 of 8)	Creative approach, ideas more important than errors, importance of evidence, clarity, communicate, grammar, lack of development
20	6	100%	Diction, evidence of planning, clarity, communicate, grammar

It may be useful to examine the essay characteristics in the table above by showing their relationship in a quadrant chart. Quadrant charts have been useful in looking at how various factors or characteristics relate to some sort of outcome. For instance, quadrant charts have been used to indicate how factors such as clear leadership and investments in research relate to successful IT companies (Schulte et al., 2005). Similarly, here a quadrant chart can be used to display how a combination of

characteristics leads toward a description of a college-ready essay. Below is a quadrant chart that shows how the features of essays related to judgments of college readiness.



*Figure 6.* Quadrant showing how overall features of essays related to college readiness.

As shown in Figure 6, grammar and mechanics played a role when essays were not showing college readiness; however, essays that consistently showed college readiness were seen to have both clear and coherent organization and word usage, as well as have the ability to communicate ideas by using evidence in creative ways. The upper left and lower right represent components that are related to college readiness; however, in both cases essays characterized in these ways did not receive total agreement about indications for college readiness. For instance, the upper left quadrant represents essays that possibly contained good ideas, but the organization may have lacked clarity;



therefore, some participants did not rate them as indicating college readiness. On the other hand, the lower right represents essays that may have shown control over grammar, structure, and clarity, but the essays lacked creativity and did not move ideas forward in ways that indicated college readiness.

In my full analysis of the interview transcripts, I developed four key code families that informed the summary shown in figure 6. It should be noted that these code families (communicating ideas, clarity and structure, grammar and mechanics, and writing processes) align fairly well with the KSUS standards and the HSPA scoring criteria. In particular, my analyses of the transcripts revealed that two features of essays were most frequently cited as being the reason for making a judgment of college readiness: the ability of the piece to communicate ideas and the ability of the piece to create clarity and structure to those ideas. Interestingly, writing instructors from different educational contexts differed slightly in regard to the weight of each of these two important characteristics when judging college readiness. In the next section I will discuss the differences in opinions across participants from the various educational contexts and will use some examples of instructor judgments to describe those differences. The goal of such a discussion is to further examine whether the HSPA provides a clear signal about college readiness across educational contexts. In other words, the HSPA may signal a potentially differential “frame of mind” about readiness for college-level writing based on instructors’ teaching contexts, which is similar to finding from Smith (1992) and Conley, et al (2006).

### Differences in Judgments across Educational Contexts

The second question of this part of the study relates to examining HSPA sample responses to determine if some patterns existed when broken down into teaching contexts. It is important to note that these results are not intended to be statistically generalizable about how high school and postsecondary instructors differ in their judgments about college readiness. To answer such a question about these populations in general, a much larger sample size would be necessary and could not be accommodated in the time allowed to complete this study. However, because I collected in-depth data on instructors' judgments of college readiness, I can describe how instructors from across different educational contexts judged these essays and what patterns exist that could possibly be used to characterize expectations.

#### *Overview of Patterns of Judgments*

First, it is important to look at the overall judgments of those in each of the teaching contexts. Across all essays, the high school and university instructors said that 58% of the essays showed college readiness, while community college instructors judged that 72% of the essays in the packet demonstrated college readiness. The table below provides the ratings that each of the instructors gave to the essays by score level. For example, the first row presents essays scored a 6 in the *2008 HSPA Criterion-Based Holistic Scoring: A Writing Handbook* (NJDOE, 2008a). In the packet, there were four 6-level essays. In total, three high school instructors provided judgments on the four 6-level sample essays included in their packet, resulting in a total of 12 judgments from high school instructors. Of these 12 possible judgments, eleven indicated college readiness, or

in only one case did a high school instructor judge that a 6-level essay did not meet expectations for a college-level student.

Table 11

*Percentage of Essays Judged to Show College Readiness by Teaching Context and HSPA Score*

Score	N samples	High School		2 Year		4 Year	
		N ratings	Ready	N ratings	Ready	N ratings	Ready
6	4	12	11	12	12	8	8
5	4	12	10	12	12	8	7
4	4	11	9	12	11	8	5
3	4	12	4	12	8	8	3
2	2	6	0	6	0	4	0
1	2	6	0	6	0	4	0

As shown in Table 11, community college instructors were more likely to judge samples as demonstrating college readiness than their high school or university counterparts. This was especially true for samples at the 3-level. The high school and university instructors were less likely to call a 3-level sample college-ready, while the community college participants called 3-level essays college-ready eight out of 12 times. There were four essays scored a 3 in the set of twenty samples: two were persuasive essays and two were narrative/speculative essays. As mentioned in the previous section, the community college instructors called the five-paragraph persuasive 3-level essay ready for college. However, the fact that community college instructors were the most lenient in terms of judging an essay ready-for-college is perhaps not as dissonant as it may seem.

If the levels of education are viewed as linear progressions of advancement, it might seem odd that high school teachers were less likely to judge essays as showing college readiness than were any type of postsecondary instructor. However, these three educational contexts are perhaps not linear in terms of student expectations. High school and college instructors are likely dealing with higher-level performers, while community college instructors are possibly teaching students that were not motivated to go to 4-year institutions or lacked support for those educational opportunities due to their life situations. Without going too far into interpretations of these data in this chapter, as the next chapter will deal with interpretations of results in broader scope, it is important to note that Smith's (1992) results on the recent teaching experiences are similar to these findings. If community college instructors are generally working with lower-level students (on average) then their judgments of college readiness based on essays will be informed by those experiences. In other words, the HSPA lower-middle responses may align better with the skills and expectations for 2-year postsecondary contexts, while for those that are trying to gear students toward the 4-year institutions only the highest-level samples will align to skills required in that context.

#### *Characteristics of Essay Judgments Across Teaching Contexts*

From the results presented in the previous section, it is clear that there are some differences in the frame of mind about what it means for an essay to be college-ready based on these instructors current teaching contexts. To probe that finding in more detail, I will now consider what instructors indicated in making their judgments that might help to explain why they made such judgments. In describing this situation, I will use concrete

examples of what instructors said about particular essay samples to point out patterns in judgments.

There were several essays in the selected HSPA samples that could be called classic five-paragraph essays. Some of these essays were judged as showing college-readiness by most instructors. However, one particular essay example split the group; four judging it as showing readiness and four judging the essays as not showing college readiness. The four who judged this bare-bones five-paragraph essay as showing college readiness were currently teaching at the community college level or remedial courses at the college level.

For example, all the participants who rated this essay as demonstrating readiness justified their judgment by citing the clear organizational pattern of the piece. Eileen noted that the writer appeared to have good control over paragraph-level organization, saying that “they know the difference between a thesis statement and a topic sentence” (interview 3, line 48). Two other participants said that the paragraph transitions by using first, second, lastly lent the essay a sense of clear purpose. For instance, Greg said, “Good organization in terms of using those first, second, and last transitions” (interview 5, line 51). These classic cues were seen as very clear sign posts for the reader.

Ray went into some detail about the organization, making an argument that while the organization was predictable, it did provide a clear focus for the essay.

Ray: It’s sort of pedestrian, but again that’s usually adequate, that’s usually judged as an adequate piece of writing. I mean when we look at a holistic scale, that sort of pedestrian organization, first, second, lastly, just as they have it, and of course that’s organized. And in conclusion.

Robert: So you would rather see a nice organized piece than one that’s not, that has more details but is less organized around some themes?

Ray: I mean you could have this kind of pedestrian organization, and I don't use that pejoratively—that's just a basic way to organize things for somebody who's not that experienced of a writer, but, and they may present an adequate amount of detail, but you may have somebody that is very into the topic and they will generate more information but in what could be a scattered organizational pattern or more free form organizational patterns, but then the question of focus is not being maintained. So, I wouldn't say if it's less predictable in its organization that it's not an effective piece, but I would be looking at focus as well.

Robert: Okay.

Ray: So even if this may be competent or just adequate, I think the writer maintains a focus. (interview 6, lines 46-50)

In this dialogue, Ray is clear about his commitment to making sure essays are organized and focused. In fact, this participant makes it fairly clear that focus and clarity are valued as much as, if not more than, the ideas contained in the piece. By making the point about essays that lose focus despite the presence of good ideas, it is clear that his central focus is on creating ways for writers to clarify their positions. The ability to focus an argument is most certainly a critical aspect of success for college-level writers. The participants that judged this essay as showing college readiness were more concerned with organization and focus than they were with the development and communication of ideas.

As opposed to those who thought the five-paragraph essay was acceptable, the instructors who thought it was unacceptable focused on ideas over organization. While most of the participants who rated this essay as not showing readiness for college acknowledged that there was organization, they felt that the organizational pattern was too predictable or simplistic to indicate they had the capacity to develop a thoughtful piece of writing. For instance, Joan clarified her position on the predictable structure when saying, "A lot of students say I grade hard, but what I look for is using something

other than first, second, third, the stronger transitional words that make your essays pop. “ (interview 1, line 14). Cathy had a similar reaction and said that “it has the structure but it’s a basic structure, basic transition. To me it doesn’t indicate college readiness. And it does not elaborate and it doesn’t develop the ideas” (interview 8, line 26).

So while the piece may have structured the argument against creating an honor council, the lack of thinking evident from adopting this predictable structure caused readers to feel that the essay was not developed as they might expect from a student ready for college. Phil summarized the lack of development in a similar way.

There is evidence to suggest that the writer knows how to take an organized approach through a cookie cutter persuasive essay, which are three reasons in the middle. The student knows how to do that. They do get an hour to do these things to develop some reasons, and just to think anecdotally, or to think about personal experiences that might relate to the topic, or to think about a similar experience that might relate to the argument. Again I'm seeing the language of superlatives that to me really isn't appropriate, "They will get this. They will get that." Well, how do you know? If you do know that for a fact, then show me by developing a reason. (interview 2, lines 75-76)

In this quotation, it is clear that the “cookie cutter” response would not indicate college readiness for this instructor. Phil wanted to see some kind of thinking in the writing and needed to see some production of evidence that would support that thinking. The idea that timed writing assessments create situations where students are encouraged to provide empty “blather” was established by Hillocks (2002a) and that same principle appears to be at work for this instructor. The writer of the five-paragraph essay says that certain outcomes will occur if the school board adopts an honor council, but the essay does not provide sufficient detail to support their claims. In this way, we can understand that supporting claims with sufficient evidence is an important aspect of college readiness.

Therefore, in terms of alignment, the HSPA would need to open a greater potential for students to build essays that contain evidence.

A similar sentiment was stated by the college professor who felt this five-paragraph essay was not a successful indication of college readiness. She said that “a machine or a person who is looking at this test situation might say that’s pretty good, but it’s not going to make it. It doesn’t say anything. It’s just words” (interview 7, line 103). Earlier in our interview, this participant discussed her position on having electronic reading engines assign scores for placement and other large-scale performance-based writing assessments. This instructor is suspicious of all such technologies because, while they have complex statistical models and the ability to recognize grammatical patterns associated with higher level essays, computer programs don’t actually understand the meaning of what is being addressed in a piece of writing. This instructor was essentially referring to this same sort of empty rhetoric in this five-paragraph essay. Without anything real to communicate, this instructor felt that the student would not be able to succeed at the college level.

As mentioned in the literature review, Johnson et al. (2003) found that collegial pressure created by the state assessment led the teacher-researcher to adopt the teaching of the five-paragraph essay. Hillocks (2002a) mentioned how state assessments in writing allow for the common forms of the current traditional rhetoric, such as the five-paragraph form, to become accepted norms despite the lack of critical and creative thinking that writing in such forms encourages. The judgments about the five-paragraph essay support the claim that community college instructors were more focused on organization (as described in the previous section). Overall, the five-paragraph essay uses a form and has



a clear and predictable structure; however, the high school and college instructors thought this essay failed to indicate college readiness because it didn't develop the ideas with any detail or creativity. While the community college instructors mentioned that the five-paragraph essay wasn't completely developed, they did find that it met college expectations because it used clear organizational elements, albeit noted as "pedestrian" by one community college instructor.

Another way of looking at the focus on organization and structure can be seen by examining another essay that everyone regarded as showing acceptable skills for college readiness except for one college instructor, Kim. As opposed to the university instructor who focused on ideas, this university instructor more frequently sided with the community college instructors. Kim was teaching remedial writing at the time of our interview and was concerned with writers showing the ability to control their pieces with a clear organizational structure. With regard to the one 3-level sample that all others said indicated college readiness, she said "there were some ideas in here and it was three pages long, so he obviously has some details in here. But it was just sort of all over the place in terms of a lack of organization" (interview 4, line 44). While all other participants said this essay indicated college readiness, this instructor's focus on organization deemed it insufficient at demonstrating college-level skills.

Other instructors agreed with Kim about the lack of organization in this particular essay, but they still judged that the essay indicated college readiness. Joan said this "student has developed interesting ideas in a very limited amount of time, doesn't yet necessarily know how to organize information so effectively, but has what to me suggests the basis for developing them later in a sophisticated fashion" (interview 2, line 108).

Also, Cathy said “there was a little bit of clean up that needed to be done with it. One thing, I did think that it was good that was using a text to problem connection, using literature to argue for their problem” (interview 8, line 40). So here we see that all instructors acknowledge the weakness of the piece, but they were willing to tolerate some weaknesses over others. These high school instructors valued the ideas in the piece over its lapses in organization. However, for the instructor currently teaching remedial writing at the college level, organization was a key factor and one that she required for an essay to demonstrate readiness for college. This fact, again echoing Smith’s (1992) finding, provides some evidence that the most recent teaching experiences inform how instructors define their expectations.

In remedial courses and at the community college level, the focus is on establishing a set of tools for the writer to use in later courses, while the focus of university level writing is on the production of ideas (Perin & Charron, 2003). In essence, these results show that the HSPA may provide a more consistent signal for the type of skills required by community college instructors. The HSPA prompts, scoring guides, and sample responses align better with aspects of measuring clarity, organization, and structure in writing. However, the HSPA has a weaker alignment in terms of measuring student ability to communicate ideas—a key value for high school and university instructors—and, therefore, signals 4-year college readiness better at the highest level of the student response scale.

#### Can the HSPA Indicate Preparedness for College-Level Writing?

At the end of the interview, after having gone through their judgments on each sample essay, I asked participants a few summary questions. These questions were

designed to glean their overall impressions of whether the HSPA writing test could indicate readiness for college. Since participants typically had previous experience with the HSPA or similar large-scale writing tests, they often used their teaching experience or experiences as a scorer of tests when answering these questions. Also, because participants had just gone through their judgments on each essay and had previously done the coding tasks involving the HSPA scoring criteria, they were well-versed in the details related to the HSPA writing test. In the next section I will go over their impressions on whether the HSPA can, as it is currently designed, indicate readiness for college. I will cover aspects that they referenced as successful, aspects of the test that they deemed unsuccessful, and go over what they viewed as potential solutions in the alignment of the HSPA with college expectations.

*Positive Aspects: HSPA Measures Basic Competencies*

The HSPA is best suited to address some general competencies according to one community college instructor, Ray. “I think minimal competence needs to be there. Quality and correctness in writing. Adequate sense of organization and development. So HSPA speaks to that across the board” (interview 6, line 184). This was echoed by others. Harriet, a college instructor, thought that all standardized writing tests trained students about how to use some basic writing tools such as “vocabulary, how to have voice in your writing” (interview 7, line 40); however, she went on to discuss how looking at the training to use these writing tools produced a sense of emptiness regarding the meaning of writing.

The focus on basic competencies is also revealed in the results already discussed in this chapter. Organizational elements are easier for the HSPA to measure. Those

whose focus was on the organization of pieces were more likely to say that essays met their expectations of college-level students. However, for the high school and one of the university instructors who valued students' abilities to develop ideas, the HSPA is less likely to provide the information needed. Those who focused on ideas in general were less likely to say that these essays met their expectations.

A few participants mentioned that the HSPA did allow some insight into the way that a high school student can develop an idea. While all participants acknowledged the constraints of the testing situation, there was agreement that writing the persuasive essay was a valuable task and that it aligned with college expectations in some respects. One high school instructor, indicating that the timed aspect of the test situation isn't so unrealistic, said "the persuasive one, if they don't have that down pat for an hour of writing, that could give a university good enough idea about them coming" (interview 1, line 112). She thought it was similar to the SAT, which is also typically taken in 11<sup>th</sup> grade, and in her opinion the persuasive prompt aligns with the SAT writing task and could give colleges some indication of whether students need remediation.

Another high school instructor indicated something similar and went on to articulate potential ways that the timed situation also causes some problems for determining how well students can develop a complex idea or use evidence effectively.

It measures whether you can write a persuasive essay in 60 minutes so I guess you have to be um quick-thinking. You have to be organized and able to write in a pretty capable manner. Because that's not something you can fake. You have to come in and you have to, I mean there are ways to fake passing the HSPA, but I don't mean that in a bad way. I mean, we tell them all the time don't worry about what you really think. Make up your statistics. Make up your reasons. It just has to have your development. It doesn't have to be true. But to me that doesn't prepare a kid to how do I then take my own argument and sustain my own argument in college. (interview 8, line 111)

This sentiment was echoed by others. While the HSPA does analyze development of an essay, it measures the ability to develop an essay in a way that is outside of the understanding of audience and outside of the context of source material. While the HSPA does ask students to develop an idea, the scoring criteria do not take into account the relative value of the evidence they selected to use.

In particular, because the test asks students to write without much context, the majority of participants thought that the HSPA writing test would not be able to speak to the ways in which college students are expected to develop ideas. Indeed, when asked to indicate positive aspects of the HSPA, participants had relatively little to say; however, participants had much more insight about how the test failed to meet college expectations and what could be done to improve the situation. This was a particularly interesting finding because instructors maintained this position on the HSPA despite having indicated that sample responses did demonstrate college readiness.

*Negative Aspects: HSPA Fails to Measure Development of Ideas*

While basic skills related to grammar and organization were thought to be appropriately measured by the HSPA, most participants mentioned that the HSPA failed to align with their expectations for how entering college students should be able to communicate their ideas. Harriet, a university instructor, mentioned that the strict adherence to forms contributed to students' inability to think creatively. In her opinion, students were focused on superficial aspects of the form rather than on communicating ideas and using a variety of forms to meet the communicative needs of the situation.

"You can lock students into a writing pattern. You know like, like in my conclusion do I have to restate my thesis kind of thing" (interview 7, line 36). This particular instructor

made generalizations about public education and said that she felt the entirety of high school writing instruction (not just the assessments) failed to gear students toward the thinking that is required in college courses.

That gap between the high school and university—like I said some can make the leap and some are saying oh my, this is so hard for me. I wrote the five-paragraph essay. This is what I did and this is all I know. So when I say under-prepared they are not really challenging them and very little critical thinking goes into the writing and it's just very sad. Really it's frighteningly sad for me and I just don't think high schools are doing their job. I'm talking about public high schools, most students that come in through private high schools are better but not by miles but they are better prepared. They are better prepared to make adjustments and break some molds. (interview 7, line 36)

While Harriet's opinion is constructed from her personal experiences with students at her school, it does show that there is a potentially wide-spread opinion about the failings of public high school writing instruction. The adherence to forms and lack of critical thinking is at the core of this instructor's viewpoint of the misalignment of high school and college writing. While this quote was addressing curriculum and students on a general level, the overall idea of misalignment is supported by a range of data collected in this study about the HSPA assessment. From looking at the characteristics of essays, I found that the focus on organizational elements was a major reason for identifying college readiness. However, some felt that even the highest level essays couldn't be deemed college ready because they failed to appropriately develop and support their ideas. Also, when looking at the scoring rubric, I found elements in the rubric that addressed idea-development at a lower depth of knowledge than the corresponding college success standard. Overall, the HSPA may not be accurately geared toward the importance of developing ideas.

This impression about the lack of ideas was mentioned by the high school instructors as well. The high school instructors mentioned that they geared their own instruction to what they conceptualized as a college standard—that is, a focus on ideas and supporting ideas with the use of appropriate evidence. However, the HSPA was not geared to measure students’ abilities to communicate and support ideas, as shown here from Cathy.

It’s not indicated whether you are able to make judgments about what kinds of evidence are good evidence. I mean it does in minor ways. Like if you know statistics are a good thing to use as evidence in terms of hierarchy of evidence, but it seems to me that it’s so often used in these contrived situations. If I look at college in terms of English writing, you are going to have to be able to analyze text in depth and take that text and sustain your argument and there’s nowhere where it [HSPA] does that. (interview 8, line 115)

This high school instructor would likely disagree with the perception that public high schools are not focused on teaching students to communicate critical thinking. During our interview, Cathy mentioned that she doesn’t gear her instruction to the standards of the HSPA, but instead uses a combination of standards that asks students to support their thinking with evidence, writing research papers, and also does not tolerate grammatical errors as much as the HSPA would allow. As shown in her quotation above, Cathy has defined specific expectations of what college writers must be able to do. When she thinks about the HSPA, it does not align with her own view of how college students should use evidence to appropriately communicate ideas.

Another high school instructor, Phil, had similar thoughts about the ability of the HSPA to measure students’ articulation of ideas supported with good evidence. When asked whether he thought the HSPA measured the skills he would expect at the college level, he said that the HSPA was okay for a high school graduation requirement, “but if

you are talking about the great leap between high school and college... I don't think it has any validity for me" (interview 2, line 230). Earlier in our interview, this participant already established that he felt the problem with the HSPA had to do with the use of evidence. In his construction of college expectations, there is a heavy reliance on research and critical thinking about content, but on the HSPA there is no opportunity for working with evidence because of the nature of the prompts and the lack of time to do research within the testing situation.

In a collegiate environment, you are going to take the time, do your research, gather and hopefully discern between credible sources. You are going to confer with others. You are going to revise. You are going to do all these things through stages and you are probably going to learn the hard way like I did freshman year of college. This is when the students, I think, learn how to write in a rigorous environment because especially in schools where we teach the tests in the high school, it's very dangerous. It's not natural and it isn't natural for us. I don't think as educators [we can] judge too much about what a student can and cannot do in a rigorously timed environment when we are not even giving them evidence. We are not even providing them with the tools to be successful. It's only the ones who are most adept at being creative and are most adept at gleaning from their personal experience who can even begin to convey a collegiate level of writing. (interview 2, lines 219-221)

Here again we see that this high school instructor is thinking about the college environment in terms of the importance of evidence and critical thinking. In his opinion, the HSPA doesn't provide this opportunity because there is no chance to engage with evidence in the test situations. The prompts provide a starting point, but they do not include the option to substantiate claims with outside information.

It is interesting to note that this instructor believed that high-scoring students were those who could produce evidence based on personal experience. This same sentiment was voiced earlier by multiple participants. At the core of test alignment is a validity question, and that validity question seems to indicate weakness when multiple instructors



point out that the quickest creative thinkers are also likely those earning scores at the top of the scale. This begs the question of whether the test is measuring writing proficiency or simply a general ability to think quickly and critically.

As these interviews have shown, critical thinking needs to be part of college-level writing. Perhaps the assessment needs to provide opportunities for developing that type of critical thinking. Because time is constrained, the act of writing for a test is in fact very different from the act of writing for college. While success in the test condition may correlate to success on the task itself, it is clear based on these descriptions of the writing portion of the HSPA that the performance-based assessment does not represent the act of writing. In fact, one might consider it a completely different construct within the domain of writing. If at least some aspect of college-level writing is about research, critical thinking, and selecting appropriate evidence to support an argument, then the HSPA does not align with those aspects of college-level writing.

The community and remedial college instructors noted that the HSPA failed to indicate how students would perform when asked to synthesize text to create an argument. Kim said that “I would say the majority of college work is more reaction or reading-based essays rather than just what do you think of the school board’s new policy” (interview 4, line 116). This participant went on to say that they often have to teach their remedial students how to summarize and then synthesize that summary into their argument. The HSPA doesn’t require an in-depth articulation of an argument related to a text; the reading and writing portions of the test are administered and scored separately.

Other participants saw similar weaknesses in the HSPA. Eileen said, “You have no way of knowing what, if they can do any kind of research or document anything. Or

synthesize information. Or write a summary of anything” (interview 3, line 199). This overall need for the HSPA to address working with evidence and text was mentioned across all participants in one way or another, showing that incorporating writing with text interactions is clearly an important college-level expectation.

Overall, a generally agreed upon aspect of misalignment between the HSPA and college expectations was the lack of opportunity for evidence or text interactions to support an argument. Supporting an argument is clearly an important aspect of college-level work, but the HSPA writing assessment attempts to measure writing in the abstract. The focus on the HSPA is on the use of writing forms, but fails to acknowledge the connection between content and form. Also, the HSPA did not provide students with the time or reference text to produce reasonable evidence for an argument. The highest performers on the HSPA used their personal experience, but perhaps that reliance on quick thinking under-represents the construct of writing expected at the college level, seeing as revision and reflection are noted to be important aspects of college writing (Yancey, 1999).

*Negative Aspects: Narrative Irrelevant and Genres Questioned*

In addition to the lack of alignment with college-level expectations with regard to communicating ideas, participants generally agreed that the narrative writing portion of the HSPA was less relevant to college writing. Kim said, “I’m not sure that having them do a narrative is necessarily that valuable in terms of determining the college-level skill” (interview 4, line 110). In general, participants thought the persuasive writing task aligned better with the types of writing expected at the college level; however, as mentioned above, the persuasive task still did not allow for the equivalent level of text

synthesis and use of appropriate evidence that would be expected at the college level.

Neither task effectively aligns with real college writing assignments or expectations.

Some participants thought the narrative tasks were culturally irrelevant, or that the tasks were not engaging because of their abstractness from the real life of students. Joan said,

As for the picture prompt, sometimes it's a little tough for the kids to connect to it because they might not understand what in the world is going on in that picture and it might take them that much longer to think about what to write. They start writing and run out of time because that's 30 minutes. (interview 1, line 112)

Joan went on to clarify that one potential problem with the narrative writing prompts may be their lack of appeal to students from diverse backgrounds. "The community I teach in is very diverse. I get a large majority of the students from Africa, India, Pakistan, and Egypt and there's a lot that they cannot connect with" [in the prompts] (interview 1, line 122). Narrative writing is perhaps more difficult to force into a strict five-paragraph response, therefore the nature of the students' relationship to the task or prompt likely has an even greater reliance than it does for the persuasive essay. Likewise, studies by Langer (1984) and Sullivan (1997) have shown that knowledge of the subject and cultural background knowledge have various effects on student performance.

However, there was an alternative view to the narrative prompts. One high school instructor, Phil, mentioned that the narrative prompt might be a better way to judge writing in the abstract because persuasive writing demands research and use of cogent evidence. Given this impression the narrative is a more natural or isolated writing task where the student is expected to only use their inner resources. The difference between these two opinions points out that there is a weakness in relating the HSPA writing assessment to natural writing conditions. On one side of the argument, narrative writing

prompts may not make sense to students, and therefore their motivation and ability to succeed may be compromised. Alternatively, writing a persuasive argument without access to real resources is not at all like expository writing at the college or even high school level which requires an argument to be based on relevant evidence or a reaction to a text, and, therefore, the narrative prompt is superior given the testing situation.

The state of New Jersey recently announced that the 2010 writing assessment will not include the narrative picture prompt. It will be replaced by some type of expository or descriptive writing prompt that was still unclear to the participants of this study at the time of our interviews (as mentioned in the next section). It is important to note, however, a final thought on the issue of having narrative and persuasive, or any types of writing categorized in these generic terms. Some participants felt that having different forms, including narrative, gave students an opportunity to use different organizational structures. Ray mentioned the importance of having students understand “flexibility in organizational patterns based on the kind of writing that the student is being asked to do. So obviously you can organize a narrative paper one way and can organize a persuasive essay another way and a higher-level argument essay a third way” (interview 6, line 178). Again, from this community college instructor’s perspective, there was a focus on organization and structure and therefore knowing how to perform in the generic writing situations was reported as valuable.

On the other hand, the high school and one of the college instructors were more concerned with the creation of meaning. In comparing the generic responses that are acceptable in a testing situation to the types of writing that is allowable in Harriet’s

portfolio assessment, it is clear that the college experience (at least at this institution) is focused on the idea of making meaning, not simply on applying a form.

We let them use lab notes, I let them use song lyrics if they had something they wanted to put in there and they are proud of their poetry. You could tell a lot by a student's poetry. There's something in that. But it's time consuming. You know a test is much simpler. But to me the portfolio tells me much more about a student than just let's sit down and do this. (interview 7, line 46)

This was not so much about being flexible to use forms, but in using elements of forms to communicate an intended meaning to an intended audience. In this quote, the portfolio project is about showing a variety of aspects of the student as a writer, and the instructor mentioned that students write reflections on their portfolio pieces. By including various forms of writing and writing reflectively about them, the focus at the college level for this instructor is focused on creating a meaningful set of materials. This is different from a process of applying a formulaic structure within a timed situation. The difference may be a subtle one, but I think it is an important one that underscores the results presented throughout this study: there is a divide in the field over how organization relates to the articulation of ideas and the creation of meaning.

### *High School Teachers Felt a Lack of Transparency*

Before talking about how instructors voiced solutions for creating a test that aligned with college expectations, I want to make a final note about the high school teachers' reaction to the test. As I found in two previous pilot studies, high school instructors found that there was a lack of transparency for them with regard to the way the HSPA operates. This lack of transparency was noted in two distinct ways: first, the test was not related to the curriculum they were trying to enact on the high school level; and second, the politics and details of the test administration were not clear. This lack of

transparency appeared to erode participants' perceptions of their roles as professionals in the high school classroom.

The HSPA was criticized for its lack of relationship to the high school curriculum on several levels. Phil articulated the problem as an issue with the test situation itself. "We also know that the best writers in the world are the best revisers. They are the ones who take the time to improve upon their work" (interview 2, line 237). So the exclusion of revision directly contradicts what these instructors do with their students and how they are teaching them to be successful writers. This participant went on to say how a change in the test could be made to focus energy on a more relevant and complete goal. If the test aligned with a view of writing that allowed for revision and the use of evidence, "we now have liberty as a school district to fit that into our curricula and to make sure that our students know how to do that" (interview 2, line 237). However, within the current system, teaching empty forms does not fit the district's curriculum or best practices in the field, so there's an inherent mixed signal being sent to instructors on how to incorporate the test into their curriculum.

The issue of signaling is important to note because as referenced earlier, the HSPA sends mixed signals of success to students via the scoring criteria and via their sample responses showing readiness for college. If the test does not gear students toward advanced educational opportunities, then students should understand that their success in high school does not signal preparation for success in college. Specifically for writing, Cathy articulated that the lack of consistent signaling on the writing process causes her to question the validity of the test.

We spend so much time teaching them that writing is a process. You have to spend a lot of time to plan it and write and then go back to it and rewrite it. And

then that's it, one shot, do you want school uniforms, there's no way of preparing for that or even showing that you can go back and work on it again and take another look at it. It's kind of counter intuitive to teach it as a process and then turn around and say this is it. We are not going to indicate, or we are not going to look at all whether you've engaged in a process, just whether you can write it in the confines of this setting. There's something to be said about be able to perform on the spot, but I think that you should also take a look at some of the other work that is being done to get a better look at what a student can do in their entirety. (interview 8, line 129)

This quote clearly shows that this instructor perceives that there is lack of consistent signaling between the test and her pedagogy, about what it means to be a good writer, or how to engage in best writing practices. This lack of consistent signaling as students move through high school also plays out when students arrive at college.

The mixed signals between college and high school were articulated by Harriet. In talking about the portfolio project she assigns, she mentioned that “when [students] build their portfolio they have to include a reflective letter and what I’m getting in the reflective letter is ‘I wrote this in high school... My teacher gave it an A, but now that I’m reading it again, I really don’t think that it deserved an A.’” (interview 7, line 34). Or this same instructor mentioned the mixed message a different way when students are surprised about their grades in her course. She hears “that old song, ‘I was in honors and now I can’t get above a C in your class.’ Well, you might need to reflect on that observation” (interview 7, line 23). These moments show that students in this transition understand and voice the frustration over the mixed signals. Determining whether they are “good” at writing is complicated by many factors, but one of them is likely that the curriculum at the high school level is torn between focusing on the HSPA requirements versus focusing on the necessary skills for writing in college (or simply best practices in writing).

Another way that the HSPA does not align with the high school curriculum is that there is a lack of focus on content. This correlates with the results mentioned earlier dealing with students not having opportunities to use real evidence when making an argument. However, this aspect also relates back to a lack of transparency between the test and the high school curriculum. At the high school level, teachers are typically teaching writing in relation to a set of texts of some variety or another. However, the HSPA does not embody this same focus. Cathy articulated the problem as follows.

A weakness is that there's no core of content [associated with the HSPA]. So there's nothing to say that when they graduate, they know what this device is and that device is and this period of literature and that. And I know that that is probably not very popular, but I do believe you need some core content knowledge. Instead of just everything being extemporaneous you know off the top of your head. (interview 8, line 117)

While establishing a uniform core set of content would likely result in problems across schools and districts, clearly the teachers who took part in this study felt a little lost without some connection to real content. Writing might need to be defined on local levels in schools and districts so that there is content embodied in the writing work that students do. Of course, in classrooms students do work with content, but the lack of transparency when it comes to the test obscures how success on the HSPA equates to past and future educational endeavors that students are likely to face.

In addition, there was an overall lack of clarity for the high school teachers in terms of their understanding of changes in the assessment or how prompts are constructed. One instructor mentioned that people who write the prompts should form more panels and visit more schools so that there would be a greater understanding of what students do in class. Also, she went on to articulate the problem of changing the test without fully informing or preparing teachers and students: “come March it’s something



different [than the picture prompt] and that determines whether you graduate or not” (interview 1, line 128). When new elements of the test are planned, this teacher perceived that students and teachers were at risk of not being prepared for the change; therefore, students may require retesting or risk not graduating. Indeed, this is a serious consequence.

### *Solutions Suggested by Participants*

When I asked participants about the ability of the HSPA to indicate college readiness, all the instructors provided insightful suggestions for improving writing assessments like the HSPA. Overall, the instructors agreed that the HSPA (and other writing assessments like it) ought to contain some way for students to react to a text and create an argument with evidence; specifically, a few participants suggested simply making the HSPA prompts more text-oriented, while others wanted to break the mold of the single-draft test in favor of a portfolio model.

The current writing prompts on the HSPA ask students to respond to a situation. Of course, the prompt itself does provide some text from which the student can orient their writing; however, the prompt is typically short. For instance, the text for the prompt about writing a letter to the school board about a proposal to form student honor councils follows.

#### Writing Situation

Your school board has proposed creating an Honor Council of students. This council would be responsible for disciplining students accused of violating school rules. The proposal has caused controversy in your school.  
You decide to write a letter to the school board expressing your opinion of the proposal.

#### Directions for Writing

Write a letter to the school board either supporting or opposing the Honor Council. Use reasons, facts, examples, and other evidence to support your position.

Clearly this three sentence scenario provides no real text to synthesize. The situation is a purely hypothetical one, and the student must assume one position or another in the letter. The directions explicitly state that the student should support one side or another; however, a thoughtful consideration of the issue may lead students to consider the benefits and costs of creating an honor council. All participants viewed this type of writing situation to be inauthentic and irrelevant to college writing situations.

Eileen said that throughout her decades of teaching she has “had to teach more about how to summarize and synthesize” (interview 3, line 201). The focus on the need for students to synthesize multiple texts was articulated similarly by Ray. He mentioned that most of their assignments focus on synthesizing a reading.

For instance we might give them two writers that are discussing an issue like global warming but they are on opposite sides. And ask them to evaluate the two arguments and ask them to explain which they find most compelling and of course that task has to be based on evidence from both texts. So that’s something again. It doesn’t have to be something, a skill that is very, very developed at the entry point. That’s something that occurs during those courses. (interview 6, line 180)

While this participant was talking about the essay assignments they tend to ask students to do in their courses, he did go on to mention that the HSPA would align better with college expectations if it integrated reading and writing. In this participant’s view, the writing task would still be oriented toward the creation of first-draft writing, but would include the important aspect of using real evidence to make a cogent argument. Other instructors across educational contexts agreed with the overall sentiment that the current

writing assessment prompts should have some level of text engagement, as that would better simulate the academic tasks required at the college level.

While all participants agreed that some textual evidence in a writing test situation would increase the ability for the HSPA to align with college expectations, some participants took a more global approach and suggested reconsidering the first-draft writing test paradigm. Harriet suggested that there be a combination between a HSPA-like test and a portfolio. “I think that I like the idea of portfolio building so that you can see the overview of the student’s work. A senior year portfolio or a senior year project along with some kind of assessment might be a good way of looking at what a student’s worth is in writing” (interview 8, line 46). I will not go into detail on the benefits of portfolio assessments because there are several definitive volumes on the topic (cf. Sunstein, 2000); however, looking at student writing in natural conditions appears to have several benefits, and for this instructor, a portfolio was mentioned as a good mechanism for aligning the HSPA with college expectations.

Phil created a vision for a computer-based portfolio system. His vision for such a system accounted for so many of the weaknesses in the current writing test paradigm that it warrants reproducing and explaining here. Phil’s idea essentially involved multiple testing sessions in which students go to a computer lab (for test security and consistency reasons) to brainstorm and research, draft, and finally revise a full piece of writing. Phil supports this vision by saying that providing time for revision best reflects a comprehensive view of education between what success means in secondary and postsecondary settings. Unfortunately, Phil and others agreed that “There is very little

connection between learning and assessment with these state tests” (interview 2, line 238).

Phil went on to explain that using technology could accomplish two educational goals that have previously been separated between secondary and postsecondary domains. In his view of a new assessment, secondary teachers could directly teach to the test because the test would represent the type of learning and skills that students would be expected to have at the postsecondary level. In talking about a writing test that allowed for the entire writing process, Phil said “we can prepare them to be successful on such an exam, and we can prepare them to be ready for college” (interview 2, line 241).

The sentiment and vision for a new type of writing test as articulated by Phil was completely singular; no other participant considered the situation with as much breadth and depth as this participant. While I’ve tended to focus in these findings on moments of convergence of data, this vision for a new testing paradigm is so complete that it warranted inclusion on the basis of how the speaker has gathered sources of conflict between teaching and testing to create a complex solution.

In the explanation of his vision and quotations presented above, we can see that Phil considered all aspects of the writing process: reading, researching, planning, drafting, and revising. By creating an authentic situation, the vision for a computer-based portfolio project would allow students to develop and communicate ideas as well as show their skills with grammar and organization. Also, the computer-based aspect of the scenario described would allow for data to be captured on how students research, how they plan, how their plans match to their drafts, and how their drafts match to their revised essays. By creating a longer project as an integrated assessment, this vision solves

the majority of the misalignments discussed previously in these results. With such a project, students could use evidence more effectively and they could be given the opportunity to revise.

In addition to solving content-oriented issues, this vision solves a more systemic conflict between teaching and testing. Creating an integrated assessment task would allow for instructors to feel engaged with the state testing implementation and understand that the test is actually geared toward what they know to be best practices for writers. Such an integrated task would provide clearer signals to students, teachers, and administrators (at the secondary and postsecondary level) about what it means to produce a successful piece of academic writing. In doing so, there would be better alignment between the state's assessment and college expectations.

#### Integrating Alignment of Scoring Criteria and Sample Responses

Now that I've provided in-depth information about how the scoring criteria align with college standards and how benchmark samples align with college expectations, I will consider how these two data sources converge to indicate how the HSPA aligns with postsecondary expectations. Looking at the scoring criteria allowed for the collection of data on how the state defines success relating to the test and how that definition of success on the test relates to college standards. By looking at the sample responses, I described how performances on the test relate to the writing abilities that are expected of college writers. While the data collection focused on different aspects of the HSPA, a similar story emerges about how the HSPA aligns with college expectations.

From looking at the scoring criteria, I found acceptable levels of alignment for most areas of the Webb (1999) alignment criteria. However, there were some aspects of

the scoring criteria that did not align with the KSUS standards. Most notably, the HSPA scoring criteria did not address the KSUS standard E (communicating ideas, concepts, emotions, and descriptions) with an equivalent depth of knowledge. The HSPA scoring criteria did have elements that linked to the KSUS standard on communicating ideas, but it did not do so with an equivalent level of complexity.

Similarly, the analysis of participants' judgments of sample essays indicted that students did not have an adequate opportunity to show an ability to develop ideas. As shown when looking at the five-paragraph essay in-depth, half of the participants thought that an essay with adequate organization but very little development of ideas and concepts was not acceptable at the college level. In addition, all participants thought that the HSPA did not have the ability to align with college-level work that involved synthesizing ideas from texts to support an original argument or opinion. Participants felt that the design of the HSPA should be improved to include ways for students to interact with texts when writing.

Therefore, while the HSPA scoring rubric does address communicating ideas, and while the sample responses sometimes indicated students' communication of ideas at the college level, both the scoring rubric and the interviews with participants suggest that the HSPA does not address this standard with the same level of complexity. Students are not working with text while writing for the HSPA, and this was noted as being a critical aspect of the college writing experience. Also, the participants acknowledged that evidence used for a persuasive essay didn't have to be real, but that they simply needed to acknowledge using any sort of evidence derived in the moment. As with the use of an empty organizational structure embodied in the five-paragraph essay responses, the use of

evidence without real engagement in the meaning-making process shows that the HSPA writing assessment does not align with core elements of college writing expectations.

While the scoring rubric analysis and the analysis based on the sample responses indicated that the HSPA does not address the full complexity of a writer's ability to communicate ideas, the two analyses indicated that the HSPAs focus on categories of organization and structure in writing was satisfactory. The KSUS standard D (using conventions to write clearly) was covered by the HSPA and was matched with a similar level of complexity. Likewise, the interviews with participants showed that sample essays could appropriately show whether students could use organizational elements to produce clarity in their writing. Clarity of writing was noted as a contributing factor when a piece was found to demonstrate college readiness. Community college instructors tended to look at clarity more than ideas, but all instructors thought that some degree of clarity was necessary for a piece to demonstrate college-level expectations.

Another convergence is seen in the way that grammar and mechanics contributed to college readiness as a part of the scoring criteria and the sample responses. In both analyses, the emphasis on grammar and mechanics was focused on lower levels of complexity. Grammar and mechanics were important when looking at lower level essays that did not indicate readiness for college. The mentions of spelling, capitalization, and grammar were minimal on the HSPA scoring rubric, thus the scoring rubric did not indicate that grammar and mechanics were key elements for success on the test. Also, by definition of the depth of knowledge levels, rubric elements and KSUS standards on grammar and mechanics were on the lower end of the depth of knowledge complexity

scale; however, it should be noted that the KSUS presented several spelling and capitalization conditions that were not covered by the HSPA rubric.

Nevertheless, both the scoring rubric and the sample response analyses indicated that grammar and mechanics were important to essays demonstrating college readiness, but that such elements were identified at a minimum threshold of competency. If an essay demonstrated inadequate control it was indicative of a lack of college readiness. Therefore, the HSPA seems to perform sufficiently in providing a consistent signal on grammar and mechanics; however, the rubric might be improved to indicate some types of errors that should not be tolerated in essays.

While there are these convergences, some additional data arose from looking at alignment of the writing assessment with an analysis of student responses. By looking at responses, additional details surfaced that would not have been uncovered had I relied only on the Webb methodology of aligning the scoring criteria. From looking at alignment in terms of student performances, I was able to understand the importance of using evidence, developing ideas, and enacting revision and planning processes. In particular, examining the sample responses shows that the current writing assessment paradigm should be revised to include opportunities to synthesize texts in making arguments. Also, the nature of the disconnect between teaching and assessment was more explicitly described and potential solutions for revising the HSPA writing assessment would not have been identified had I only performed the Webb alignment study on the scoring criteria.

The Webb criteria, however, was not completely irrelevant. The analysis of the scoring criteria demonstrated some general degree of alignment, and the rating of



responses at the lower and upper end of the scale indicated general alignment. Papers in the middle range were harder to classify as showing college readiness and there were elements of the KSUS that were never addressed by the scoring rubric. Therefore, the matter of alignment is not a definitive mark; instead, alignment requires that policymakers, parents, students, teachers, and the public understand more precisely how the test instruments and its sample responses align with expectations. Creating authentic arguments with textual support was seen as an important aspect of college readiness, but the HSPA does not address this skill, nor does the HSPA address revision or the ability to research a topic thoroughly. Even so, while these aspects don't line up, it doesn't mean that the HSPA is completely out of alignment. These results provide a descriptive starting point to further analyze the HSPA in terms of its ability to signal college-level writing abilities.

Most importantly, these results point out the persistent division in the field over the validity of structured responses to generic prompts in showing a student's ability to address more complex issues, such as those that will be confronted in a college atmosphere. The persistence of current-traditional rhetoric (CTR) in writing essays for tests shows that some students may not be prepared for the more complicated rhetoric associated with college-level writing. Students in college will need to use evidence to produce arguments, synthesize multiple texts to produce new ideas and concepts, and revise their work accordingly. The reliance on the five-paragraph essay will typically not suffice in these college situations (Hillocks, 2002a).

Gearing students to only use basic forms and accepting these tight forms that communicate little do not send a consistent signal to students, teachers, and policy

makers about what it means to be successful in both high school and college. The KSUS requires that students understand different audiences and specifically mentions some of the specific writing situations in which college writers should be proficient. The HSPA rubric, while it addresses elements of communication, fails to fully tackle the various situations in which writers need to communicate their ideas. The HSPA rubric calls for structure, and the five-paragraph structure is an accepted structure; however, participants acknowledge that such predictable rhetoric would need to be expanded upon once in college.

Clearly there is a need to refine the theory of writing that underlies the HSPA. To align the HSPA with high school and college teaching, some work with text, expansion of rhetorical moves to communicate an argument, and some opportunities for revision would need to be included in the assessment. This would allow students to expand their writing forms beyond the CTR five-paragraph essay. The allowable genre of test writing needs to be expanded to address real writing situations. From doing the interview-based study, I learned of specific solutions that would create tests that move beyond the constraints of the CTR test genre—namely, giving students texts to research and allowing time for thoughtful planning and revision. The HSPA aligns with aspects of college writing, especially grammar, mechanics, and the basics of organization, but the test could be revised to more accurately signal to students what type of writing will be required at the college level.

## CHAPTER 5. DISCUSSION

The two main questions that guided this study were: “How do the New Jersey High School Proficiency Assessment (HSPA) writing scoring criteria align with college expectations?” and “How do benchmark sample responses from the HSPA align with college expectations?” In order to explore these questions, I recruited eight writing instructors from the high school, community college, and university level and asked them to provide judgments about elements of the HSPA. First, I used the Webb (1999) method for analyzing content alignment of the scoring criteria of the HSPA against the Knowledge and Skills for University Success (KSUS). In order to provide a comparative frame of reference, the study participants also judged how the College Level Examination Program’s (CLEP) writing composition scoring criteria related to the KSUS. Second, I asked instructors to provide judgments on a set of 20 sample HSPA essays. Instructors indicated which essays demonstrated college readiness, and I met with them to discuss their judgments. In analyzing instructor judgments, I used the KSUS as deductive codes and found inductive code patterns in the interview transcripts.

In this chapter, I will discuss and summarize the results of this study. In doing so, I will also describe the implications of this research and how this research relates to the development of new writing assessment paradigms. In the final section of this chapter, I will discuss how I maintained validity as well as the study’s limitations.

### Summary

The first main question that guided this study examined how requirements for success on the HSPA and CLEP assessments related to the KSUS, which are consensus standards for success for first-year college students. By using the Webb (1999) method

for content alignment, several factors related to college-readiness signaled by the HSPA scoring criteria were made explicit. The KSUS contain six general areas (or standards) in which students should be proficient in order to be successful in the first year of college.

1. Successful students apply basic grammar conventions in an effort to write clearly,
2. Successful students know conventions of punctuation and capitalization,
3. Successful students know conventions of spelling,
4. Successful students use writing conventions to write clearly and coherently,
5. Successful students use writing to communicate ideas, concepts, emotions and descriptions to the reader, and
6. Successful students both use and prioritize a variety of strategies to revise and edit written work to achieve maximum improvement in the time available.

Each of these areas of readiness will be discussed herein.

*How Do the New Jersey High School Proficiency Assessment (HSPA) Writing Scoring Criteria Align with College Expectations?*

Similar to previous studies (Brown & Conley, 2007; Milewski et al, 2005), the content alignment analysis indicated that the HSPA writing assessment fails to measure all aspects of writing as stated in the KSUS—especially those noted in the literature as the most important to college success. In particular, the HSPA writing assessment is not designed to measure students’ abilities in making revisions. This is an obvious constraint of the current writing assessment model that relies on a timed first-draft piece of writing as the measure of student performance. Students might revise during the 30- or 60-minute

writing test sessions for narrative or persuasive writing tasks, but whether or not they made meaningful revisions is not captured by the HSPA scoring criteria.

Administering writing assessments as first-draft essays seems acceptable when thinking of the operational constraints of testing in a paper-and-pencil format. However, because more tests are moving toward an online format, further research should focus on the use of multi-session tests or formative assessment systems that record and measure revision practices. Yancey (1999) showed the importance of revision to the writing process, and Acker & Halasek (2008) provided some information on using e-Portfolios to measure student revision practices. Additional work in this area needs to be done so that assessment systems as a whole align better with college success standards and best practices in teaching writing (Hillocks, 2008).

While revision was not measured, other KSUS standards were measured by the HSPA writing rubric. In particular, the HSPA writing rubric fully covered, with equivalent depth, range, and balance, the following KSUS standards.

- Successful students apply basic grammar conventions in an effort to write clearly,
- Successful students know conventions of punctuation and capitalization, and,
- Successful students use writing conventions to write clearly and coherently.

However, the HSPA rubric did not fully measure the depth of knowledge required in the KSUS standard about communicating ideas (Successful students use writing to communicate ideas, concepts, emotions and descriptions to the reader.). When using the Webb (1999) methodology for examining content alignment, instructors rated aspects of the HSPA rubric and the KSUS according to a depth of knowledge hierarchy with “1”

representing the lowest and “4” representing the highest level of complexity. Instructors’ ratings indicated that matching elements of the KSUS were rated at the highest level of complexity, while the same elements of the HSPA were rated at a lower level. For example, the instructors’ ratings indicated that the HSPA scoring criteria did not require that students use a variety of logical arguments to persuade readers, engage the topic with some insights, and use of evidence that anticipates readers’ biases. Such specific elements from the KSUS are difficult to include on a holistic scoring rubric like the one developed for the HSPA. However, it is clear that some of more complex standards stated in the KSUS are not presently measured by the HSPA.

Some additional research needs to be accomplished that would examine how the scoring criteria might be adjusted so that more complex areas of writing can be explicitly measured by the HSPA and other large scale writing assessments. It should be noted that these results indicated that the CLEP scoring rubric did a better job of aligning with the KSUS in terms of representing students’ abilities to communicate ideas. The CLEP rubric included statements that student writing should contain “insightful” ideas that related to the topic. Even general and slight references to the quality of student writing had an impact on participants’ judgments of the depth of the knowledge of the CLEP rubric. Therefore, this study provides an explicit way that the HSPA rubric could be improved with a moderate revision that would increase its alignment with college expectations.

It should be noted that spelling as articulated in the KSUS was also not measured with equivalent depth of knowledge by the HSPA. However, I believe this result is not very important for several reasons. First, several specific factors related to spelling were mentioned in the KSUS (e.g., knowing about common misspellings such as its and it’s),

while the HSPA simply asks that readers examine spelling in general when scoring HSPA responses. Also, as tests move online spelling would be very easy to measure. Therefore, the more important challenge involves designing appropriate writing tasks and scoring criteria that measure student abilities in communicating ideas and including revision practices in future-generation writing assessments.

*How Do Benchmark Sample Responses from the HSPA Align with College Expectations?*

During the second part of this study, the eight instructors provided college-readiness judgments on a set of 20 sample HSPA essays. After they provided their judgments, I met with each participant individually and recorded their thoughts about each essay as well as their overall impressions of the HSPA's ability to indicate college readiness in writing. The interview transcripts provided a rich source of data to further probe elements of the HSPA that may or may not signal college preparedness.

Participants' judgments and their talk about each essay indicated that organizational elements of an essay and the ability of an essay to communicate ideas were the most important factors related to college readiness. In fact, the judgments of the instructors from this study align well with the overall sense of the KSUS, the National Survey of Student Engagement (2008), as well as research and theory in the field (Berlin, 1986, Hillocks 2002a). Participants found that the HSPA samples demonstrated how students could organize an essay; however, participants noted that the HSPA samples failed to accurately represent the proficiency in communicating and supporting ideas at the level that will be demanded of students in college. Indeed, the importance of communicating a strong message and demonstrating critical thinking is necessary; however, HSPA essays were found to indicate these skills only at the highest scoring

levels. Therefore, the HSPA does provide some indication of college readiness in terms of the ability to communicate ideas, but the tasks should be examined so that there are more opportunities for students to use evidence to support an argument.

To highlight the problems in the middle of the scoring scale, I described the case of one particular essay that falls into the classic five-paragraph essay genre. Participants were evenly split about whether this essay indicated college readiness. Those at the community college level or who taught remedial writing in college thought the piece did indicate readiness, while those at the high school and credit-bearing university level thought it failed to indicate college readiness. Those who believed it showed readiness were generally ambivalent, saying that the essay was borderline, but they thought the ability to organize the essay and focus on key points showed the core skills that college students need to be successful.

However, those who thought this essay failed to meet college readiness expectations cited that this essay demonstrated an inability to communicate ideas. Composition and language theorists such as Russell (1997) and Bakhtin (1986) would agree with this later group of participants, and such theory would support an interpretation that a five-paragraph essays fails to understand aspects of the academic genre. Thus, a writing sample that is merely employing a form without working with real evidence to support an argument demonstrates a lack of ability to participate in the socially-constructed academic discourse.

In this regard, the HSPA fails to provide a consistent signal to students and teachers in terms of proficiencies in communicating ideas at the college level. The HSPA writing samples indicated that there are some essays that can use a very standardized



response and still show some relative success on the test. However, those same writing techniques might be seen as a need for remediation at the college level. Nevertheless, it should be noted that HSPA tasks at the upper end of the scoring scale were found to indicate college readiness by the majority of participants (at least 7 of 8). As such, the HSPA does provide a consistent signal of college readiness for students that score at the 5- or 6-level. Unfortunately, despite multiple attempts to collect data from contacts at the New Jersey Department of Education, I could not determine the scoring distribution for students across the state. Nevertheless, some additional research and assessment policy work should be done to determine how to increase the potential use of real evidence as part of the writing assessment.

When looking across teaching contexts, there were some clear patterns in instructors' judgments. Community college instructors tended to focus on organizational elements over the depth or breadth of ideas communicated in a piece of writing. Conversely, high school and university instructors tended to be more lenient in terms of organizational mishaps, but instead made judgments based on ideas presented in a piece. Similar to Smith's (1992, 1993) findings, these results show that teaching context does have a relationship to expectations. In terms of the HSPA's ability to signal college readiness, these results may be interpreted as showing that the HSPA aligns better with the skills and expectations expected at the community college level rather than the university level.

In the final sections of my analyses, I looked more broadly at instructors' perceptions of whether the HSPA sample responses met their expectations for college readiness. Instructors' expressed that the HSPA did provide some information about basic

skills in grammar and organization, but the instructors in this study unanimously indicated that the HSPA failed to accurately measure whether students could synthesize text and use evidence to support an argument. Instructors cited that the design of the single, first-draft writing prompt assessment did not allow for enough interaction with text and the use of evidence. This overall impression from participants was echoed earlier in the results chapter when I described participants' judgments of sample essays. Participants' judgments of college-ready essays relied on the use of evidence. Thus, the assessment design interacts with college readiness judgments and shows a weakness in terms of the ability for the HSPA to indicate college preparedness. The ability to synthesize arguments from across texts has been noted as a critical part of writing at the college level (Berlin, 1986; Hillocks, 2002a), but the HSPA tasks were not viewed as providing that opportunity. In fact, instructors offered suggestions that would alter the assessment format so that critical data on using evidence and synthesizing texts could be added to the writing assessment design.

The results of the interview-based analyses provided some supportive evidence and some additional evidence in comparison to results from the analysis of the HSPA scoring rubric. The HSPA scoring rubric did not contain a measure of revision, and the scoring rubric also did not measure the communication of ideas with an equivalent level of complexity as stated in the KSUS. The interview-based analysis supports these claims. Participants' judgments provided insight into what types of characteristics were missing when considering the communication of ideas. In particular, the interview-based information shows that HSPA essays that used evidence were indicated as demonstrating college readiness. Also, participants' interviews indicated that the assessment format

should be adjusted so that more opportunities for using real evidence and opportunities for revision would be included in the test situation. By adding opportunities to use evidence and to revise, a modified HSPA would support a more consistent signal across the test situation, best practices in teaching at the secondary level, and expectations at the postsecondary level. However, including such elements implies that several practical and theoretical issues need to be considered.

### Implications

The United States Department of Education (2010) recently announced an initiative to improve the quality and college alignment of educational assessments. The results of this study speak directly to the characteristics of the new brand of assessments that are described by the *Race to the Top* initiative.

To fully meet the dual needs for accountability and instructional improvement, however, States need assessment systems that are based on standards designed to prepare students for college and the workplace, and that more validly measure student knowledge and skills against the full range of those standards and across the full performance continuum. Further, States need assessment systems that better reflect good instructional practices and support a culture of continuous improvement in education by providing information that can be used in a timely and meaningful manner to determine school and educator effectiveness, identify teacher and principal professional development and support needs, improve programs, and guide instruction. (p. 5)

This study was designed to investigate how the New Jersey high school assessment in writing aligned with college standards, and it uncovered that the state assessment system in writing is only partially signaling preparedness for college. In addition, the results of this study provide valuable, specific information about how the assessment system in writing can be changed to better reflect good instructional practices. It is appropriate to

acknowledge how this study relates to the *Race to the Top* initiative in more detail as a way to show some potential implications of this study.

First, as stated in the summary above, this study shows how the New Jersey state assessment system in writing signals student preparedness for college. This study provides a careful analysis of the scoring criteria on the writing assessment and compared those criteria against the KSUS. Overall, the HSPA provides a fairly consistent message about good writing, as evidenced by both the scoring rubric's ability to measure most aspects of the KSUS standards. In addition, the analysis of sample responses shows that the highest-level HSPA samples were consistently rated as showing readiness for college. However, it is also clear that some aspects of the assessment need to be revised to increase the complexity with which students are required to demonstrate their ability to effectively communicate ideas. In addition, the interview-based analysis of the instructors' judgments of sample HSPA responses also indicates the importance of revision and the need to increase the opportunities during the test situation when students synthesize text evidence to support an argument.

Of these suggested revisions, some are likely to be more practical than others. For instance, revising the structure of assessment to account for student revision is likely difficult in a state-wide testing program. Having students testing for longer periods of time or more frequently wouldn't be likely to yield greater information about student proficiency at an equivalent rate to what it would cost in terms of dollars and student time away from normal instructional activities. However, increasing the ability for students to interact with text would be more likely within a single administration assessment. The assessment would simply need to include more text for students to analyze and use as

support. The problems of measuring of revision and text synthesis abilities are not easily solved. Perhaps ongoing work in innovative assessment system design, as summarized by the Center for K-12 Assessment and Performance Management (2010), could inform how to use teacher-based formative assessment information on revision and text synthesis abilities as part of a more complete data collection effort.

As there currently is little information on how to use formative and summative assessments to build more complete profiles of student proficiencies, an ambitious research agenda should be undertaken to investigate how to revise the data collection on student writing associated with the HSPA. In particular, some thinking should be done about possible ways to collect information on revision that would be efficient and fair. In addition, some thinking should be done around the possibility of using current technologies to allow students to take a writing assessment over multiple sessions or allowing students to use multiple texts, which would provide an opportunity for the use of evidence to support the communication of ideas and arguments.

#### Considerations for Future Research

These results show that the HSPA administrators should consider a research agenda on how to practically implement new forms of writing assessments; however, some additional aspects should be considered about how these results relate to previous research. Hillocks (2008) provided a review of research on the impact of written comments on the quality of student writing, and the results are somewhat mixed. Hillocks (2008) reported that a study by Gee (1972) and Hillocks (1986) provided some indication that teacher comments show little or no statistical evidence of increasing the quality of student writing. These findings point out that, while it may be difficult to teach and even

harder to measure, revision is viewed as an important part of the writing process. Hillocks (2008) reports that other studies (Hillocks 1982; Sperling & Freedman, 1987) have shown that feedback that is consistent with other educational supports, such as in-class assignments on revision strategies, did have a statistically significant positive impact on student writing.

The research studies about the impact of revision on the quality of student writing are important to consider when undertaking a research agenda about increasing opportunities for revision in large-scale writing assessments. Any research agenda that looks to increase revision as part of a writing assessment needs to first perform a careful literature review to determine what might be appropriate goals of a revision measure. It may be that including revision might not provide a great deal of new information about student abilities to write overall (and may correlate highly with student performance if only given a single draft opportunity). However, policymakers and educators should nevertheless consider adding revision opportunities to the writing assessment because of its potential link to representing real classroom practices and student learning.

When considering adding opportunities for synthesizing texts to writing assessments, considerations of introducing construct-irrelevant tasks into the writing assessment should be studied (Messick, 1986). Based on these results, it would appear as though the construct of writing as tested in the assessment does not align with a critical aspect of the domain of interest. Namely, in writing composition courses in college, students need to demonstrate proficiency with synthesizing a reading while creating written arguments. Therefore, the construct of the writing assessment should be defined with both reading and writing in mind, and reading is therefore not a construct-irrelevant

skill. Nevertheless, a thorough research agenda should be developed that seeks to understand other boundaries of construct-irrelevant skills that may or may not interfere with the ability to measure writing (and reading) expected at the college level.

The implications of this research on the HSPA scoring criteria and sample responses are not entirely novel. Previous research (Murphy & Yancey, 2008) has shown that there are multiple studies that question the validity of timed, first-draft writing assessments. In particular, studies by Langer (1984), Sullivan (1997), and Murphy, Carroll, Kinzer, and Robyns (1982) have shown that knowledge of the subject, cultural background knowledge, and task interpretations have a varying impact on student performance. As cited throughout this dissertation, Hillocks (2002a) performed a multi-state study of writing assessments and found that they typically did not encourage students to engage in a critical form of rhetoric. This research study on the alignment of the HSPA writing assessment with college expectations corroborates with these previous studies and calls into question the validity of using the impromptu writing sample as the measure of writing proficiency. If new goals for writing assessments are to be defined as the *Race to the Top* initiative promotes, then researchers and policymakers can use the results of this study to direct their efforts to study how the communication of ideas can become an integral part of the writing assessment paradigm.

The implications of this research are not limited to the current conceptions and applications of writing assessment. This research should also be considered when thinking about how to design assessments intended to measure skills and knowledge associated with students competing in the 21st century global and digital environment. The Partnership for 21st Century Skills (2009) advocates a vision of education where the

traditional focus on reading, writing, and arithmetic are merged with the four core skills: critical thinking, communication, collaboration, and creativity. This study demonstrated a weakness with the current high school writing test paradigm in its incapacity to accurately measure students' abilities to communicate ideas and use critical thinking to support an argument in a first-draft piece of writing. Further research should be done that would seek to define foundational aspects of new assessment paradigms that might better align with and foster proficiency in 21st century skills.

New Jersey has already recognized the importance of encouraging teaching and learning with 21st century skills in mind. Recently, the New Jersey DOE created an office to coordinate reading and mathematics initiatives (Partnership for 21st Century Skills, no date). A similar coordinated effort should be undertaken for the language arts. In particular, New Jersey might consider how new assessment formats, such as those being explored by the National Assessment of Educational Progress (Bennett, Persky, Weiss, & Jenkins, 2007), might be used to situate writing assessment within problem solving situations. As noted in this study, participants thought that students needed a greater opportunity to engage with their writing subject matter through research and revision. A problem-solving, technology-rich writing assessment should be explored to better understand how high school students are prepared to perform in the global and digital environment.

### Validity and Limitations

The validity of this study is supported in several ways. The results that were described in chapter 4 (and summarized above) did in fact converge around a common set of results. As mentioned above, the HSPA was noted to show a weakness in sending a



consistent signal regarding the complexity of communicating ideas as part of the writing scoring rubric. A similar finding was indicated in the analysis of interview transcripts when all participants mentioned that the HSPA should be improved by requiring students to produce writing that synthesizes multiple texts. By looking for themes across the two data sets involved in this study, conclusions about the HSPA's ability to signal college readiness are strengthened.

In addition to the search for corroborating evidence, the validity of this study is supported by the participants who provided the data. Each participant was an instructor with a great deal of experience in both the classroom as well as with assessment scoring procedures. Acquiring data from these participants provides a rich source of data that is informed by years of practice. In addition, participants generally agreed about some aspects of both the scoring rubric coding and the sample essay ratings. The raters achieved a .95 intraclass correlation on their depth of knowledge ratings, and the agreement rate on the sample responses was 88%, which ranged from 50 to 100 % agreement on various essays. While the study's design attempted to examine perceived differences in opinion on whether essays show college readiness, a relatively high degree of interrater agreement shows that raters' judgments converged around consistent patterns.

Validity was also maintained through checking with participants about the accuracy of transcripts of interviews and the initial results created from the transcripts. I emailed transcripts to each participant and asked for their feedback. In addition, I emailed them an overview of my results so that I could hear their feedback. Unfortunately, likely due to the busy nature of teachers' schedules, I did not hear a response from the

participants. As Coffey and Atkinson (1996) noted, this is often the case with member checking. However, during the interview process, I did record several very positive remarks from participants. In general, participants thought that a study like this one, which examines the nature of writing done on the HSPA to the expectations of college-level writing, provided a good way to begin a dialogue about how the assessments relate to learning. Participants voiced that they would like a forum in which to interact on a professional level with the HSPA requirements so that the test could better speak to the richness of their curricula, which is similar to results by Brown (2004). Lastly, I support the validity of this study through a process of searching for negative evidence in the transcripts. By looking at both convergence and divergence in the transcripts, I was able to provide a complete picture of the interviewees' perspectives. This process of providing a complete picture of interviewees' perspectives is evidence in my use of their words to describe important characteristics of essays and their divergent judgments about the case of the five-paragraph essay presented in chapter 4.

The use of particular methods also contributes to the validity of this study. By using the Webb (1999) alignment methodology, this study uses a generally agreed upon method for analyzing alignment between a test and a set of standards. Also, by using interviewing, coding, and qualitative data analysis techniques (Marshall & Rossman, 1999; Patton, 1990), this study allows for judgments about alignment of the HSPA to college expectations to maintain a focus on in-depth information and details of writing performance that inform the central aspects of expectations for college writers.

Nevertheless, there are some limitations. For instance, the sample of teachers was relatively small and cannot be generalized to the population in New Jersey or to

commonalities across teaching contexts. In addition, because this study only asked for the judgments of instructors about aspects of college readiness, it did not take into account potential data from classroom observations or surveys about classroom practices, as the Porter (2006) method for surveying the enacted curriculum may have done. While the comparison to classroom practice could provide an avenue for further research, the current study does suggest that there is some degree of alignment between the HSPA and college expectations, both in terms of the requirements for success on the test as represented in the scoring rubric and on the sample essay responses. However, this study also shows that there are specific weaknesses with the current writing paradigm used on the HSPA.

## REFERENCES

- Achieve, Inc. (2007). *Aligned expectations? A closer look at college admissions and placement tests*. Retrieved October 4, 2008 from <http://www.achieve.org/AlignedExpectations>.
- Acker, S. & Halasek, K. (2008). Preparing high school students for college-level writing: Using ePortfolio to support a successful transition. *The Journal of General Education*, 57 (1), 1-14.
- ACT (2007). ACT National Curriculum Survey 2005-2006. Iowa City, IA: ACT.
- Agee, J. (2004). Negotiating a teaching identity: An African-American teacher's struggle to teach in test-drive contexts. *Teachers College Record*, 106, 747-774.
- Anagnostopoulos, D. (2003). Testing and student engagement with literature in urban classrooms: A multi-layered perspective. *Research in the Teaching of English*, 38, 136-176.
- Anastasi, A. (1986). *Psychological testing* (5th ed.). New York: Macmillan.
- Assaf, L. (2005). One reading specialist's response to high-stakes testing pressures. *The Reading Teacher*, 60 (2), 158-167.
- Bakhtin, M. M. (1986). The problem of speech genres. In C. Emerson & M. Holquist (Eds.) *Speech genres and other late essays* (pp. 60-103), (V.W. McGee, Trans.) Austin, TX: University of Texas Press.
- Berlin, J. A. (1988). Rhetoric and ideology in the writing class. *College English*, 50, 477-494.
- Berlin, J. A. (1996). *Rhetorics, poetics, and culture*. West Lafayette, IN: Parlor Press.
- Blank, R.K., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science. Results from Survey of Enacted Curriculum Project*. (Final Report to the National Science Foundation on Contract No. REC98-03080). Washington, DC: Council of Chief State School Officers.
- Borko, H. (1997). New forms of classroom assessment: Implications for staff development. *Theory into Practice*, 36, 231-238.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles, and Practice*, 11, 305-322.

- Brown, R. S. & Conley, D. T. (2007). Comparing state high school assessments for success in entry-level university courses. *Educational Assessment*, 12 (2), 137-160.
- Callan, P. M., Finney, J. E., Kirst, M. W., Usdan, M. D., & Venezia, A. (2006). *Claiming common ground: State policymaking for improving college readiness and success* (National Center report No. 06-1). San Jose, CA: National Center for Public Policy and Higher Education.
- Callahan, S. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing*, 6 (1), 5-40.
- Camp, R. (1993). Changing the model for the direct assessment of writing. In B. Hout, and M. M. Williamson. *Validating Holistic Scoring for Writing Assessment* (pp. 45-78) Cresskill, NJ: Hampton.
- Center for K-12 Assessment and Performance Management (2010). *Next generation assessment systems proposed under the race to the top program*. Retrieved October 23, 2010 from <http://www.k12center.org/rsc/pdf/15051-K12Cntr-RTTTbro-Digital.pdf>.
- Coffey, A. & Atkinson, P. (1996). *Making sense of qualitative data: Complementary strategies*. Thousand Oaks, CA: Sage.
- College Board. (2009). *CLEP official study guide 2009: English composition*. New York: The College Board.
- Conley, D. T. (2003). *Mixed messages: What state high school tests communicate about student readiness for college*. Eugene, OR: Center for Education Policy Research.
- Conley, D. T., Aspengren, K., & Stout, O. (2006). *Advanced placement best practices study: Biology, chemistry, environmental science, physics, European history, U.S. history, world history*. Eugene, OR. Educational Policy Improvement Center.
- Creswell (1998). *Qualitative inquiry and research design; Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Dekker, T. and Feijs, E. (2005). Scaling up strategies for change: Change in formative assessment practices. *Assessment in Educations: Principles, Policy, and Practice*, 12, 237-254.
- Elbow, P. (1996). Writing assessment in the 21st Century: A utopian view. I. L.Z. Bloom, D. A. Daiker & E. M. White. (Eds). *Composition in the Twenty-First Century: Crisis and Change*. (pp. 83- 100). Carbondale, IL: Southern Illinois University Press.

- Finnegan, R. J. (2005). "It just doesn't reflect what we are doing:" A case study of navigating pedagogical beliefs in the context of state assessments. Division G, American Educational Research Association.
- Firestone, W.A., & Mayrowetz, D. (2000). Rethinking "high stakes:" Lessons from the U.S. and England and Wales. *Teachers College Record*, 102, 724-749.
- Firestone, W. A. Monfils, L., Camilli, G., Schorr, R. Y., Hicks, J. E. & Mayrowetz, D. (2002). The ambiguity of test preparation: A multimethod analysis in one state. *Teacher College Record*, 104, 1485-1523.
- Firestone, W. A., Schorr, R. Y., & Monfils, L. F. (2004). *The ambiguity of teaching to the test*. Mahwah, NJ: Erlbaum.
- Flores, B. B. & Clark, E. R. (2003). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education* [On-line], 6 (3). Retrieved January 12, 2007 from <http://cie.ed.asu.edu/volume6/number3/>.
- Flower, L. & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*. 32, 365-387.
- Gambell, T. & Hunter, D. (2004). Teacher scoring of large-scale assessment: Professional development or debilitation? *Journal of Curriculum Studies*, 36, 697-724.
- Gee, T. C. (1972). Students' response to teacher comments. *Research in the Teaching of English*, 6, 212-221.
- Gitomer, D. H. & Duschl, R. A. (2007). Establishing multilevel coherence in assessment. In P. Moss (Ed.), *Evidence and decision making* (pp. 288-320). Malden, MA: Blackwell.
- Harwayne, S. & Calkins, L.M. (1991). *Living between the lines*. Portsmouth, NH: Heinemann.
- Heath, S.B. (1983). *Ways with words*. New York: Cambridge University Press.
- Higgins, B., Miller, M. & Wegmann, S. (2006). Teaching to the test...not! Balancing best practice and testing requirements in writing. *Reading Teacher*, 60, 310-319.
- Hillocks, G. (1982). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English*, 16, 261-278.

- Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: National Conference on Research in English/ERIC Clearinghouse on Reading and Communications Skills.
- Hillocks, G. (2002a). *The testing trap: How states writing assessments control learning*. New York: Teachers College Press.
- Hillocks, G. (2002b). Rhetoric in classrooms: Prospects for the twenty-first century. In Williams, J.D. (ED). *Visions and Revisions: Continuity and Change in Rhetoric and Composition*. (pp 219-248). Carbondale, IL: Southern Illinois University Press.
- Hillocks, G. (2008). Writing in secondary schools. In C. Bazerman (ED) *Handbook of Research on Writing*. (pp. 311-329). Mahwah, NJ: Erlbaum.
- Hoffman, J. V., Assaf, L. C., and Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *Reading Teacher*, 54, 482-493.
- Hoffman, N., Vargas, J., Venezia, A., & Miller, M. S. (2007). *Minding the gap: Why integrating high school with college makes sense and how to do it*. Cambridge, MA: Harvard Education Press.
- Hout, B. (2002a). *(Re)Articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.
- Hout, B. (2002b). Toward a new discourse of assessment for the college writing classroom. *College English*, 65 (2), 163-180.
- Johnson, T. S., Smagorinski, P., Thompson, L., & Fry, P. G. (2003). Learning to teach the five-paragraph theme. *Research in the Teaching of English*, 38, 136-176.
- Kirst, M. W., & Braco, K. R. (2004). Bridging the great divide. In M.W. Kirst & A. Venezia (Eds.) *From high school to college* (pp. 1 – 30). San Francisco, CA: Jossey-Bass.
- Kirst, M. W., & Usdan, M. D. (2007). The history of the separation of K-12 and postsecondary education. In N. Hoffman, J. Vargas, A. Venezia, & M.S. Miller (Eds.) *Minding the gap* (pp. 55-64). Cambridge, MA: Harvard Education Press.
- Langer, J. (1984). The effects of available information on response to school writing tasks. *Research in the Teaching of English*, 18 (1), 27-44.
- Le, V. (2002). *Alignment among secondary and post-secondary assessments in five case study states*. Santa Monica, CA: RAND. (Monograph report MR-1530.0-EDU).
- Lincoln, Y. & Guba, E. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage.

- Marshall, C. & Rossman, G. B. (1999). *Designing qualitative research*. Thousand Oaks, CA: Sage.
- Milewski, G. B., Johnsen, D., Glazer, N., & Kubota, M. (2005). *A survey to evaluate the alignment of the new SAT writing and critical reading sections to curricula and instructional practice*. (College Board Research Report No. 2005-1). New York: College Board.
- Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 24, 367-378.
- Messick, S. (1989). Validity. In R. Linn (Ed.). *Educational measurement*. (pp. 13-103) New York: Macmillan
- Mislevy, R. (2007). Validity by design. *Educational Researcher*, 36, 463-469.
- Moss, P. (2007). *Evidence and decision making*. Malden, MA: Blackwell.
- Murphy, S., Carroll, K., Kinzer, C., & Robyns, A. (1982). A study of the construction of the meaning of a writing prompt by its authors, the student writers, and the raters. In J.R. Gray & L.P. Ruth (Eds.), *Properties of writing tasks: A study of alternative procedures for holistic writing assessment* (pp. 386-468). Berkeley, CA: University of California, Graduate School of Education, Bay Area Writing Project. (ERIC Document Reproduction Service No. ED230576).
- Murphy, S., & Yancey, K.B. (2008). Construct and consequence: Validity in writing assessment. In C. Bazerman (ED) *Handbook of Research on Writing*. (pp. 365-385). Mahwah, NJ: Erlbaum.
- National Assessment Governing Board (2009). *Making New Links, 12th Grade and Beyond: Technical Panel on 12th Grade Preparedness Research Final Report*. Washington, DC: National Assessment Governing Board.
- National Survey of Student Engagement. (2008). *Promoting engagement for all students: The imperative to look within*. Bloomington, IN: Indiana University, Center for Postsecondary Research.
- New Jersey Department of Education. (2008a). *Criterion-based holistic scoring: A writing handbook*. Trenton, NJ: NJ DOE.
- New Jersey Department of Education. (2008b). High School Proficiency Assessment Spring 2008: Executive Summary. Retrieved October 21, 2009 from <http://www.state.nj.us/education/schools/achievement/2009/hspa/>.



- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.
- Perin, D & Charron, K. (2003). Trends in community college assessment and placement approaches: Implications for educational policy. Division J, American Educational Research Association.
- Pomplun, M. (1997). State assessment and instructional change: A path model analysis. *Applied Measurement in Education*, 10, 217-234.
- Porter, A. C. (2006). Curriculum assessment. In J.L. Green, G. Camili, & P.B. Elmore. (Eds). *Complementary Methods in Education Research*. Mahwah, NJ: Erlbaum.
- Porter, A. C. (2002) Measuring the content of instruction: Uses in research and practice. *Educational Researcher* 31(7), 3-14.
- Porter, A. C., Kirst, M. W., Osthoff, J., Smithson, J. & Schneider, S. (1993). *Reform up close: A classroom analysis*. (Final report to the National Science Foundation, Grant No. SPA-8953446.) Washington, DC: National Science Foundation.
- Porter, A. C.. & Smithson, J. (2001). *Defining, developing, and using curriculum indicators*. Research Report Series RR-048. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2003). Benchmarking and alignment of standards and testing. *Educational Assessment*, 9, (1&2), 1-27.
- Resnick, L., Besterfield-Sacre, M., Mehalik, M., Sherer, J. Z., Halverson, E. (2007). A framework for effective management of school system performance. In P. Moss (Ed.), *Evidence and decision making* (pp. 155-185). Malden, MA: Blackwell.
- Resnick, L. B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform in changing assessments. In B.K. Gifford & M.C. O'Connor (Eds.) *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: National Commission on Testing and Public Policy.
- Rex, L. & Nelson, M. (2004). How teachers' professional identities position high-stakes test preparation in their classrooms. *Teachers College Record*, 106, 1288-1331.
- Roach, A. T., Niebling, B.C., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessments: Implications and applications for research and practice. *Psychology in the Schools*, 45(2), 158-176.

- Roth, J., Crans, G. G., Carter, R. L., Ariet, M., & Resnick, M. B. (2001). Effect of high school course-taking and grades on passing a college placement test. *The High School Journal*, 84 (2), 72-87.
- Russell, D. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication*, 14, 504-554.
- Salahu-Din, D., Persky, H., and Miller, J. (2008). *The nation's report card: Writing 2007* (NCES 2008-468). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- Schulte, R.W., Thompson, J., Natis, Y.V., Pezzini, M., Sinur, J., Kenney, L.F., et al. (2005). *Magic quadrant for integration backbone software*. (Research Rep. No. G00127186). Stamford, CT: Gartner, Inc.
- Silva, E. (2008). *Measuring skills for the 21st Century*. Washington, DC: Education Sector Reports.
- Smith, W. L. (1992). The importance of teacher knowledge in college composition placement testing. In J.R. Hayes, R.E. Young, M.L. Matchett, M. McCaffrey, C. Cochran, T. Hajduk. (Eds.) *Reading empirical research studies: The rhetoric of research* (pp. 289-316). Hillsdale, NJ: Erlbaum
- Smith, W. L. (1993). Assessing the reliability and adequacy of using holistic scoring of essays as a college composition placement technique. In M.M. Williamson & B.A. Hout (Eds.) *Validating holistic scoring for writing assessment: theoretical and empirical foundations* (pp. 142-205). Cresskill, NJ: Hampton Press.
- Sperling, M. & Freedman, S. W. (1987). A good girl writes like a good girl: Written response to student writing. *Written Communication*, 9, 343-369.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stone, C. A. and Lane, S. (2003). Consequences of a state accountability program: Examining relationships between school performance gains and teacher, student and school variables. *Applied Measurement in Education*, 16 (1), 1-26.
- Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge, MA: Cambridge University Press.
- Strong American Schools. (2008). *Diploma to nowhere*. Washington, DC: Author.
- Sullivan, F. (1997). Calling writer's bluffs: The social production of writing ability in university placement testing. *Assessing Writing*, 4(1), 53-82.

- Sunstein, B. (2000). Be Reflective, Be Reflexive, Beware: Innocent Forgery for Inauthentic Assessment. In B. Sunstein & J. Lovel (Eds.) *The portfolio standard: How students can show us what they know and are able to do.* (pp. 3-14) Portsmouth, NH: Heinemann.
- United States Department of Education (2010). Race to the top assessment program: Application for new grants: Comprehensive assessment systems. CFDA number 84.395B. Washington, DC: United State Department of Education.
- Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessment in mathematics and science education* (Research Monograph No. 6). Madison, WI: National Institute for Science Education.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Madison, WI: National Institute for Science Education.
- Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.
- Webb, N. ., Herman, J. L., & Webb, N. L. (2007). Alignment of mathematics state-level standards and assessments: The role of reviewer agreement. *Educational Measurement: Issues and Practice*, 26 (2), 17-29.
- Wiggins, G. & McTighe, J. (2005). *Understanding by design.* Alexandria, VA: Association for Supervision and Curriculum Development.
- Williams, W.C. (1985). *Selected poems.* New York: New Directions.
- Wixson, K. K., Fisk, M. C., Dutro, E., & McDaniel, J. (2002). *The alignment of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.
- Yancey, K. B. (1999) Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, 50, 483-503.

## APPENDIX A. PARTICIPANT INSTRUCTIONS

### Instructions for Alignment Study Participants

Thank you for participating in this alignment study of the HSPA writing assessment. This study will occur in two stages. First, you will perform some analyses on your own, and then we will meet to discuss some of the judgments you made. When performing analyses on your own, some specific processes will be used to make judgments about the HSPA and CLEP scoring rubrics and how they address the Knowledge and Skills for University Success (KSUS). The KSUS are consensus standards for university success that have been adopted by many universities and are used here to provide a common language for gauging how standards are embodied by the assessment. You will also read and provide a judgment about 20 sample HSPA essays. Below is an explanation of the analyses you will perform on your own.

### Alignment Analysis of Scoring Rubrics

In this packet you will find the scoring rubrics used for the HSPA writing assessment and CLEP English composition test. Please read both of these scoring rubrics to gain general familiarity with them. After you have read the rubric, go to the alignment coding sheets that follow.

Your first task will be to assign each of the KSUS standards with a Depth of Knowledge (DOK) rating in the shaded cell next to the standard/objective description. The DOK rating provides a way to examine the level of skill required by each standard. We will come to consensus (via a phone or in-person meeting) on how to apply the DOK levels prior proceeding with the coding process.

After you provide a DOK rating for each of the KSUS standards and objectives, you will provide a DOK rating for each element of the HSPA scoring rubric. You will also provide a DOK rating for each element of the CLEP scoring rubric.

After you have provided DOK ratings for the standards and the rubrics, you will then indicate where each element of HSPA scoring rubric addresses a KSUS standard or objective by listing the KSUS standard in the primary or secondary objective columns on the HSPA or CLEP coding sheets. Each element of the scoring rubric may address more than one KSUS standard. Mark all that apply. If the element of the rubric does not address a standard, skip that column and do not mark anything on the coding sheet. You will repeat this process of indicating “hits” for CLEP rubric as well.

### Judgment of Sample HSPA Responses

In the packet there are 20 sample HSPA responses. All of these are actual responses from the 2008 administration of the HSPA. For each essay, you are going to make a dichotomous judgment. Imagine that you are using this essay to make a judgment about college readiness. Indicate on the cover sheet whether the essay provides sufficient evidence that the student should or should not enter credit-bearing college-level writing courses. You can write any notes about the essay in the space provided, and doing so may help when we discuss your judgments in our follow-up interview.

## APPENDIX B. DEPTH OF KNOWLEDGE LEVELS FOR WRITING

Norman L. Webb

March 28, 2002

### ***Level 1***

Level 1 requires the student to write or recite simple facts. This writing or recitation does not include complex synthesis or analysis but basic ideas. The students are engaged in listing ideas or words as in a brainstorming activity prior to written composition, are engaged in a simple spelling or vocabulary assessment or are asked to write simple sentences. Students are expected to write and speak using Standard English conventions. This includes using appropriate grammar, punctuation, capitalization and spelling. Some examples that represent but do not constitute all of Level 1 performance are:

- Use punctuation marks correctly.
- Identify Standard English grammatical structures and refer to resources for correction.

### ***Level 2***

Level 2 requires some mental processing. At this level students are engaged in first draft writing or brief extemporaneous speaking for a limited number of purposes and audiences. Students are beginning to connect ideas using a simple organizational structure. For example, students may be engaged in note-taking, outlining or simple summaries. Text may be limited to one paragraph. Students demonstrate a basic understanding and appropriate use of such reference materials as a dictionary, thesaurus, or web site. Some examples that represent but do not constitute all of Level 2 performance are:

- Construct compound sentences.
- Use simple organizational strategies to structure written work.
- Write summaries that contain the main idea of the reading selection and pertinent details.

### ***Level 3***

Level 3 requires some higher level mental processing. Students are engaged in developing compositions that include multiple paragraphs. These compositions may include complex sentence structure and may demonstrate some synthesis and analysis. Students show awareness of their audience and purpose through focus, organization and the use of appropriate compositional elements. The use of appropriate compositional elements includes such things as addressing chronological order in a narrative or including supporting facts and details in an informational report. At this stage students are engaged in editing and revising to improve the quality of the composition. Some examples that represent but do not constitute all of Level 3 performance are:

- Support ideas with details and examples.
- Use voice appropriate to the purpose and audience.
- Edit writing to produce a logical progression of ideas.

### ***Level 4***

Higher-level thinking is central to Level 4. The standard at this level is a multi- paragraph composition that demonstrates synthesis and analysis of complex ideas or themes. There is evidence of a deep awareness of purpose and audience. For example, informational papers include hypotheses and supporting evidence. Students are expected to create compositions that demonstrate a distinct voice and that stimulate the reader or listener to consider new perspectives on the addressed ideas and themes. An example that represents but does not constitute all of Level 4 performance is:

- Write an analysis of two selections, identifying the common theme and generating a purpose that is appropriate for both.

## APPENDIX C. KSUS OBJECTIVES RATING SHEET

KSUS Objectives	DOK Rating
A. Successful students apply basic grammar conventions in an effort to write clearly.	
A.1. identify and use correctly and consistently parts of speech, including nouns, pronouns, verbs, adverbs, conjunctions, prepositions, adjectives and interjections.	
A.2. use subject-verb agreement and verb tense consistently and correctly.	
A.3. demonstrate consistent, correct and appropriate pronoun agreement and the use of different types of clauses and phrases, including adverb clauses, adjective clauses and adverb phrases.	
B. Successful students know conventions of punctuation and capitalization. They:	
B.1. use commas with nonrestrictive clauses and contrasting expressions.	
B.2. use ellipses, colons, hyphens, semi-colons, apostrophes and quotation marks correctly.	
B.3. capitalize sentences and proper nouns correctly.	
B.4. consistently avoid run-on sentences and sentence fragments.	
C. Successful students know conventions of spelling. They:	
C.1. use a dictionary and other resources to spell new, unfamiliar or difficult words.	
C.2. differentiate between commonly confused terms, such as “its” and “it’s” or “affect” and “effect.”	
C.3. know how to use the spellchecker and grammar check function in word processing software while understanding the limitations of relying upon these tools.	
D. Successful students use writing conventions to write clearly and coherently. They:	
D.1. know and use several prewriting strategies, including developing a focus; determining the purpose; planning a sequence of ideas; using structured overviews; and creating outlines.	
D.2. use paragraph structure in writing as manifested by the ability to construct coherent paragraphs and arrange paragraphs in logical order.	
D.3. use a variety of sentence structures appropriately in writing, including compound, complex, compound-complex, parallel, repetitive and analogous sentence structures.	

D.4. present ideas to achieve overall coherence and logical flow in writing and use appropriate techniques such as transitions and repetition to maximize cohesion.	
D.5. use words correctly; use words that mean what the writer intends to say; and use a varied vocabulary.	
D.6. demonstrate development of a controlled yet unique style and voice in writing where appropriate.	
D.7. use a style manual, such as the Modern Language Association (MLA) or the American Psychological Association (APA) to apply writing conventions and to create documentation formats in a manner consistent with the manual.	
E. Successful students use writing to communicate ideas, concepts, emotions and descriptions to the reader. They:	
E.1. know the difference between a topic and a thesis.	
E.2. articulate a position through a thesis statement and advance it using evidence, examples and counterarguments that are relevant to the audience or issue at hand.	
E.3. use a variety of methods to develop arguments, including compare-contrast reasoning; logical arguments (inductive/deductive); and alternation between general and specific (e.g., connections between public knowledge and personal observation and experience).	
E.4. write to persuade the reader by anticipating and addressing counterarguments, by using rhetorical devices and by developing an accurate and expressive style of communication that moves beyond mechanics to add flair and elegance to writing.	
E.5. use a variety of strategies to adapt writing for different audiences and purposes, such as including appropriate content and using appropriate language, style, tone and structure.	
E.6. distinguish between formal and informal styles, for example, between academic essays and personal memos.	
E.7. use appropriate strategies and formats to write personal and business correspondence, including appropriate organizational patterns, formal language and tone.	
E.8. use appropriate strategies to write expository essays that employ supporting evidence; use information from primary and secondary sources; incorporate charts, graphs, tables and illustrations where appropriate; anticipate and address readers' biases and expectations; and explain technical terms and notations.	
E.9. use strategies to write fictional, autobiographical, and biographical narratives that include a well-developed point of view and literary elements; present events in logical sequence; convey a unifying theme or tone; use concrete and sensory language and pace action.	
F. Successful students both use and prioritize a variety of strategies to revise and edit written work to achieve maximum improvement in the time available. They:	
F.1. employ basic editing skills proficiently to identify obvious mechanical errors, clarify and improve the structure of the piece and sharpen language and meaning.	

F.2. review ideas and structure in substantive ways to improve depth of information and logic of organization.	
F.3. reassess appropriateness of writing in light of genre, purpose and audience.	
F.4. use feedback from others to revise written work.	



# APPENDIX D. HSPA SCORING RUBRIC RATING SHEET

Put as many objective references as apply in primary or secondary cells

HSPA Rubric Elements	DOK (1-4)	Primary Obj	Secondary Obj	Secondary Obj	Notes
A.1. Communicates intended message to intended audience.					
A.2. Relates to topic with a single distinct focus. At the highest level, unified and coherent. Well-developed.					
A.3. Logical progression of ideas.					
A.4. Transitions. At the highest level, fluent and cohesive					
A.5 Appropriate details and information: at the highest level, details are effective, vivid, explicit, and/or pertinent					
B.1. Tense formation					
B.2 Subject-verb agreement					
B.3. Pronouns usage/agreement					
B.4. Word choice/meaning					
B.5. Proper modifiers					

C.1. Variety of type, structure, and length of sentences. At the highest level, precision and/or sophistication.					
C.2. Correct sentence construction. At the highest level, very few, if any, errors.					
D.1. Spelling					
D.2. Capitalization					
D.3. Punctuation					

# APPENDIX E. CLEP SCORING RUBRIC RATING SHEET

Put as many objective references as apply in primary or secondary cells

CLEP Rubric Elements	DOK (1-4)	Primary Obj	Secondary Obj	Secondary Obj	Notes
A 6 essay demonstrates a high degree of competence and sustained control although it may have a few minor errors					
addresses all elements of the writing task effectively and insightfully					
develops ideas thoroughly, supporting them with well-chosen reasons, examples, or details					
is well focused and well organized					
demonstrates superior facility with language using effective vocabulary and sentence variety					
demonstrates general mastery of the standard conventions of grammar, usage, and mechanics but may have minor errors					

## APPENDIX F. HSPA SCORING RUBRIC

***New Jersey Registered Holistic Scoring Rubric***

Used for the "Writing to Speculate" (Picture Prompt) and Persuasive Writing Items

	Inadequate Command	Limited Command	Partial Command	Adequate Command	Strong Command	Superior Command
Score:	1	2	3	4	5	6
<u>Content and Organization</u>	May lack opening and/ or closing	May lack opening and/ or closing	May lack opening and/ or closing	Generally has opening and/or closing	Opening and closing	Opening and closing
	Minimal response to topic; uncertain focus	Attempts to focus May drift or shift focus	Usually has single focus	Single focus	Single focus Sense of unity and coherence Key ideas developed	Single, distinct focus Unified and coherent Well- developed
	No planning evident; disorganized	Attempts organization Few, if any, transitions between ideas	Some lapses or flaws in organization May lack some transitions between ideas	Ideas loosely connected Transitions evident	Logical progression of ideas Moderately fluent Attempts compositional risks	Logical progression of ideas Fluent, cohesive Compositional risks successful
	Details random, inappropriate, or barely apparent	Details lack elaboration, i. e., highlight paper	Repetitious details Several unelaborated details	Uneven development of details	Details appropriate and varied	Details effective, vivid, explicit, and/ or pertinent
<u>Usage</u>	No apparent control Severe/ numerous errors	Numerous errors	Errors/ patterns of errors may be evident	Some errors that do not interfere with meaning	Few errors	Very few, if any, errors
<u>Sentence Construction</u>	Assortment of incomplete and/ or incorrect sentences	Excessive monotony/ same structure Numerous errors	Little variety in syntax Some errors	Some variety Generally correct	Variety in syntax appropriate and effective Few errors	Precision and/or sophistication Very few, if any, errors
<u>Mechanics</u>	Errors so severe they detract from meaning	Numerous serious errors	Patterns of errors evident	No consistent pattern of errors Some errors that do not interfere with meaning	Few errors	Very few, if any, errors
Non-Scorable Responses	<b>NR</b>	No Response	Student wrote too little to allow a reliable judgment of his/her writing.			
	<b>OT</b>	Off Topic/Off Task	Student did not write on the assigned topic/ task, or the student attempted to copy the prompt.			
	<b>NE</b>	Not English	Student wrote in a language other than English.			
	<b>WF</b>	Wrong Format	Student refused to write on the topic, or the writing task folder was blank.			
<b>Content/ Organization</b>		<b>Usage</b>		<b>Sentence</b>		<b>Mechanics</b>

		Construction	
<ul style="list-style-type: none"> <li>Communicates intended message to intended audience</li> <li>Relates to topic</li> <li>Opening and closing</li> <li>Focused</li> <li>Logical progression of ideas</li> <li>Transitions</li> <li>Appropriate details and information</li> </ul>	<ul style="list-style-type: none"> <li>Tense formation</li> <li>Subject- verb agreement</li> <li>Pronouns usage/ agreement</li> <li>Word choice/ meaning</li> <li>Proper Modifiers</li> </ul>	<ul style="list-style-type: none"> <li>Variety of type, structure, and length</li> <li>Correct construction</li> </ul>	<ul style="list-style-type: none"> <li>Spelling</li> <li>Capitalization</li> <li>Punctuation</li> </ul>

## APPENDIX G. CLEP SCORING RUBRIC

### CLEP English Composition with Essay Examination

Readers will assign scores based on the following scoring guide. The essays must display the following characteristics in response to the assigned task.

A 6 essay demonstrates *a high degree of competence and sustained control* although it may have a few minor errors. A typical essay in this category

- addresses all elements of the writing task effectively and insightfully
- develops ideas thoroughly, supporting them with well-chosen reasons, examples, or details
- is well focused and well organized
- demonstrates superior facility with language, using effective vocabulary and sentence variety
- demonstrates general mastery of the standard conventions of grammar, usage, and mechanics, but may have minor errors

A 5 essay demonstrates *a generally high degree of competence* although it will have occasional lapses in quality. A typical essay in this category

- addresses the writing task effectively
- is well developed, using appropriate reasons, examples, or details to support ideas
- is generally well focused and well organized
- demonstrates facility with language, using appropriate vocabulary and some sentence variety
- demonstrates strong control of the standard conventions of grammar, usage, and

A 4 essay demonstrates *clear competence* with some errors and lapses in quality. A typical essay in this category

- addresses the writing task competently
- is adequately developed, using reasons, examples, or details to support ideas
- is adequately focused and organized
- demonstrates competence with language, using adequate vocabulary and minimal sentence variety
- generally demonstrates control of the standard conventions of grammar, usage, and mechanics, but may have some errors

A 3 essay demonstrates *limited competence*. A typical essay in this category exhibits ONE OR MORE of the following weaknesses:

- addresses only some parts of the writing task
- is unevenly developed and often provides assertions but few relevant reasons, examples, or details
- is poorly focused and/or poorly organized
- displays frequent problems in the use of language
- demonstrates inconsistent control of grammar, usage, and mechanics

A 2 essay is *seriously flawed*. A typical essay in this category exhibits ONE

OR MORE of the following weaknesses:

- is unclear or seriously limited in addressing the writing task
- is seriously underdeveloped, providing few reasons, examples, or details
- is unfocused and/or disorganized
- displays frequent serious errors in the use of language that may interfere with meaning
- contains frequent serious errors in grammar, usage, and mechanics that may interfere with meaning

A 1 essay is *fundamentally deficient*. A typical essay in this category exhibits ONE

OR MORE of the following weaknesses:

- provides little or no evidence of the ability to develop an organized response to the writing task
- is undeveloped
- contains severe writing errors that persistently interfere with meaning

0 Off topic (i.e., provides no evidence of an attempt to respond to the assigned topic), in a language other than English, merely copies the topic, consists of only keystroke characters, or is nonverbal.

## APPENDIX H. WRITING PROMPT AND FIVE-PARAGRAPH ESSAY SAMPLE

### Writing Situation

Your school board has proposed creating an Honor Council of students. This council would be responsible for disciplining students accused of violating school rules. The proposal has caused controversy in your school.

You decide to write a letter to the school board expressing your opinion of the proposal.

### Directions for Writing

Write a letter to the school board either supporting or opposing the Honor Council. Use reasons, facts, examples, and other evidence to support your position.

### Student Response

Dear School Board,

The students at [blank] school have a good and fair way of disciplining students without an Honor Council. As a student at this school, I strongly disagree with the proposal of creating an Honor Council of students. The kids in this school aren't just students, they are also great friends, and this can cause problems within the school. Also, selected students will receive unfair power and disciplining other students isn't our responsibility.

First, many student on the Honor council will base their accusations and opinions depending on how well or how close they are with the person being accused. Most students will favor their friends and not give them the fair discipline. The disciplining will not be as harsh because students are not very responsible, and they don't usually do what's right or fair, but what's "cool".

Second, the honor council students will receive more power than others. This may cause separation between the students and the school. Friendships might fall apart just because of too much power given to a few students. If the school isn't careful, it may give power to the wrong people. Students with a lot of control will contribute nothing to our school.

Lastly, it is not the students' job or responsibility to discipline other students. Discipling students is the administrations job. They know all of the rules and they know how to be fair. High school students have many other things to worry about. Such as, homework, tests, and personal relationships. Students also work. It has been state that 89% of all high school students have a part-time job. Many of those students are honor students, and having them in the honor council would just take them away from their studies and jobs.

In conclusion, I strongly oppose the honor council of students. We are just students, not the administration. Friends will always come first. There will be unfair power and this isn't the students responsibility. There is no need for the Honor council. High school student have their own problems to worry about, we don't need to worry about anyone else's disciple problems.

Sincerely,



## APPENDIX I. ADDITIONAL TABLES AND FIGURES

Table I.1

*Summary of alignment between HSPA and KSUS using all rater judgments*

KSUS Standard	Categorical Convergence	Depth of Knowledge Consistency	Range of Knowledge	Balance of Knowledge
A. basic grammar conventions	Yes. Total matches to the rubric = 16	Yes. 100% of matches were at or above the DOK level of the objective.	Yes. 100% of the objectives were matched to the rubric.	Yes. Index of .87 for proportion of matches to objectives.
B. conventions of punctuation and capitalization.	Yes. Total matches to the rubric = 19	Weak. 48% of matches were at or above the DOK level of the objective.	Yes. 100% of the objectives were matched to the rubric.	Yes. Index of .99 for proportion of matches to objectives.
C. conventions of spelling	Yes. Total matches to the rubric = 7	No. 12% of matches were at or above the DOK level of the objective.	Yes. 67% of the objectives were matched to the rubric.	Yes. Index of .98 for proportion of matches to objectives.
D. write clearly and coherently	Yes. Total matches to the rubric = 29	Yes. 82% of matches were at or above the DOK level of the objective.	Yes. 85% of the objectives were matched to the rubric	Yes. Index of .94 for proportion of matches to objectives.
E. communicate ideas, concepts, emotions, and descriptions	Yes. Total matches to the rubric = 26	Weak. 44% of the matches were at or above the DOK level of the objective.	Yes. 88% of the objectives were matched to the rubric.	Yes. Index of .95 for proportion of matches to objectives.
F. revise and edit	No. Total matches to the rubric = 0	No.	No.	No.

Table I.2.

*Clarity and Organization Family*

Code Name(s)	Positive Example	Negative Example
Clarity – focus	Because they can focus. They can organize.... They have a beginning middle and end. (Interview 3, line 123)	My problem here is focus. It kind of drifts, but I'm not sure what is going on here. Um. You know, just back and forth with big gaps in information. (interview 3, line 141)
Clarity – paragraph structure 5 paragraph, Predictable structure	There's a structure. There's a thesis. There's lead sentences supported by details. Reasonably good paragraph development. (interview 6, line 9)	topic sentences were not as strong as I would have liked to have seen (interview 2, line 63)
Transitions	when there working with... transitional words and phrases. This student does that, and he or she uses that type of writing to flow from one idea to the next mostly. (interview 2, line 113)	transitions... this brings me to my third reasoning... okay, right now I try to teach them to stay away from another, third, in conclusion. (interview 1, line 26)
Clarity – logical flow & Clarity over creativity	The story follows along logically in a chronological order or logical order. And, a little bit of a surprise ending, but a surprise ending that makes sense, not an out of the clear blue sky surprise ending... (interview 4, line 90)	there's no organized thought process here on paper. (interview 2, line 45) the imagery was very striking, but there were a lot of ... places where the story didn't seem to connect to the previous line (interview 4, line 88)
Clarity – diction	And they understood an audience. They knew not to use inflammatory language or an insulting tone. It was acceptable vocabulary. (interview 3, line 13)	it just has severe errors in the writing and diction. (interview 6, line 56)
Clarity – voice	they use a lot of voice. It almost seems the student is right there, saying let me tell you this... (interview 1, line 46)	I am saying the student would have to begin to recognize again appropriateness of tone as well as other things. (interview 6, line 68)
Lacking understanding of genre	n/a	simply put it's almost accusatory. It's not appropriate ....very informal almost like a wandering email type of letter. (interview 2, line 131)

Table I.3.

*Communicating Ideas Family*

Code Name(s)	Positive Example	Negative Example
KSUS – communicate ideas	Deal with complexity of the issue nicely... examines various topics within this issue so it's a fairly comprehensive response (interview 6, line 13)	They had a structure but it's not a substantiated argument (interview 8, line 20)
Importance of evidence	Nice points that um... that the audience can very much relate to... uses some statistics. (interview 3, line 44)	I'd expect that each of the body paragraphs would contain evidence, even if anecdotal evidence for reasons and not just outright claims, or bare claims, that's all I really see here are bare claims (interview 2, line 45)
Lack of development	n/a	In writing you have to put description so you can capture the emotions of the characters. (interview 1, line 76)
Ideas more important than errors	If the ideas can be there in a timed writing environment, then that to me suggests college readiness, regardless of structure. (Interview 2, line 107)	n/a
Lack of creativity	n/a	it is written in a pedestrian style but has college potential. Pedestrian meaning it is average there is nothing impressive about it and it doesn't reflect critical thinking ... it seems very high school. (interview 7, line 95)
Understanding audience	Clear audience awareness. They even impressed me as they used sarcasm and not inappropriately to insult the audience (interview 3, line 81)	You might understand what you just wrote, but I don't. And if I don't, you have not met the audience's need. (interview 7, line 127)
Critical thinking	n/a	It is somewhat inventive in the sense that the student did think about what this issue might mean ... for the school board and these honor students (interview 2, line 131)

Table I.4.

*Grammar and Mechanics Family*

Code	Example
Ideas more important than errors	<p>“there are noticeable writing errors for a very strong paper, but again I don’t they think they in any way compromise the effort that the writer makes. We are aware of them as we are reading but it doesn’t really interfere” (interview 6, line 76).</p> <p>“we are taught to count organization and structure over language. So, um, you are only supposed to count it if it impedes your understanding of it” (interview 8, line 30).</p>
Grammar	<p>“a real sense of fluency with the language. (interview 6, line 34).</p> <p>“they have an hour to do this, and I thought they showed pretty good control at least when they go into their freshman year with this, this would be a first draft and they’ll have some time to improve it based on how the instruction is going or the requirements of that course” (interview 1, line 6).</p>
Grammar easier to teach	interview 1 said, “Grammar, you drill it in them. You can get books to them. There are so many things for grammar and structure and organization, but not a lot on voice” (line 50).
Grammar easier to teach with narrative assignments	We found and continue to find there is a need to increase fluency first of all before we move on. We just can’t assume those are going to be there, and if they are there they are in varying degrees of competence. Some people are comfortable producing text and other are not as comfortable producing text, so that is one of the rationales for included some kind of experiential kind of writing, some personal kind of writing. Certainly by the mid-point or by the third essay assignment in the comp level it should be moving toward the more expository kind of writing that we are more familiar with. (interview 6, line 102)
Language issues, ESL	“Serious and severe writing errors that interfere with meaning.”

Table I. 5

*Process Family*

Code	Example
Creative approach	To me, there is a certain level of creativity that a student needs, particularly with prompts, such as the one seen on the HSPA where you're asked to use evidence, but you're not given any (interview 2, line 65).
Importance of revision	if we could get this person and have them, you know, do some re-writing, they would have a good essay" (interview 7, line 108).
Evidence of planning	"What we see in this essay that jumps out at me was the use of contrast, for one. That, to me, is indicative of planning— that the student is considering but both sides of an issue" (interview 2, line 23).

Table I.6

*Causes and Solutions Family*

Code	Example
Focus on basics	I think minimal competence needs to be there. Quality and correctness in writing. Adequate sense of organization and development. So HSPA speaks to that across the board (interview 6, line 184).
Persuasive more relevant	the persuasive one, if they don't have that down pat for an hour of writing, that could give a university good enough idea about them coming (interview 1, line 112)
Quick thinking but not relevant to teaching of writing	It measures, um, whether you can write a persuasive essay in 60 minutes so I guess you have to be um quick thinking. You have to be organized and able to write in a pretty capable manner. Because that's not something you can fake. You have to come in and you have to, I mean there are ways to fake passing the HSPA, but I don't mean that in a bad way. I mean we tell them all the time don't worry about what you really think. Make up your statistics. Make up your reasons. It just has to have your development. It doesn't have to be true. But um to me that doesn't prepare a kid to how do I then take my own argument and sustain my own argument in college. (interview 8, line 111)
HSPA not preparing/not relevant to college	That gap between the high school and university like I said some can make the leap and some are saying oh my, this is so hard for me. I wrote the five-paragraph essay. This is what I did and this is all I know. So when I say under prepared they are not really challenging them and very little critical thinking goes into the writing and it's just very sad. Really it's frighteningly sad for me and I just don't think high schools are doing their job. I'm talking about public high school, most students that come in through private high schools are better but not by miles but they are better prepared. They are better prepared to make adjustments and break some molds. (interview 7, line 36)
Does not acknowledge importance of evidence	It's not indicated whether you are able to make judgments about what kinds of evidence are good evidence. I mean it does it minor ways. Like if you know statistics are a good thing to use as evidence in terms of hierarchy of evidence, but it seems to me that that it's so often used in these contrived situations. If I look at college in terms of English writing, you are going to have to be able to analyze text in depth and take that text and sustain your argument and there's nowhere where it [HSPA] does that. (interview 8, line 115)
Not relevant to the process, revision, not relevant to pedagogy	We spend so much time teaching them that writing is a process. You have to spend a lot of time to plan it and write and then go back to it and rewrite it. And then that's it, one shot, do you want school uniforms, there's no way of preparing for that or even showing that you can go back and work on it again and take another look at it. It's kind of counter intuitive to teach it as a process and then turn around and say this is it. We are not going to indicate, or we are not going to look at all whether you've engaged in a process, just whether you can write it in the confines of this setting. There's something to be said about be able to perform on the spot, but I think that you should also take a look at some of the other work that is being done to get a better look at what a student can do in their entirety. (interview 8, line 129)
Flexibility with organizational patterns/genres	flexibility in an organizational patterns based on the kind of writing that the student is being asked to do. So obviously you can organize a narrative paper one way and can organize a persuasive essay another way and a higher level argument essay a third way" (interview 6, line 178).
Lack of transparency	come March it's something different [than the picture prompt] and that determines whether you graduate or not" (interview 1, line 128).

## CURRICULUM VITAE

Robert J. Finnegan

**Education**

Loyola College, 1994, BA, English/Writing  
Naropa University, 1997, MFA, Poetics

**Principal Occupation**

Educational Testing Service, 1998-2004, Assistant Reading Assessment Coordinator,  
National Assessment of Educational Progress

Educational Testing Service, 2004-2009, Website Content Coordinator, National  
Assessment of Educational Progress

Educational Testing Service, 2009-present, Director, NAEP Reporting

**Publications**

Donahue, P.D., Finnegan, R.J., Lutkus, A.D., Allen, N.L., Campbell, J.R. (2001). *The nation's report card: Fourth-grade reading 2000*. Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 2001-499).

Finnegan, R.J. (1997). *Voicing Anonymity*. Unpublished master's manuscript. Naropa University, Boulder, CO.