

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Dissertations, Theses, and Student Research:
Department of English

English, Department of

April 2008

THE SCIENTIFIC MANAGEMENT OF WRITING AND THE RESIDUE OF REFORM

Eric Turley

University of Nebraska - Lincoln

Follow this and additional works at: <http://digitalcommons.unl.edu/englishdiss>



Part of the [English Language and Literature Commons](#)

Turley, Eric, "THE SCIENTIFIC MANAGEMENT OF WRITING AND THE RESIDUE OF REFORM" (2008). *Dissertations, Theses, and Student Research: Department of English*. 11.

<http://digitalcommons.unl.edu/englishdiss/11>

This Article is brought to you for free and open access by the English, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations, Theses, and Student Research: Department of English by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

THE SCIENTIFIC MANAGEMENT OF WRITING
AND THE RESIDUE OF REFORM

by

Eric D. Turley

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: English

Under the Supervision of Professor Chris W. Gallagher

Lincoln, Nebraska

May, 2008

The Scientific Management of Writing and the Residue of Reform

Eric D. Turley, Ph.D.

University of Nebraska, 2008

Adviser: Chris W. Gallagher

Working at the intersection of composition, writing assessment, and school reform, this dissertation draws on an archival study of Progressive Era educational journals and a year-long qualitative study in a small urban district to examine the ways standardized writing tests are implemented as tools in public school reform. In the first half of the dissertation, I argue that administratively-minded Progressive Era school reformers, in a response to a “writing problem” framed around teacher inefficiency, designed tools for teachers to measure writing “objectively” in their classrooms; however, these tools were quickly used against teachers by administrators interested in efficiently managing schools based on the ideology of Frederick Taylor’s scientific management. In the qualitative study, I examine how Butler Public Schools (BPS) created a Writing Graduation Exam (WGE), as a response to a perceived threat of accountability from outside the district. In both historical and contemporary contexts, I demonstrate how the framing of the problem led to the implementation of standardized tools and school structures that work to manage teachers, students and writing. Juxtaposing the historical and contemporary sites, I trace how a piecemeal version of scientific management is embedded in large-scale writing assessment in order to yield efficient and accountable teachers, students, and writing in schools.

The second half of the dissertation exposes moments of fissure that exist within BPS’s system of accountability. Through the fissures, teacher’s subjugated knowledges (Foucault) surface and provide a critique of the intended and unintended consequences of the WGE. In chapter two, I explore counter-narratives of the “official story” to illuminate the intended and unintended consequences of BPS’s school reform; chapter three examines disruptions in a scoring session over concerns of test validity; and chapter four exposes how teachers negotiate their judgment between disparate theories of writing.

I contend that excavating teachers' subjugated knowledges serves as a potential space to reimagine current practices of school accountability and writing assessment.

Acknowledgements

To Chris W. Gallagher, who invited me into the field of composition and modeled how academic work might be “real.” Thank you for the countless hours dialoging in your office, your careful response to my writing and thinking, and the critical collegueship you have provided throughout my graduate career.

To Deborah Minter and Amy Goodburn, in whose courses many of the ideas of this project originated and who served as readers for this project. Thank you for pushing my reading, writing, and thinking while encouraging me to seek out and wrestle with complexities.

To Joy Ritchie, whose courses broadened my thinking about teaching, research, and community activism. Thank you for providing opportunities for praxis and a space for me to reflect on the relationship between the academy and the community.

To Paul Olson, who in my first seminar at UNL told me to pursue “what puts a fire in your belly.” Thank you for showing me what it means to be a civically minded scholar.

To the Graduate College at the University of Nebraska, thank you for providing me the time and financial support to conduct a year long qualitative study.

To the ten teachers, two principals, and two district administrators (who will remain nameless, due to IRB protocol), who welcomed me into your buildings and classrooms. Thank you for sharing you time, experiences, and professional knowledge with me.

To Whitney Douglas, Dana Kinzy, and Sandy Tarabochia, who have read and responded to various drafts of this project. Thank you for the support and insights you have provided this project, as well as your friendship outside of Andrews Hall.

To Ramon Guerra and Tyrone Jaeger, with whom I have navigated graduate school for the past six years and whose work I very much respect. Thank you for conversations over coffee about academic and nonacademic issues.

To my parents, Vince and Linda, thank you for your constant support and love.

To Alex & Katelyn, who daily remind me of the importance of literacy and education. Thank you for providing me breaks from this project to play baseball, restaurant, and take walks to the park.

And, most of all, to Julie, whose love and support made this possible. Thank you for all the sacrifices you have made to help make one of my dreams come true.

Table of Contents

<i>Vignette:</i>	<i>Inside the Black (Mail)box</i>	1
Introduction:	Writing Assessment and the Residue of Reform	4
<i>Vignette:</i>	<i>Fire Alarm</i>	28
Chapter 1:	Writing in an Era of Efficiency: Toward a Science and Management of Writing	30
<i>Vignette:</i>	<i>A District without Schools</i>	68
Chapter 2:	Creating a System of Writing: The (his)Story and Counter-narratives of the Writing Graduation Exam	72
<i>Vignette:</i>	<i>Do You See What I See?</i>	106
Chapter 3:	Disruptions in Scoring Writing: Possibilities for Conversations on Reliability and Validity	108
<i>Vignette:</i>	<i>Compliment, Question, Critique: Repeat</i>	137
Chapter 4:	Complexities in Teachers' Judgments of Writing	139
Epilogue		178
Appendices		188
Works Cited		198
End Notes		208

Inside the Black (Mail)box

As I opened the tattered black mailbox at the end of my driveway I saw three pieces of mail. I quickly reached in, shut the door and began my return to the house. Two letters were from Vanguard, one for me and one for my wife, updating us on our IRAs. The third piece was the Community News, the quarterly publication of the Butler Public Schools. Entering my house I opened the Community News, Being a teacher, I found this more interesting than the IRA updates, which I threw on the kitchen counter for my wife to look over when she came home. The headline of the small newsletter announced, “A year that made history for Butler Public Schools” and indeed it was. The district passed a \$250 million bond, student enrollment was at an all time high, and the district reported the highest graduation rate ever. On the back page were pictures of four teachers who won local, state, and national teaching awards. The 2005-2006 school year was very successful for Butler Public Schools.

On the center of the front page straddling the fold was a blue box announcing, “Butler public schools high school seniors taking the ACT have scored the highest in BPS history, dramatically improving in most subjects.” The announcement continued, “on a range of 1 to 36, BPS composite scores for 2005-2006 increased from 22.6 to 23.0. BPS students also made major increases in English (from 21.8 to 22.4), reading (from 22.7 to 23.4) and science (from 22.5 to 22.7). Last year almost 1,300 BPS students took the ACT, America’s most widely accepted college entrance exam.” It was hard to miss the blue box and if the average citizen didn’t understand this was big news, the two graphics depicting fireworks on the right side of the article may have helped tip them off.

The middle pages of the newsletter consisted of different graphs and charts that illustrated the academic success of the district. On the left page was a series of bar graphs illustrating the percentage of 4th and 12th grade students meeting state standards for math and English—all of the graphs extended into the 80 and 90 percent range. On the bottom of the left page the district achievement tests charted progress over the last four years for 3rd and 6th graders. In every category the students' average scores were higher than they were over each of the past four years. On the next page, the passing rates of the graduation exams were all in the upper 90's: 97 for reading, 95 for math, and 96 for writing. High school graduation rates going back 18 years were plotted on a graph; the last four years marked a clear upward movement to the new record graduation rate of 80.4%. Finally, the ACT scores which were covered on the first page were put into a graph comparing Butler Public Schools with the state and national averages—the state higher than the national average and BPS even higher than the state.

It was hard not to be impressed with the success of the local school district. The numbers and charts were more than enough to convince me that some good things must be going on in the schools. And when the increasing scores were accompanied with pictures of students and teachers smiling and learning together it made sense that the district had positive news to share with the community it serves.

The two letters from Vanguard were still on the counter. I knew what was inside; normally I don't look at the statements and leave them for my wife, but today I decided to look. At the top of the page was my account total, not enough to retire on, but a start. I immediately looked at the last column the most recent value and was happy to see that it was higher than the value six months ago—a clear sign of financial progress. The

backside provided a breakdown of each fund that is part of my portfolio, similar to the breakdown of the standardized test scores into their smaller components. Here I could look at a lot of numbers that I really didn't understand, but on the far right of the page were the columns that mattered, and in two of the three columns the share price had gone up, causing my investment to increase in value.

It ended up being a good mail day. The schools in my town were increasing their scores and my retirement fund, though still small, was growing. I think a lot of people read their mail the way I did that day. If the numbers are going up—whether it is investment funds or school scores—it must be a good thing. But numbers can be deceiving. There is pressure put on people within organizations—either a corporation or a school district—to assure their stakeholders that their numbers will continue to rise. Corporate collapses and malpractices like those of Enron, Arthur Anderson, and WorldCom have uncovered the elaborate schemes created to increase profits and shareholders' investments. And at the same time stories circulate throughout the media of teachers cheating on standardized tests to get the highest test scores out of their students. While corporations and schools are not apple to apple comparison, education in the twenty-first century has clearly adopted an accountability model based off a corporate structure that focus on measuring output and using data to value what goes on in schools. So much attention is focused on the numbers that it begs the questions: What are these numbers? Where do they come from? Who do they serve? And who do they hurt? And on a very practical note, why can I read my Vanguard statements and the school newsletter the same way?

* * *

Introduction

Writing Assessment and the Residue of Reform

Several times a year, my mother, an elementary school teacher, and I discuss the merits and drawbacks of standardized tests. Truth be told, more often than not we focus on the drawbacks. Our conversations have paralleled discussions in professional literature and national debates on education: standardized tests narrow the curriculum, create a mismatch between the test and the school's curriculum, take time away from real learning, create a test-prep game, compare but don't improve schools, are designed by people who are not teachers, and the list goes on. When I was teaching high school my mom and I would compare how similar our experiences were preparing students for the 5th grade and the 10th grade state writing tests. We discussed how the test encourages students to fill in the form of a five paragraph essay, how real thinking and research is pushed aside by identifying three main points to support a topic the student might not even be interested in writing about, and how our students don't enjoy writing for these tests.

Despite the fact that the majority of my high school colleagues claimed not to teach the five paragraph essay, in the spring, just prior to the state writing test, you could walk the halls and hear teachers drilling the five paragraph form. And not just in English classrooms; our school, after all, believed in writing across the curriculum. Teachers reminded students to think of three examples, provide detailed support and make obvious transitions between points. In addition to the formulaic writing, there were "tricks" to

doing well on the test such as: write as neatly as possible, try to stretch your paragraphs to the next line so they look longer, use remaining time to go back and incorporate larger words or combine sentences to make more complex sentences.

I despised the state tests in the spring, mainly because of how I responded to them. I stopped my students from doing what I thought was good writing and thinking and spent two to three weeks having my students write responses to prompts that couldn't have been more vague: What makes a good friend? Who is your role-model and why? Should the school allow open campus for lunch? Should all high school students be required to work a part-time job? It wasn't that these prompts couldn't be used to do interesting writing, but I used them because they were similar to the prompts my students would see on the state tests. Furthermore, I made sure my students could write in the way the state defined as good writing. We spent time reading the example papers of top scoring essays, analyzing them, trying to figure out how to write well for the test, and mimicking their moves in our practice test sessions. Then, after the exam was over, we returned to the writing we were engaged in prior to the weeks of test-prep.

I was frustrated with what I allowed these writing tests to do to my teaching, but I truly loathed the day the local newspaper printed the scores of all the schools in the county in rank order. It was never a surprise that the wealthiest schools scored higher than my school and the poorer schools scored lower. What seemed important to my administration was that our scores were higher than the community adjacent to ours—bragging rights I guess, they may have beat us in football, but we got them in persuasive writing. While being in the top half of the county was important to the school I worked for, showing improvement from the year before was also stressed. But I didn't understand

why it was important that our scores be higher than the previous year. Sure we wanted our students to do well and show improvement, but what *did* it say about our school, students, and teachers if the average score of the class of 2001 was .08 higher than the average score of the class of 2000 on a different test? After all it was different students taking a different exam. Furthermore, how did our school improving .08 in a year help the students who took the test in 2000 become better writers? They were already on to different courses and electives. Some of them would never take an English class or a writing intensive course again in their high school experience. These were questions I found myself asking, but not finding an answer.

Approaching this dissertation, I bring with me a decade of teaching writing at the secondary and postsecondary level. Though I'm acting the role of a researcher, the questions that form the foundation of this project are those of a teacher. Yet, these questions are not mine alone. I also bring to this dissertation conversations over my professional career that I've had with my colleagues (at both levels) who are concerned about how students learn to write, writing pedagogy, the intersections of policy, theory and practice, and the politics of literacy and education. This dissertation continues these conversations through an exploration into the origins of standardized writing tests, the ways standardized writing tests serve as tools in writing and school reform, and the impact standardized writing tests have on the culture of a school. Specifically, I aim to address the following sets of questions:

- What are the origins of standardized writing tests within the fields of English Education and Composition? Where did they come from? Who

created them? What were their purposes? And what problems were they trying to solve by implementing them?

- Why do politicians, policymakers, administrators, and teachers turn to standardized writing tests to solve their writing problems?
- How do standardized writing tests impact student writing, teachers' pedagogies and theories of writing, and school culture?

The Scientific Management of Writing

The first phrase of my dissertation title, “The Scientific Management of Writing,” comes from Frederick Taylor’s work on industrial efficiency, the socio-cultural and educational impact that resulted from his industrial reform, and the impact it had on the discipline of English and the teaching of writing. In both Shop Management (1903) and The Principles of Scientific Management (1911), Taylor, an industrial engineer, established an ideology—both a theoretical and a systematic approach—to create efficient workplaces throughout society. While Taylor’s industrial reform focused primarily on steel factories, his principles of scientific management were taken up and advocated by social and industrial reformers who applied them to social institutions such as the military, government, schools, universities, and homes in order to make work more efficient. In fact, the legacy of Frederick Taylor’s ideas on efficiency have led historian Peter Drucker to argue, “it is Taylor, and not Marx, who should stand alongside Darwin and Freud as the principal shapers of modern sensibility” (cited in Kliebard 49).

Taylor shaped ‘modern sensibility’ through the use of science as a means to justify the organizational management of a factory. Central to this process was the ability

to measure worker output in temporal or quantitative units. Prior to scientific management, industrial management was often unaware of how much work could be accomplished by laborers in a week, day, or hour. The creation of a standard output of work and tools to measure worker output served as the missing link to knowing if a factory was running on maximum efficiency. Taylor focused on how work could be broken down into smaller movements and measured to assure that each worker did not waste energy or time. For Taylor, finding the most efficient way to accomplish a task was a scientific process which relied on objective measurement. Once the best (most efficient) way to accomplish the task was established, workers were trained to perform the task within the parameters of the established science.

At the same time Taylor's scientific management was making waves within industry, the discipline of English was taking shape through the creation of the National Council of Teachers of English (NCTE) in 1911. Soon, Taylor's ideas were referenced in the pages of English Journal as well as other educational journals, by administratively minded reformers who sought to run efficient schools. These reformers critiqued the teaching of writing for its subjectivity and argued for the need to measure writing accurately using scientifically designed tools to effectively and efficiently measure student writing. Once established, the creators of the tools professed them as "objective" and more reliable than the judgment of an individual teacher. Furthermore, the ability to generate "objective" data from the new tools allowed administrators to compare the effectiveness of their teachers and the ability of their students with other schools in the city, as well as across the state and country.

Although scientific management has its roots in industrial reform, notions of measurement and management are very much a part of the fields of writing assessment and composition today. Theoretical and technical conversations in contemporary writing assessment often focus on the assessment's reliability and validity—two terms adopted from the field of educational measurement. Since the late 1970's, compositionists have questioned and studied what *is* and *isn't* measured by standardized writing assessments. These questions have led compositionists to assist in creating tests that measure student writing rather than students only filling in bubble sheets (White), and examine the impact a writing assessment has on minority students (White and Thomas). Additionally, new theories for writing assessment have been imagined to account for shifts in psychometric theory on validity while also considering the local educational context of learning (Huot). Other scholars have called for a paradigm shift away from ideas of objectivity and science presented in psychometric theory and toward hermeneutical or interpretive frameworks of doing writing assessment (Broad, Wilson), while others have called for a separation from traditional concepts of psychometrics and advocate creating new theories along constructivist theories of education and social epistemic theories of composition pedagogy (Lynne). Visible in the field of writing assessment is the work of some compositionists to shift away from viewing writing assessment as a scientific and objective form of measurement. Instead, there is momentum to build new theories and models of writing assessment that are context-specific and in line with current theories of writing and pedagogy.

While a shift is occurring in post-secondary conversations on theories of writing assessment, historically standardized tests have been used within institutions as a

management tool to place students within an institution's curriculum or used as an exit or graduation exam. These exams help writing program administrators manage the large number of students moving throughout secondary and postsecondary institutions.

Additionally, standardized writing tests have been required by policy makers, specifically in a K-12 context, as a form of mandating school accountability. In both contexts, these exams often manage the work of teachers who are held responsible for students' performance. The pressure to 'teach to the test' has resulted in a recent boom of publications on "test writing" or "on demand writing."¹ These texts are marketed to assist teachers in preparing their students for the variety of standardized writing tests they may encounter in their educational careers. Because standardized writing tests often define writing in narrow and prescriptive ways, teachers conform their instruction to the values of the tests to ensure their students succeed (Hillocks). Thus, one consequence of using standardized writing tests within an educational setting is the creation of a culture of management whereby teachers manage students writing in accordance to the test while, simultaneously, they, too, are managed by the mandates of the test.

The relationship between measurement and management is a central component of this dissertation as these practices continue to be used to reform schools and writing. A deeper examination of Frederick Taylor's systematic design of scientific management provides insight into the ideology of social efficiency and the ways this ideology was imported into schools during the Progressive Era. Through the creation of standardized tools to measure writing and teachers the residue of Taylor's ideology and practices are embedded within contemporary theories and practices of writing assessment today.

The Residue of Reform

Writing and writing assessment do not happen in a vacuum. Rather, both are part of a complex network of stakeholders, socio-political forces, and educational institutions. Therefore, when discussing writing assessment, attention also must be paid to the political nature of school reform and the teaching of writing. In “Power and Agenda Setting in Writing Assessment,” Edward White illustrates the ways various stakeholders approach writing assessment, including teachers, researchers and theorists, testing firms and governing bodies, and students. White makes clear the differing perspectives and agendas each group brings to debates on writing assessment and argues for a negotiation between the various stakeholders. Earlier in his career, White warned compositionists of the overt ways politics and power influence writing assessment, “assessment is power and power is a root political issue...No one should imagine that a test is above politics or that an assessment program is outside the political arena” (246).

Writing assessment *is* political. It’s political because it is typically implemented through a policy created by a school board or a state board of education. It’s political because it defines writing in specific ways, and uses scoring devices to measure writing accordingly. It’s political because there are often consequences for doing poorly on the assessments. And it’s political because at its core, writing assessment is about an effort to change or reform districts and schools, teaching and learning, teachers and students.

The American educational system is a product of reform efforts of the past. As a society, we have created a system of education on which clings the residue of reform.

David Tyack and Larry Cuban offer a useful definition of reform:

When we speak of educational reforms, we mean planned efforts to change schools in order to correct perceived social and educational problems. Sometimes broad social crises trigger school reforms, and sometimes reforms were internal improvements initiated by professionals. Diagnoses of problems and proposed solutions change over time. But whatever the reform, it usually entailed a long and complex set of steps: discovering the problems, devising remedies, adopting new policies, and bringing about institutional change. (Tyack and Cuban 4)

The process of correcting “perceived social and educational problems” illuminates the political nature of writing assessment. For educational reform to occur, problems must be defined and solutions proposed. Over time these reforms become part of the “organizational framework” of schools (86). Tyack and Cuban use the term “grammar of schools” to explain how schools have come to exist and function within American society. “The grammar of schooling is a product of history... result[ing] from the efforts of groups that mobilize to win support for their definitions of problems and their proposed solutions” (86). The implementation of such reforms serve to create a predictable and orderly fashion of organizing and managing a school and over time become permanent structures within schools unnoticed to those who inhabit the schools (86).

The residue of reform efforts of the past are still with schools today, often unnoticed or unquestioned by those who inhabit schools. Like the graded school, credit unit, bell systems, and urbanization of schooling, standardized writing assessment has become part of the grammar of schooling. Tyack and Cuban sum up the permanence of

these reform efforts, “they [the grammars of schools] become fixed in place by everyday custom in schools and by outside forces, both legal mandates and cultural beliefs, until they are barely noticed. They become just the way schools are” (86). Every spring, typically because of state mandates, students in schools across the country take standardized writing tests. Typically these tests ask students (often 4th, 8th, and 11th grades) to write a timed response to a prompt within a certain mode (narrative, expository, persuasive). The essays are scored off-site typically by the state or a testing company. As a response to the state test, some school districts have created their own standardized writing tests to serve as an independent assessment or to prepare their students for the state test. The testing instrument and practice, along with the theories that guide it, are almost indistinguishable across context: standardized writing tests have become a part of the grammar of schools, and a fixed component of teaching writing.

Toward a History of the Present

An Archival Study

Tyack and Cuban argue what “has held the grammar [of schooling] in place is not so much a conscious conservatism as it is unexamined institutional habits and widespread cultural beliefs about what constitutes a ‘real school’” (88). In order to understand how standardized writing tests have become embedded in schools and universities, as part of the work done in “real schools,” I undertook an archival study to excavate the “institutional habits” and “cultural beliefs” of the Progressive Era wherein the first writing assessment tools were created.

I began my archival research with the first pages of English Journal. In my initial reading of volumes one through fifteen I was interested in understanding how standardization impacted the teaching of writing in the Progressive Era. I read each issue taking notes on the topic and main ideas of each article. I documented each piece that dealt with standardizing grading practice, standardizing curriculum, standardizing schools and standardizing disciplinary practices. I found tools used to standardize the work of composition teachers such as writing scales, check lists, and standard tests. And I listened on the pages to the voices of teachers, school administrators, and college professors who called for, responded to, or resisted various forms of standardization. Completing my first read through the first fifteen years of English Journal, I realized that standardization was a response to something—a problem, chaos, inefficiency. My first reading provided insight into how standardization was taken up in various forms and in various ways including supervising teachers, articulating educational standards, achieving equity for students, and yielding efficient schools. Through this first reading I began to see that standardization was really a response to a perceived problem by those who managed schools.

A year later, I returned to the stacks seeking to understand the problems that standardizing writing, teachers' judgments, and schools attempted to solve. Who framed these problems? And why did they propose solutions that ultimately led to standardized writing tests? To answer these questions I again worked through the first fifteen years of English Journal, this time quicker spending more time with specific articles and authors. I also began to look outside of English Journal to journals like Teachers College Record, Pedagogical Seminary and Educational Administration and Supervision where

conversations about writing standards, writing scales, and standard tests were also taking place and being cited by the authors of English Journal. Additionally, I attempted to find as many of the writing scales cited in English Journal that I could in order to study the actual tools implemented in classrooms and schools to measure writing with a standardized tool. Finally, I read the major writings of Frederick Taylor along side the conversations of English Journal and the writing scales to consider the cultural moment in which these conversations were occurring.

A Qualitative Study

In this project, I am not just interested in a historical look at standardized writing tests, but if the residue from Progressive Era reform efforts still lingers in schools today. Therefore, I conducted a qualitative study of a school district that uses a standardized writing test as a graduation requirement. Through this study I hoped to understand both the ways Progressive Era school reform efforts exist as part of schools today, as well as the ways the standardized writing test affect the teaching of writing and a school's culture.

Butler Public Schools, located on the Great Plains, serves a growing city of over a quarter million people. Butler is both the capital city and home of the state's flagship university. The school district serves over 32,000 students from pre-school to twelfth grade and is comprised of 36 elementary, 11 middle, and 6 high schools, along with an alternative school and 9 programs offered in separate facilities. Two out of the six high schools were chosen as primary sites for the qualitative study—Wilson and Marshall High Schools. While the two high schools are in the same district, they have strikingly different characteristics which provide two different contexts to examine the impact of

the Writing Graduation Exam (WGE) on each school and the teachers who inhabit each building. Marshall High School, the oldest school in the district, has served the community for 135 years, while Wilson is a newer school in its fifth year. Wilson uses a 4:4 block schedule and Marshall runs a traditional eight period day. Marshall is a more diverse school than Wilson and, overall, Marshall's students pass the WGE at a slower rate than Wilson students. The charts below serve to illustrate the student demographics and passing rates of the two schools.

	Marshall High		Wilson High	
	Total	%	Total	%
Total Student Population	1,555		1,910	
White, not Hispanic	1,029	66%	1,773	93%
Black, not Hispanic	231	15%	59	3%
Hispanic	174	11%	37	2%
Asian / Pacific Islander	73	5%	32	2%
American Indian / Alaskan Native	48	3%	9	0.5%
Free and Reduced Meal	645	41%	237	12%
ELL	183	12%	11	1%
Mobility Rate		26%		8%

	Marshall High	Wilson High
WGE Passing Rate		
Total % of 12th passed WGE	90.2%	98.7%
Total % of 11th passed WGE	81.0%	90.9%
Total % of 10th passed WGE	67.0%	87.2%
Total % of 9th passed WGE	49.2%	70.0%

In the spring prior to the year of fieldwork, I contacted Mark Thurman², then Coordinator of the Writing Graduation Exam, about the possibility of conducting a study on the impact of the WGE on the teaching of writing in different schools. Over a series of meetings and discussions with Mark, he helped me to imagine how my study might take place within the district. I was interested in working with Marshall and Wilson High Schools and Mark helped establish communication with the school principals and the English department chairs of each school.

I attended the first department meeting of the school year at both Wilson and Marshall High Schools where I introduced myself and my research project and invited the teachers to participate in the study. The opportunity sample consisted of ten teachers, five from each building along with the building principals and two district level administrators. Throughout the study I relied on interviews, observations of classrooms and district meetings as well as the collection of documents, data which I braid together in the dissertation. Here I attempt to explain the various threads of the data I used to analyze.

The first part of the qualitative study consisted of a series of interviews with the ten English Teachers, both school principals, and two district level administrators—the Coordinator of the Writing Graduate Demonstration Exam and the Secondary English / Language Arts Curriculum Specialist. I interviewed all participants in the fall of 2006 and the winter of 2007. The first round of interviews consisted of three different scripts of questions: one for the teachers, one for the school principals, and one for the district administrators. The second round of interviews were divided between half scripted questions and the other half were customized for individual teachers based on my

observations. The interviews ran between forty-five and ninety minutes. All of the interviews were transcribed and shared with the participants.

Along with conducting interviews, I spent five to seven hours, four days a week, sitting in on English classes at both high schools. In all, I observed eleven different courses: English 9, English 9 Honors, English 10, English 10 Honors, General English, Composition, AP Language and Composition, American Literature and Composition, Journalism, Women's Literature, and Creative Writing. For some of these courses I observed two different sections within the same school, for other courses like General English and Composition I was able to observe sections at each high school. In addition to classroom observations, I attended several district meetings for the Writing Graduation Exam. The work of these meetings included table leaders selecting anchor papers and training materials for scoring the WGE, the actual scoring of the WGE, as well as the standard setting for the WGE.

Finally, I collected documents from classrooms such as handouts teachers provided students for writing assignments (such as planning documents, graphic organizers, and rubrics, as well as examples or readings). I also collected documents from the district such as technical manuals, example tests, and the WGE rubrics. Observing the classroom as well as the district meetings and coupling the observational research with the interviews and documents provided a lens into the ways the WGE gets played out in classrooms, schools, and the district as a whole.

Throughout the 2006-2007 school year I conducted a concurrent analysis of the data as I collected it. After spending four days in the schools I used Fridays and Saturdays to review my observations and type up sketches and memos based on my field

notes. Through this review of data I was able to create questions for the second round of interviews as well as to conduct more focused observation when I returned to the classrooms.

Starting over spring break I began a comprehensive analysis of the two rounds of interviews. I approached the data with the following broadly constructed questions:

- What impact does the WGE have on Butler Public School District?
- What impact does the WGE have on Wilson and Marshall High Schools?
- What impact does the WGE have on teachers and their classroom practice?
- What impact does the WGE have on students and their writing?

Using ATLAS.ti software to assist in data analysis, I coded all 28 transcripts resulting in 107 categories. I looked to Kathy Charmaz's principles of constructivist grounded theory, specifically her "constant comparative method" of data analysis:

The constant comparative method of grounded theory means (a) comparing different people (such as their views, situations, actions, accounts, and experiences), (b) comparing data from the same individuals with themselves at different points in time, (c) comparing incident with incident, (d) comparing data with category, and (e) comparing category with other categories. (515)

Working through the categories I began to condense and combine categories that overlapped and drew lines of demarcation between categories that contrasted and contradicted with others. In chapters two, three, and four, I focus on different sites of the district. By comparing data between different people, of the same individual, and between incidents, I was able to capture and name various ways the WGE impacted the

district, schools, teachers, and students, the tensions presented within the WGE system, and the ways some teacher and administrators resisted the WGE.

While I draw on archival and qualitative methodologies in this dissertation, I do not intend for these methodologies to be seen in isolation; rather, I am interested in looking across time and the discipline of English to explore the ways science and management have been key components in writing assessment and the ways they have shaped the institutions students, teachers, and administrators work in. Through this dual pronged methodology I am attempting to write what Michel Foucault calls a “history of the present” by exploring the systematic architecture of standardized writing tests within a school district, as well as the capability for teachers and administrators to claim agency within the system (30-31). While I strive to explore the tension between structure and agency, I, too, am interested in uncovering the institutional habits and ideas about writing that standardized writing assessments have infused within schools as well as the ways teachers who inhabit schools today resist these institutional habits and ideas.

Overview of the Dissertation

Critics of standardized test and test-based accountability have argued that these educational tools (some prefer the term weapons) are used against teachers and schools to narrow educational conversations to scores and testing (Kohn 28), deskill teachers and remove their judgment from curricular decisions (Ohanian 99), and focus instruction only on the content areas which are tested (Resnick & Resnick, Smith). In her study of the Texas Accountability System, Linda McNeil argues the “logical consequences” of test-based accountability are designed to deskill teachers, re-stratify access to education and

de-democratize public education (270). For McNeil, accountability performed through standardized tests becomes a technical enterprise with a discourse used to resist criticism:

Finally ‘accountability’ as a closed system admits no critique. Criticisms of the system that have any currency and that generate response are those regarding the technical components of the system...Questions about technical tinkering are tolerated [while] the technical language of accountability silences those professionals who want to stay in public education because it takes away the legitimacy for any other, counterlanguage to shape school practice. This problem is exacerbated as a system becomes so “aligned” that professionals in the system are locked into compliance if they are to stay in the system.” (263)

The systemization of accountability perpetuates teachers acquiescing their classrooms to the tests if they want to stay in the system of education. George Hillocks, in his study of large-scale state writing assessments in five states, makes a similar claim to McNeil.

Hillocks found, “In every state, when teachers have little knowledge of writing, the testing system tends to become the knowledge base for teaching writing” (102).

Hillock’s study is the most comprehensive study on standardized writing assessments to date, and his findings point to teacher’s instructing students in formulaic writing as a means to write in ways that are rewarded by the state rubric.

Critics of standardized tests illustrate the ways state policies or institutional structures manage the work of teachers and administrators working within a system of accountability. Through this dissertation I aim to accomplish two purposes which will

extend and complicate the work of researchers concerned with the effects of standardized writing tests on schools, teachers, and students.

First, I engage in an archival study of Progressive Era writing assessment tools in order to illustrate the ways the ideology of efficiency and the reform efforts of Frederick Taylor's scientific management are embedded in the tools which attempted to objectively measure writing and manage teachers. Contemporary practices of writing assessment, when used as an accountability mechanism, are guided by the use of similar tools as those designed in the Progressive Era and carry the ideology of scientific management into schools.

Chapter one examines the framing of the "writing problem" by Progressive Era reformers who critiqued the teaching of writing in schools. Their definition of the "writing problem" included a distrust of teachers' subjective judgment of writing, a concern on the intense labor of teaching composition, and questions over the kinds of writing students should do in secondary education. Defining the "writing problem" in terms of Progressive Era notions of efficiency, I trace how the ideology of Frederick Taylor's scientific management moved into writing assessment allowing administrators to manage their schools and writing along scientific (and therefore efficient) principles. Coupling Taylor's industrial management ideology with the new field of educational measurement provided the foundation to design "objective" tools (writing scales and standard tests) to measure writing. In the end, Taylor's principles of scientific management were never fully enacted in schools; rather a piecemeal version of Taylor's practices were enacted by administrators looking to efficiently measure and manage their schools and teachers.

From the archival study I move to a contemporary context exploring how Butler Public Schools attempted to respond to their own “writing problem” by implementing a standardized writing test. Chapter two alternates between two different narratives of the WGE. The first tells the “official story,” the collective narrative teachers and administrators tell of the way the test was created, a series of remedial classes was formed to help students pass the exam, how district personnel are utilized to maintain the WGE, and the improvements in writing the district has seen since its implementation eight years prior to the study. Telling this institutional narrative allows for the history and structure of the system to be exposed within the district and illustrates the residue of Progressive Era ideology of efficiency still lingers in writing assessment and schools today. The second telling moves beyond the institutional narrative of a systemic solution to an accountability problem, shared through counter-narratives teachers and administrators told of the WGE system. The counter-narratives work to disrupt the official story to show how the system affects the people working inside, while also complicating the relationship teachers and administrators have with the district and each other. Examining the counter-narratives teachers and administrators tell reveals what Foucault calls “subjugated knowledges.” Excavating subjugated knowledges provides a recovery of teachers’ knowledges which the district’s “official story” ignores or overlooks, and complicates the efficiency narrative of the WGE system.

Second, rather than only examine the ways the system of accountability manages teachers through standardized tests (as McNeil, Hillocks, and other have before me), I conduct a multi-level analysis of the ways the teachers disrupt and critique the system in various sites including the district level, holistic essay scoring sessions, and classrooms.

In doing so I argue that exploring the fissures within the system of the WGE allows for the recovery of teachers' subjugated knowledges and provides a more complex rendering of teachers' experience working within an educational system of accountability.

Chapter three turns to examine a BPS holistic scoring sessions of student papers and the efforts to maintain test reliability. I trace recent conversations on validity and reliability theory prior to raising a series of moments when the district scoring system fissures or breaks down. While these moments are brief, typically when teachers and administrators question or critique the process or their role, it is in these moments that teachers and administrators are voicing concerns of test validity. These questions of validity point to larger issues about teaching writing, the function of the district assessment system and the overall education students receive. I contend that in these moments of fissure, writing teachers can begin to reconsider practices of writing assessment centered on validity.

Chapter four moves from a district scoring session to the classrooms of two teachers where I study the ways the WGE impacts teachers' judgments of writing. While many scholars argue the implementation of standardized tests lead to teachers "teaching to the test," I argue that debate ignores the way power functions systematically on teachers. Drawing on Foucault's technologies of control I trace how disciplinary knowledge is enacted on teachers and impacts their instruction and assessment. Through this reading teacher judgment is eroded and replaced by disciplinary knowledge and the pressure of getting students to pass the exam. However, I also use Foucault's method of discontinuity allowing for the gaps and fissures within interviews and classrooms to illustrate the ways these two teachers draw and enact their subjugated knowledges despite

the attempt to discipline them to the values of writing on the WGE. In the end, I argue teachers neither fully submit to the WGE nor do they fully enact their professional judgment of writing; rather, in their classroom they work between incongruent theories of writing and instruction.

In the Epilogue I take up the ways the fissures within the WGE system can lead to prompting new questions about writing and writing assessment which ultimately can be used to frame debates on school reform and writing.

Composing and Listening

Researching and writing about teachers is a tricky kind of work. Cultural narratives tend to script teachers at either end of a spectrum spanning from the inept practitioners lacking knowledge, skill, and personality (those who can, do; those who can't, teach) to heroic redeemers who overcome great odds: the educational system, resistant students, socio-cultural divides, in order to transform their students' lives (representations of teachers in movies such as Mr. Holland's Opus, Dead Poet's Society, Dangerous Minds, and Finding Forrester). I resist constructing my representations toward either end of this spectrum; instead, I find Ellen Cushman's activist research methodology to provide an ethical goal for representing the work and lives of teachers:

Overall, activist research sets for itself the goal of portraying participants in respectful ways while also conveying the texture and complexities of participants' lived conditions. [...] At the same time as participants must be respected, activist's research resists creating representations drawn from the perspective of liberal idealism. Liberal idealism might lead to

Pollyannish representations or naïve glorifications of individuals' acts of agency. [...] Activist research demands we describe the tension filled place where individuals' acts wrangle with larger structural constraints...

(37)

As Cushman wrote, exploring the ways individuals advocate their agency within structural constraints can lead to “unflattering, perhaps problematic” representations of participants (37). These unflattering representations might not fit neatly when measured against the common cultural narratives of teachers; however, they provide for an examination of the complexities and contradictions of teachers' lived experiences in schools.

In chapters two, three, and four I draw from interview transcripts, recreate scoring sessions, and reconstruct moments from teachers' classrooms. In chapter two I compile teachers' and administrators' voices to create two different narratives, an “official story” and a series of counter-narratives. The goal of chapter two is to juxtapose two narratives, to do so I present quotes from interviews with teachers. The selected quotes only touch on the complex stories and perceptions each individual told and are not meant to represent the quoted teacher as only holding one perspective; rather, the quotes represent a theme several teachers and administrators discussed over a series of interviews. The centerpieces of chapter three are conversations between participants taken from field notes of table leader meetings and scoring sessions. In chapter four I focus on two teachers' judgments of writing, drawing on data from interviews along with observations of their classrooms. Across these chapters I attempt to provide two readings of what is occurring at each site of the WGE system. Rather than give a one-sided reading, which

might get me close to one of the binaries on the spectrum of cultural narratives, the dual-read serves to illuminate the complexities of teachers' lived experiences and the ways they navigate the educational structures they inhabit.

As part of my methodology I strive to share and create knowledge with the participants who allowed me into their classrooms and schools. At this point in time the fourth chapter has been cycled back to the teachers to read, respond to and critique. I have incorporated one of the teacher's responses into chapter four. Chapters two and three are currently being read by all the participants of the study. Their responses will be included in forthcoming revisions of this project.

I have worked hard listening to the participants in my studies talk about their work and lived experiences within schools. As Chris Gallagher expressed in Reclaiming Assessment, "listening to teachers, taking them seriously, ought to be the work of anyone who thinks she has something important to say about education" (5). I believe there is something important to say about education, composition, and writing assessment in the pages of this dissertation; however, it could not have been articulated without first listening to the teachers and administrators who participated in this study.

Fire Alarm

The shrieking sound of the fire alarm streamed out the doors as I exited Marshall High School. The students, excited to be out of English 10, huddled in small groups against the fence that divided the school grounds from the nearby street. I stood close to Andrea, watching her compare her roster to the students who moved about finding friends from other classes to celebrate the disruption in their day.

“It must have been pulled.” She said. “Usually we’re done by now if it’s a drill.”

Though it was an unseasonably warm December day, I was pleased I grabbed my coat. Sliding my arms into the sleeves I could hear the sirens of the fire truck approaching the school.

“So how is your research going?” Andrea asked.

“Pretty well.”

“Have you started writing your diss yet?”

“No. I think I just figured how I am going to put it all together. I want to look at how we got to where we are with standardized writing assessment so I need to do some more historical work.”

“When do you think that all started?” Andrea’s eyes opened wide, “Because, I have a theory about all of that.” She paused, waiting for me to share my thoughts.

“I think it goes back to the Progressive Era when ideas of industrial management, behavioral psychology, and educational measurement all converged onto schools. That’s some of the historical work I might try to do in my dissertation.”

“You know, that’s interesting because in all my curriculum and instruction classes I took at the university for my master’s degree we only talked about John Dewey. He was the only historical figure we ever read and spent time on. And I have a problem with that because education is not like the way Dewey wrote about it. Or maybe I should say the people who run schools don’t read Dewey. Seriously, do administrators read Dewey? Because I don’t think they do. I don’t think it would be like this if they did.”

The fire alarm stopped and the all clear signal was given. The students started moving back toward the building, but I wanted to continue the conversation with Andrea. I wanted to hear more about the tension that she has experienced as an English teacher. I wanted to know if she thinks Dewey’s ideas would work in schools today. I wanted to know why in all of my education classes Dewey was the main historical voice we read? I wanted to find out the answers to my questions about the history of standardized writing tests, the disciplines of English and education. As I thought all of this, I found myself entering the building and like the students and teachers around me I went back to the classroom and to the work that I was engaged in prior to the disruption.

* * *

Chapter 1

Writing in an Era of Efficiency:

Toward a Science and Management of Writing

At the turn of the twentieth century the United States was preoccupied with efficiency. The pursuit of efficiency in industry, government, schools and homes was the byproduct of several socio-cultural forces converging in a particular historical moment including: immigration, the urbanization of cities, the compression of space and time by new technologies like trains and cars, as well as the expansion of industry. Responding to the chaos of an evolving society, efficiency served as a socio-political term around which reform efforts were levied. Comprehensively, efficiency was concerned with the elimination of waste—energy, time, material, and economic. The call for efficiency was made across issues and institutions from President Roosevelt’s decree to increase national efficiency by conserving national resources, to the railroad industry’s standardization of time to improve transportation, to periodicals such as Outlook running a series on improving efficiency in the home (Taylor iii, Callahan 44).

With the formation of the National Council of Teachers of English (NCTE) in 1911, the publication of English Journal in 1912, and annual conferences thereafter, the discipline of English took shape in an era of efficiency. In Radical Departures, Chris W. Gallagher maps the imprint Progressive Era educational reformers made on the field of composition. In his historic rendering, the young field of composition, like education more broadly, came to a divide during the early twentieth century over what was meant

by progressive education. “All early progressives shared an abiding faith that education could serve as a lever for social progress. The fundamental difference between the two main groups, however, hinged on their answer to the question, Progress toward *what?*” (12). Gallagher takes up the terminology educational historians David Tyack and Lawrence Cremin established for the two predominant groups who differed on how to best answer the question, progress toward what? “Pedagogical progressive,” most often associated with the work of John Dewey, viewed education as a preparation for democratic living and social action. The other group, “administrative progressives” comprised of efficiency experts, corporate leaders, and other professionals from the business world attempted to use their social status to influence schools. They believed education should serve the interest of industrialism and business, and should be structured similar to an efficient business. While the two versions of progressive education carried with them distinctly different views, purposes, and goals of education, Gallagher warns against solely viewing Progressive Era education reform as a dichotomous distinction of pedagogies—pedagogical and administrative—which mask a complex historical reading and understanding of the field of composition.³

Gallagher’s reading of the administrative progressives focuses on the incorporation of education along the lines of business-industrial practices which equated efficiency with economy and sought to “improve human capital as a means of economic growth” (Gallagher 14, quoting Spring 153). Gallagher illuminates labor practices, managerial strategies, and disparate factions of reformers involved in the struggle over composition and education more broadly. In this chapter I extend Gallagher’s historical reading of Progressive Era reform efforts in another direction to provide a history and

critique of the ways the ideology of Frederick Taylor’s scientific management crept into the teaching of writing as educational reformers aimed to systematically reform schools through the creation of objective and measurable standards while implementing a hierarchy of management used to maintain accountability to the science around which the system was created.

I begin this chapter by defining the “writing problem” of the Progressive Era, not a single problem, but a series of critiques aimed at the teaching of writing in public schools by university faculty and members of the burgeoning community of educational measurement. I then examine the influence and limitation of Taylor’s work in industrial reform before moving to consider how the solution to the “writing problem” took up a piecemeal version of scientific management and applied it to the teaching of writing in schools. The solution to the “writing problem” was the creation of “scientific” tools produced by efficiency experts to measure writing. Using data collected from the scientific tools, school administrators were able to manage their schools, teachers, and writing based on the “objective” measurements the tools provided.

The “Writing Problem”

In the Progressive Era, writing was a problem for many teachers, administrators, college professors, and businessmen. While not all of these groups framed the problem the same way, inefficiency was at the heart of the problem yielding a call to reform writing in schools along more efficient practices. Therefore, the Progressive Era “writing problem” was not a single problem; rather, it consisted of interrelated and overlapping issues and critiques of the way writing was taught in schools. Combined, the “writing

problem” was concerned with teachers’ ability to judge and evaluate writing independent of a scientifically objective tool, the intense amount of labor that went into teaching and grading writing, and the shifting role of schools to prepare students for working in industry and business. Collectively, these concerns created a chaos around writing which administratively minded educators and reformers sought to eliminate through the adoption of efficient methods of teaching and running schools.

Historian Samuel Haber, provides a four-pronged definition of efficiency demonstrating the vagueness and utility of the term as it was taken up in the Progressive Era. Efficiency in the Progressive Era was defined as: 1) a personal attribute—an effective, disciplined, hard working individual who turned away from feeling, sympathy, and femininity; 2) a mechanical measurement, typically the input-out-put ratio of energy based on scientific laws applied to technology; 3) an economic criteria determined by accounting the input-output ration of money; and 4) social harmony or the relationship between men and “competent” leadership (Haber ix, x). The personal, mechanical, economical, and social definitions of efficiency were ingrained in the framing of the “writing problem.”

Haber’s first definition of efficiency as a personal attribute was the most prevalent critique of teaching writing as teachers’ ability to make an accurate and objective judgment was called into question. This critique was specifically leveled against women teachers who, according to critics, relied too much on emotions to judge student writing. One such critic, Earl Hudelson, a professor at the University of Indiana, called for definite standards in composition to eliminate teachers’ emotions from influencing their judgment of student writing, Hudelson wrote in English Journal:

Moodiness, or temperament, too often decides Johnnie's fate. If a teacher gets up on the wrong side of bed, she may grade a composition 10 per cent lower than she would have had she been in a merrier mood. If Johnnie is near the danger-line, his chances depend upon his teacher's mood. While a standard can hardly be expected to eliminate temperament entirely, it will minimize the evils flowing therefrom. (Hudelson 594)

For Hudelson, the lack of a standard for composition illustrates an inefficient practice where neither the teacher nor the student can measure their writing accurately. His critique focuses on the personal inefficiency of teachers and inconsistent grading practices that could make for an unfair treatment of Johnnie's work.

Additionally, the pronouns in Hudelson's quote point to a gendered concern which cannot be overlooked. "She" the imagined teacher in Hudelson's example cannot control her emotions. The gendering of efficiency reflects a larger concern facing male administrators as they created a systematic approach to supervising schooling. David Tyack argues that the employment of women correlates with the increase of school bureaucratization in the late nineteenth and early twentieth century.⁴ Tyack cites Lotus D. Coffman's 1911 study on the composition of the population of teachers where Coffman concludes that the increase of women in:

the teaching force has been due in part to the changed character of the management of public schools, to the specialization of labor within the school, to the narrowing of the intellectual range or versatility required of teacher and the willingness of women to work for less than menall of the graded school positions have been preempted by women; men still

survive in public school work as ‘managing’ or executive officers. (cited in Tyack 61).

Through the bureaucratization of schools and the growth of administrative degrees, a social efficiency evolved in schools by the credentialing of male administrators who oversaw a predominantly female workforce of teachers. This gendered hierarchy of schools was reinforced by the call for more efficient teachers by the administrators who were credentialed as competent leaders by various teachers’ colleges.⁵

The second component of the “writing problem” addressed the labor of teaching writing within schools and universities. A teacher’s perspective of the labor issue surrounding composition was prominently documented on the very first page of English Journal where Edwin Hopkins poses the question, “Can Good Composition Teaching be Done Under Present Conditions?” His answer, a resounding “No.” Hopkin’s essay draws on the language of the administrative Progressives and points to economic inefficiency and labor concerns around teaching writing.

For every year the complaints become louder that the investment in English teaching yields but a small fraction of the desired returns. Every year teachers resign, break down, perhaps become permanently invalidated, having sacrificed ambition, health and in not a few instances even life, in the struggle to do all the work expected of them. Every year thousands of pupils drift through the schools, half-cared for in English classes where they should have consent and encouraging personal attention, and neglected in other classes where their English should be watched over at least incidentally, to emerge in a more or less damaged linguistic

condition, incapable of meeting satisfactorily the simplest practical demand upon their powers of expression. Much money is spent, valuable teachers are worn out at an inhumanly rapid rate, and results are inadequate or wholly lacking. From any point of view—that of taxpayer, teacher, pupil—such a situation is intolerable. (1)

In The Labor and Cost of the Teaching of English in Colleges and Secondary Schools with Especial reference to English Composition, Hopkins applies cost-accounting practices to composition ascribing the inefficiency of teaching writing was a result of lacking resources and support for teachers.⁶ Hopkin’s work around the labor of composition demonstrates a concern over both the personal and economic efficiency of teaching writing.

A third component of the “writing problem” emerged over the purpose(s) of teaching writing in schools. Within the pages of English Journal a dispute emerged between viewing writing as the enculturation of young minds with rhetoric and literature—the old view—verses the business-minded progressive view that students ought to know how to correspond in the world of business. Hopkins, in his aforementioned English Journal essay, provoked the question “is or is not training in English expression necessary to a successful industrial and business future?” (Hopkins 6). The focus on correspondence through business letters became a featured topic in the pages of English Journal and the rationale directly tied to efficiency and profit. John Clapp pushed for composition to follow the business community enabling students to be proficient writers for commerce. “First of all he will be heartened by the importance which the progressive business community—and business is going to count immensely

more hereafter in the field of education—attaches to the matter of composition as a direct agency of business life. Well written letters, business men have learned, mean good operation and profit; poor letters mean inefficiency and waste” (Clapp 237).

Designing a writing curriculum to appease business leaders went as far as Sherwin Cody’s creation of a joint committee of the Chicago Association of Commerce and the Board of Education, whose report focused on the school system’s “failure to master the common ‘tools of education’ in the seventh and eighth grades, or by the time the compulsory age is completed and beginners first go out into business offices” (412). Cody’s conclusion was to find out “what the business men did want” and to “develop tests which those business men would be willing to make the basis of employment, to a certain extent, giving preferred positions and first choice to those who made the best showing on these tests” (412). The series of tests included an elementary spelling test, grammar test, punctuation test, and letter-writing test. The increased attention to “Business English” “Commerce English” and “Vocational English” in the pages of English Journal reflects a larger trend in curriculum design emphasizing vocationalism, whereby schools served society by producing a stable workforce to run an efficient society.

The “writing problem” of the Progressive Era was framed as a deficiency in efficiency—personal, mechanical, economic, and social—and provided the opportunity for solutions to be proposed which would allow for reforming the teaching of writing along more efficient principles and practices. More specifically, concerns around the management of schools with subjective and overworked writing teachers served provided a need to consider alternative ways of teaching composition and managing teachers’ work

and workloads. Additionally, the business community put pressure on schools and teachers to produce students who upon leaving school could immediately apply their writing ability to service a business or company. Framing the problem as a management issue within and outside of schools resulted in administratively-minded reforms to look for more objective and efficient ways to address the messy and inefficient process of teaching writing in schools.

For example, Ernest C. Noyes from Fifth Avenue High School in Pittsburgh, Pennsylvania wrote in 1912, “Our present methods of measuring composition are controlled too much by personal opinion, which varies by the individual. What is wanted is a clear-cut, concrete standard of measurement which will mean the same thing to all people in all places and is not dependent upon the opinion of any individual” (534). Noyes argued that a “new science of education based upon exact measurement and judgment by ascertained facts” was needed in composition (Noyes 532).⁷ He compared the present situation in teaching writing to the need industry had prior to scientific management. Pressing to standardize the measurement of composition, Noyes continued, “Our present methods of measuring composition are controlled too much by personal opinion, which varies with the individual. What is wanted is a clear-cut, concrete standard of measurement which will mean the same thing to all people in all places and is not dependent on the opinion of any individual” (532). Noyes used a thermometer as an example of a tool that produces a scientific standard by which heat could be measured regardless of the location. Thus, 90° is the same in Boston and Chicago and furthermore a Fahrenheit scale allows for comparison in heat to be made between the two cities. What composition needed, according to Noyes and his supporters, was a scale that would help

to make the assessment of writing as universal as a Fahrenheit scale makes the measurement of temperature.

Calls of reforming writing along the lines of standardization and measurement were popular within the pages of English Journal and were answered by many university professors.⁸ However, beneath this call for writing reform are the principles of Frederick Taylor's scientific management. Prior to exploring the means of solving the writing problem, an exploration of scientific management and its impact on education is useful to illustrate how the ideology Taylorism was championed for achieving efficiency within society, schools and the teaching of writing.

Taylorism: The Call for a Mental Revolution

In the introduction to The Principles of Scientific Management, Frederick Taylor's most popular work, he calls for greater national efficiency and offers a revolutionary system to achieve the goal. Prior to Taylor, efficiency had been sought for in competent men who were able to run companies or households with a mind toward increased productivity and decreased waste. However, Taylor posited finding the right (qualified or experienced) person for the job was the wrong path toward the goal of efficiency. Rather than searching for the right employee Taylor advocated creating the correct system and then training people to work accordingly within the system. "It is only when we fully realize that our duty, as well as our opportunity, lies in systematically cooperating to train and to make this competent man, instead of hunting for a man whom some one else has trained, that we shall be on the road to efficiency" (Taylor iii). Taylor's scientific management represented a major paradigm shift in industrial organization and

management. This new approach drew on scientific measurement to determine the best way to accomplish a task. After the most efficient way to accomplish each task was calculated, an entire system (made up of all the tasks within a factory working simultaneously) was set in place and kept in check by a series of managers whose job was to assure worker performance.

For Taylor there were two major problems industrial management needed to address to achieve efficient factories. The first was the need to eliminate soldiering (lazy workers) as well as systematic soldiering or the organized and subversive acts by workers to keep their “employers ignorant of how fast work can be done” (7). For Taylor, the “greatest obstacle to harmonious cooperation between the workman and the management lay in the ignorance of the management as to what really constitutes a proper day’s work for a workman” (25). Without the management knowing a standard production level for its workers, Taylor assumed a factory operated under capacity, making it wasteful or inefficient.

The second problem facing the modern factory was its dependence of “rule-of-thumb” methods created by workers. Tradesmen learned the details of their work through observing other tradesmen. This practice, according to Taylor led to many different ways to accomplish a task. The reliance on professional lore was untested, resulting in workers using more time than considered necessary to complete the task. Here Taylor explains the need to find the most efficient method rather than depend on rule-of-thumb methods:

Now, among the various methods and implements used in each element of each trade there is always one method and one implement which is quicker

and better than any of the rest. And this one best method and best implement can only be discovered or developed through a scientific study and analysis of all the methods and implements in use, together with accurate, minute, motion and time study. This involves the gradual substitution of science for rule of thumb throughout the mechanical arts.

(9)

Through scientific study, an efficiency expert could determine the best method to accomplish a task along with the proper amount of time the task should be accomplished. With this knowledge a manager could both standardize the process of work eliminating rule-of-thumb practices while simultaneously eliminating soldiering by measuring the time it took to complete the task. Taylor's time-in-motion study served as the foundation for an efficient factory.⁹

After the activities of the factory were scientifically studied and the best method to accomplish a task was calculated by the efficiency expert, a system of managers was needed to implement and oversee the system. Taylor wrote that the process of scientific management might be best deemed successful "only through the *enforced* standardization of methods, *enforced* adoption of the best implements and working conditions, and *enforced* cooperation that this faster work can be assured. And the duty of enforcing the adoption of standards and enforcing this cooperation rests with the *management* alone" (emphasis his) (Taylor 41). Taylor devised a stratified system of management to assure the factory ran within the proper boundaries of the science resulting in an efficiently run factory. Rather than have one foreman oversee workers Taylor subdivided the work of a traditional foreman by creating eight functional foreman.¹⁰ Four of the foreman worked

on the floor training the workers and inspecting the final products and the machinery while four others worked in the planning department which oversaw the planning and running of the factory. The managers were responsible for measuring the output of workers, and they too were measured by the managers above them.

Taylor's systematization of scientific management can be condensed into the following principles and practices which the owner/superintendent was to implement:

- employ an efficiency expert to study the present operation and create a new and more efficient system;
- find the best way to accomplish each task and train workers to work in the most efficient and productive way;
- set specific standards for workers to meet within a specific amount of time and pay those who meet the standard a higher wage;
- design tools to measure workers progress;
- arrange managers (foremen) to oversee specific components of the process; and
- centralize the power and decision making of the factory to the planning room away from the labor.

Ultimately, these practices removed autonomy away from the workers and functional foreman and replaced it with a scientific system that standardized, measured, and accounted for worker output. Or as Taylor summarized the effect of scientific management on industry, “in the past the man has been first; in the future the system must be first” (Taylor iv).

While Taylor was zealous in his commitment to apply scientific management across social institutions, he also warned against only implementing the mechanisms of the system (time study, standardization of tools, cost systems, etc) rather than the philosophy of the guiding principles. “When, however, the elements of this mechanism, such as time study, functional foremanship, etc., are used with out being accompanied by the true philosophy of management, the results are in many cases disastrous” (68). Or put another way, the components of scientific management are incapable of achieving efficiency without the overarching ideology that must accompany industrial reform. While the principles are means toward efficiency, what is truly necessary is an ideological shift or as Taylor referred to it in his testimony before the Special Committee of the House of Representatives a “mental revolution:”

Now, in its essence, scientific management involves a complete mental revolution on the part of the workingman engaged in any particular establishment or industry—a complete mental revolution on the part of these men as to their duties toward their work, toward their fellow men, and toward their employers. And it involves the equally complete mental revolution on the part of those on the management’s side—the foreman, the superintendent, the owner of the business, the board of directors—a complete mental revolution on their part as to their duties toward their fellow workers in the management, toward their workmen, and toward all of their daily problems. And without this complete mental revolution on both sides scientific management does not exist. (27)

Despite Taylor's warning of the need for a mental revolution whereby workers and management labored in harmony for greater profit according to the laws of the science, contemporaries of Taylor and modern historians have illustrated that most industries used a piecemeal approach to Taylor's principles of scientific management. Robert Hoxie's 1914 publication of Scientific Management and Labor he examined 35 shops that implemented scientific management and found, "no single shop was found which could be said to represent fully and faithfully the Taylor system as presented in the treatise on 'Shop Management'...and no two shops were found in which identically or even approximately the same policies and methods were established and adhered to throughout" (cited in Nelson 69). Hoxie's study provides a glimpse of the inability to reform industry along the principles of scientific management. In Daniel Nelson's historical examination of twenty-nine plants that introduced scientific management between 1901 and 1917, he found if changes did occur they "rarely if ever entail[ed] a complete reorganization along the lines that Taylor recommended" (Kliebard 49). Most often Taylorizing included modifying:

a wide-ranging revision of the physical organization of the plant, a less thorough alteration of the foreman's functions, and a modest change in the average workman's activities. To describe scientific management in these plants as a 'partial solution to the labor problem' or a radical revision of the worker's role, as Taylor, several of his followers, and many academic writers have done, is both inappropriate and misleading. (Nelson 70)

What is clear from Hoxie's and Nelson's research is how the companies they studied implemented an incomplete or piecemeal version of scientific management with varying

success. While Taylor and some of his followers claimed a complete system of scientific management was possible to create and sustain, the mechanisms of scientific management rather than the mental revolution serve as the markers of Taylor's industrial reform put into practice.

Though a full implementation of scientific management was unsuccessful; it is important to recognize the ideology within the piecemeal version to establish a system created in the name of science and used to justify an efficient form of organizational management. Ultimately the goal of these mechanisms of scientific management focused on defining work efficiently, creating efficient workers, and sustaining an efficient organization. However, as studies of factories implementing scientific management have demonstrated, a piecemeal approach to achieving efficiency was taken over the purely ideological and systematic one Taylor advocated. Similarly, this fractured approach of implementing components of scientific management can be found within the discipline of English as reformers borrowed from Taylor components of scientific management to solve the "writing problem" and form more efficient schools.

The Creation of a Scientific Tools for Measuring Writing

Educational Measurement and Expert Culture

The problem of needing an objective and standardized process for measuring writing was met with excitement by members of the burgeoning field of educational measurement and concern by those who were skeptical of the "science" of measuring writing. After critiquing teachers inability to mark papers in a uniform manner, Thomas Briggs, professor of Education at Teachers College, stated, "Such facts made the scientist

eager to devise some means that would reduce the variation in marking the same paper by a number of different judges, and that would insure more consistency in the marks assigned by a single judge on different occasions” (424). I use this quote to point to Briggs’ overt statement that marking papers more consistently and efficiently was the problem of the scientist to solve—not the teacher or principal. The role of an educational scientist or efficiency experts served to relieve schools from waste and inefficiency but also brought about an expert culture tied to the field of educational measurement that sought ‘scientific’ solutions to educational problems.¹¹

Edward Thorndike, one of the primary players in forming the field of educational measurement, drew on positivist theories of measurement which shaped both the field of educational measurement and the teaching of writing. At a moment when psychology was attempting to establish itself as a field within the American university, Thorndike and members of the Teachers College at Columbia University were attempting to build a profession of credentialed educational administrators. Thus, Thorndike set out to show psychology was a true science and could be used in educational practice (Tomlinson 369). One of Thorndike most cited lines, “whatever exists, exists in some amount. To measure it is simply to know its varying amount” might summarize best his legacy in educational measurement (379). Thorndike worked to create educational tools, typically scales that could measure the incremental growth of students. Resisting vague units of measurements such as letter grades or percentile scoring Thorndike pressed for the need to definite units of measurement and instruments that could count or calibrate student learning, “when we measured length or weight or volume or temperature or electric potential, all competent persons measured the same thing. But when we measured

achievement in first-year Latin or college algebra, even the most competent twenty teachers measured twenty different composites” (372).

Critics of Thorndike today point to his belief that educational science was so powerful that only experts were able to decide what to teach, how to teach it and how it ought to be evaluated (Giboney 170). Furthermore, as Stephen Tomlinson points out, Thorndike (like Taylor) saw a “fundamental difference between the mind of the worker and that of the expert;” the worker was grounded in perception (and rule of thumb thinking) while the expert “was able to generate objective judgment based on facts” or science (371). Thorndike, however, provided the means to achieve the end goal of efficiency in schools as Tomlinson establishes, “Thorndike maintained that educational research must identify those methods that are most effective in bringing about the social goals of schooling. Such scientifically proven practices, when combined with a system of training and supervision, could then replace the traditional rule of thumb strategies employed by the average teacher” (371).¹²

The pursuit of a science of writing aimed to move writing away from human judgment and subjectivity and toward a form of objective measurement. There were two significant attempts to achieve the goal of a more efficient approach to measuring writing. The first, writing scales, were specifically intended for teachers to use in their classes in order to provide reliable scores and to eliminate teacher subjectivity. The second, standard tests, was more concerned with the efficiency of the school and looked to be able to compare students across schools, districts, and states. Both writing scales and standard tests substituted the measurement of writing away from teachers replacing it with a more objective form of measurement created by an expert.

Writing Scales

Thorndike and his followers designed scales to assist teachers in measuring students' abilities. A scale consisted of:

a series of objective forms of exercises or definite samples of products of different quality which, by means of...statistical procedure, have been arranged in a definite order of position, usually in ascending order of difficulty or merit. In a scale, each exercise or group represents as much greater or merit than an exercise or group just below it on the scale.

(Greene and Jorgensen cited in Love 2)

Scales contained samples of student work which were quantified, calibrated, and equated so that samples on the scale could represent a numerical score that was equated in proportion to the other examples on the scale. Thorndike's first scale on student handwriting in 1910 opened the floodgates for followers to experiment in designing scales in other disciplines. By 1916 scales existed to measure arithmetic, drawing, English composition, handwriting, language, reading, oral reading, silent reading, spelling, and vocabulary (Johnston).

In creating composition scales, Thomas Briggs emphasized the role of the scientist, "Before the scientist there were two possible modes of procedure. First, he could in theory analyze effective writing into its elements and count the improvement in spelling, punctuation, choice of words, sentence structure, and the like; or, second, he could provide a means of measuring the composition as a whole, considering the impression tout ensemble" (424). Briggs dismissed both modes in favor of a more scientific method where he paraphrases Edward Thorndike:

The scientist holds that whatever qualities exist, exist in some amount, and that whenever a difference in quality can be perceived, it can be measured...the problem of the scientist, then, was to provide a stabler set of standards ranging by known steps from zero, or 'just not any,' value up as high as needed—standards that would be the same to anyone at anytime at any place (424-25).

It was Milo Hillegas, a student of Thorndike's who created the first composition scale. Hillegas' scale, A Scale for the Measurement of Quality in English Composition by Young People, (I will refer to this as the Hillegas Scale from here on) consisted of fifteen samples of writing plotted on a numerical range from 0-1000.¹³ To find the fifteen samples that would become standard markers on the scale, Hillegas and an assistant read seven thousand compositions by young people and rated them into ten classes. From the ten classes, seventy-five samples were selected and eight samples were artificially included ranging from examples that were purposely written poorly by adults to writings of Jane Austen and the Brontes when they were in their youth to compositions of college freshman The seventy-five samples were scored by "competent judges"—authors, literary workers, and teachers. Hillegas then placed the ratings on a bell curve and used the theory of normal distribution to derive the statistical value of the scale. Hillegas' scale, provided a range of compositions whose numerical scores were proportionate to quality of the composition; thus, a composition with the score of 400 was twice as good as a composition with a score of 200 and four times better than a composition with the score of 100.¹⁴

In order for the scale to be truly standardized, it needed to be more than a ranking of composition—value needed to be added to the rank. Thus, the location of a zero point was established by which the rest of the samples could be judged. Hillegas asked twenty-eight specialists (teachers, authors, and psychologists) – different from the previous groups mentioned above – to rank the twenty-seven samples. The zero point was determined by a majority of the judges declaring it the worst sample of the bunch or stating that they “considered [it] to possess absolutely no merit as [a] specimen of English composition by young people”(16).¹⁵ Upon deriving the zero point for the scale, units could be created. A unit for Hillegas’ scale was defined as the “difference in quality which exactly seventy-five per cent of the judges observe” (20). In practice this meant that if seventy-five percent of the judges believed sample A was better than sample B, a measurable unit of deviation (quantity) had been established by which the quality of composition could be arranged on the scale. The unit was similar to a degree on a thermometer, an inch on a ruler or a second on a watch providing an objective standard by which writing could be consistently measured.

The first response to the Hillegas scale came in 1914, two years after the creation of the Hillegas Scale, when Frank Ballou, the Director of the Department of Educational Investigation and Measurement of the Boston Public Schools, published a “Scale for the Measurement of English Composition” in the Harvard-Newton Bulletins, which became known as the Harvard-Newton Scales. Ballou established six principles that writing scales should be based on. Here is a brief summary:

- A composition scale should not attempt to measure too complex of a product; rather different scales need to be established for different modes of composition.
- A composition scale should consist of compositions that have the same characteristics as the composition which would be measured against it. Eighth grade compositions need to be compared to eighth grade compositions.
- A composition scale should conform to the practice of rating schoolwork or correspond to a letter and numerical grade.
- The combined judgment which created the consensus should be people most familiar with this type of work—eighth grade teachers.
- Both merits and defects must be considered in establishing the quality of a composition. An explanation for the user of the scale should be provided so they understand what the readers saw as the merits and defects of each composition.
- Finally, “teachers must accept a scale as a standard, in the same sense that they accept other standards...If a scale is to become a real standard, it must be accepted as such and used in the true sense of a standard” (9).

In its final form, the Harvard Newton-Scale consisted of four separate scales covering description, exposition, argument, and narration. Each composition was about a page long, written by an eighth grade student who was given an hour to write and another half hour for “self-correction.” This assured that the compositions were school exercises and appropriate for comparing to other eighth-grade writing. As part of the scale, three

paragraphs accompanied each sample that addressed “why each composition was better than the one below it in the scale: why it was poorer than the one above it in the scale, and also its own intrinsic worth” (47). The paragraphs were included to further help teachers using the scales determine the strengths and weaknesses of a composition.

Hillegas and Ballou were the first to attempt creating objective scales by which composition could be measured, but their work in this area was not fully embraced and quickly came under fire. First, the Hillegas Scale lacked a definition of merit. As Hillegas acknowledged, “no attempt has been made in this study to define merit. The term as here used means just that quality which competent persons commonly consider as merit, and the scale measures just this quality” (13). The method of applying a statistical analysis to the opinions of specialists served as the largest critique leveled against the Hillegas Scale and Harvard Newton Scales. Isidore Kayfetz, who conducted critical studies of both the Hillegas and Harvard Newton Scales wrote, “The scale was derived by a statistical treatment applied to opinions. The data were therefore subjective rather than objective...A scale based upon the average opinion of a large number of judges is not, strictly speaking, an objective scale” (Kayfetz 570). Furthermore James Drever referring to Thorndike’s original scale for measuring writing (which preceded Hillegas’) claimed:

Now the fundamental objection to such a scale is that it can never be in any real sense objective, nor can its use by any individual give an objective determination of the merit or value only specimen of writing.

The scale is not objective, because it is simply the average of a number of individual opinions of merit, a composite portrait of a number of subjective opinions...(cited in Kayfetz 570)

and ultimately concluded the Hillegas Scale to be “scientifically unsatisfactory.”

However, Kayfetz did praise Ballou for using writing samples that were similar to the ones student who would be measured against the scale would create, as well as creating separate scales for the various modes of writing. Kayfetz saw these revisions as attempting to align pedagogical standards with measurement.

The Hillegas and Harvard Newton Scales provided the model which other researchers attempted to revise by incorporating different components of writing or improving the technical precision of the scales. Edward Thorndike’s “Supplement to the Hillegas Scale” substituted samples of the original Hillegas Scale and improved the range of examples in the middle of the scale. The Nassau County Supplement to the Hillegas Scale was based on elementary narratives to the same prompt, and used longer pieces of writing than the original Hillegas Scale. The Breed and Frostic Scale gave sixth grade students twenty minutes to create the end to a story that was read out-loud in class. The Willing Scale separated the merit of composition into content and mechanics moving the scales away from a single quality of “merit”. The Van Wagenen Scale, divided merit into three qualities including: thought content, sentence and paragraph structure, and mechanical errors.¹⁶

Composition scale served as the first scientific tool to measure writing. Drawing on positivist theories of psychometrics premised upon the belief that everything could be quantified, composition scales were constructed using a process of consensus. The purpose of the scales was for teachers to compare their students’ work by placing it beside the samples on the scale, determining which sample it most nearly corresponded, then assigning it a numerical value which could be compared with other student writing

in the school or against an earlier piece of composition. Rule-of-thumb methods were replaced by a scientifically studied method. The implementation of writing scales served to eliminate teachers' subjective judgment of a student text and replace it with a range of scientifically determined scores.

Standard Tests

Standard tests or "school efficiency tests" served as the next step in creating a more scientific approach to education and the teaching of writing (Kayfetz 325). By testing all of the students within a school and measuring each student against a scale, proponents of standard tests argued the efficiency of a school could be determined. Standard tests differed from writing scales in that the scales were to serve as a means of standardizing classroom judgment of teachers, while the standard tests allowed for scores to be aggregated and generalizations to be made about schools. In 1923, the National Council of Teachers of English (NCTE) ran a series of articles authored by C.C. Certain, the Chairman of NCTE's Committee on Examinations. The goal of the new committee was two-fold: first to "stimulate an interest" in using standard tests in the discipline of English; and second, for schools across the country to "include standard tests in their schedules for the term beginning September, 1923" (Certain "Are" 365). Certain argued the benefit of using standard tests over traditional exams was that they could be used for both prognosis and diagnosis assuming that a schedule is "definitely planned and systematically adhered to" ("Why" 464).

Standard tests standardized the conditions under which students were asked to write. By allotting forty-five minutes for students to complete a narrative or expository essay, and providing a uniform writing prompt for students to respond to, the standard

test provided a basis for a comparison of student work across classes, schools, districts, and states. In addition to standardizing the conditions under which students were tested, the scoring procedure of scoring students' tests served to assure reliable results in scoring student tests. Scorers of standard tests, typically professors and graduate students compared the students' tests to a composition scale (typically the Hillegas, Harvard Newton, and the Nassau County). In the September article, Certain, speaking for NCTE, argued that teachers ought to become more involved in standard tests and should learn to score students' papers. Part of the scoring session would attempt to norm teachers' judgment to the scales in order to achieve reliable scoring of student essays.

[T]he leader should cut up several mimeographed copies of the 'Nassau County Supplement to the Hillegas Scale'. [...] The leader should hand the specimen compositions, two at a time, to each teacher who should decide which of the two compositions is the better. As additional compositions are given out, two at a time, the teacher should place them where they belong in relation to each other. Finally, the teacher should arrange the compositions from the highest to the lowest in the order of general quality. [...] The teacher who succeeds best in this exercise will, of course, reconstruct the scale. The exercise should be repeated until each teacher succeeds in placing the compositions in the order that they occur on the scale. ("Why" 470).

Once the teacher/scorer had the texts in the correct order they were to assign each a numerical score and check it against the Nassau County Supplement to the Hillegas Scale. After this exercise the leader would give each teacher a set of compositions which

they were to rank and rate. Certain advocated that teachers practice scoring at least three sets of examples prior to starting on the real students' compositions. And a teacher's scoring was not "satisfactory until he [brought] his average to .5 or lower" (472).¹⁷ Certain argued that scoring student papers was a skill that even experts must practice. "No one should be satisfied with merely 'learning to use the scales.' *Skill* is what is required and occasional practice is necessary for this" (473).

In addition to teachers scoring the "general quality" of student compositions, NCTE encouraged teachers to tabulate the errors in spelling and one of the following categories: handwriting, punctuation, grammar, and rhetoric. To aid in the tabulation of errors, NCTE offered check sheets for diagnosis of written English. At the end of the term teachers were to give a second exam and again tabulate errors in the same category as they did the first time. They could then compare the results of the two tests. NCTE provided a template for teachers to keep track of the results of each class. Teachers were to record the total number of words written by the class, sentence errors by the class, errors in correctness of form, errors in grammar, and the total number of errors by the class in grammar, per hundred words (477). The errors were compiled and teachers were able to compare their different classes and to create competition between them.

The NCTE Committee on Evaluation provided specific suggestions to teachers using standard tests in their classes. They focused on creating occasions to discuss the results of the standard tests with students. First, the results should be shared with the class and variable attainments within the class discussed, and comparisons should be made with schools in other cities allowing students to see how they compared. A plan should then be implemented on remedying the weaknesses of the class. In addition to a

conversation with the entire class, conferences with individual students were necessary. As a part of the individual conferences with students, teachers could diagnose students' difficulties and devise a systematic drills and exercise plan in order to help remediate the student and their writing. The Committee on Evaluation also suggested teachers schedule times throughout the term where informal tests could be given and finally, that teachers should have a publicity program to report their class results either through bulletins to the class, other teachers, or even by placing articles in the school magazine or newspaper

The use of the standard test to measure the value of a composition coupled with the calculation and recording of errors placed teachers as managers of student writing within the classroom, while also being managed by their administrators who were particularly interested in seeing comparing their schools results to others. As seen from the "suggestions" of NCTE, practicing standard tests was recommended as part of the writing curriculum and classroom activities of teachers. Obviously the more students were exposed to the standard tests the better the chance that they (students and teachers) would have higher scores. Thus, the science of writing which began with the hopes of eliminating teachers' individual judgment and creating a scientific tool to measure compositions ended with teachers matching student writing to a scale of examples and tabulating errors which attempted to improve student scores on standard tests.

The science applied to composition scales and standard tests was a process of finding consensus among experts who determined a text's worth. Units were established which enabled early proponents of educational measurement to argue that the scales were more objective than teacher judgment because they provided a means to quantify writing. However, what was being measured with the tools shifted between the scales and the

standard tests. The scales were originally created to serve teachers and provide a uniform and consistent tool to measure student writing, standard tests shifted the focus from the classroom to the overall efficiency of the school. This shift illustrates the close relationship between the science of writing assessment and the ways the tools helped administrators manage their schools.

Managing Writing, Improving Writing

In Taylor's explanation of scientific management, his goal of creating an efficient system based on the people working efficiency within the system overlooks the quality of the product being created and the actual labor workers engage in. Whether he was working with iron workers, coal shovelers, or brick layers, the question for Taylor was never about the quality of the brick wall being constructed, but only in finding the most efficient way to produce the product. Similarly, as the piecemeal ideology of Taylorism crept into Progressive Era discussions on solving the "writing problem" tools were created and "scientific" units of measurement determined to measure writing, but the focus was on the efficiency and productivity of schools and teachers not on the act and product of writing. That is pedagogical questions, or questions about student improvement in writing were not considered by advocates of the writing scales and standard test. Rather, the goal of Taylorism was to keep the people within the system working in accordance to efficient practices. Therefore, the focus was removed from the students and their writing and the placed on the teachers and their ability to get students up to a standard or on the principal's ability to get his teachers up to an appropriate level within the city or state. Focusing on systematic efficiency instead of the actual labor or

quality of the product put questions about management before questions about writing and students.

In the introduction to his writing scale, Milo Hillegas emphasizes the utility of the writing scale as a tool to help administrators measure and manage writing in order to determine the efficiency of a school system as the new tool would allow “educational administrators and investigators {to} measure and express the efficiency of a school system in terms that would carry conviction” (2). Shifting the unit of analysis from measuring writing to measuring the productivity of a school system and teachers became a common use of the scales as administrators contributed essays in English Journal discussion the ways the data gleaned from the scales and standard tests allowed administrators to remove the guesswork from reviewing their schools and teachers.

Charles Gunther, the Principal of School No. 2 in Poughkeepsie, New York reported how through an experiment he was able to scientifically study and measure teachers’ use of the Hillegas Scale. In “My experience with the Hillegas Scale” Gunther stated, “if the supervisor in English uses the Hillegas scale or some similar scale, he can supervise more intelligently the work of his teachers. No longer need he rely on vague opinions” (Gunther 535). Through his three experiments Gunther was able to show that the Hillegas scale was a more reliable tool than percentage grading and that it could be used equally as well across different schools. Yet he did not stop there. Commenting as a principal, he discussed the importance of the Hillegas scale for administrators. “A principal can [...] determine by means of the Hillegas Scale whether the teacher is a careful marker or not” (540-541). The rating of themes was not easily observed through classroom observations, yet having teachers submit their students’ scores, the

administrator could get information that he did not get from visiting classrooms. Furthermore, the teacher's use of the Hillegas Scale could be discussed during a conference with a teacher because "a principal would have something tangible to talk about in his private conference with his teacher. His supervision would be intelligent" (540).

S. A. Courtis, the Supervisor of Educational Research in Detroit Michigan, articulated the importance for administrators to have a standardized tool within their school so they could measure the efficiency and effectiveness of their teachers. For Courtis, one of the central misunderstandings of teachers was the differing roles of administrators and teachers within a school. "Supervision and teaching give rise to two such different points of view. Supervision is concerned with the efficiency of the general process, teaching with the development of the individual child" (Courtis 209). According to Courtis teachers are not concerned with comparing school systems nor determining the efficiency of different teaching methods because those issues are administrative and supervisory issues; rather, teachers are only concerned with students. This divide on perspective caused administrators like Courtis to use writing scales as a way to measure efficiency within their schools. "My interest in tests lies in the fact that by means of measurement alone do I see any hope of so remedying present inefficiency and waste that the products of our teaching effort shall be ideals, inspiration, and character, instead of a few morsels of knowledge imperfectly developed" (209) Therefore, Courtis advocated the use of the Hillegas scale to determine the efficiency of different modes of teaching.

Courtis provided a blueprint on how he saw the Hillegas Scale being implemented into the school, and it is worth quoting at length:

...every teacher, without exception, should test and mark by the Hillegas scale the ability of her pupils when she *receives* them and when she *sends them on* to the next teacher. In this marking, which you will notice is not for the student's benefit but for the teacher's, no consideration should be entered but one, the merit of the compositions as compositions. I believe every teacher should compare both the amount of the change she has produced during the term with the standard change produced by other teachers of her grade for the city as a whole, and the final ability of her class within the standard for the city as a whole. I believe, further, that when a teacher finds by such measurement that she is getting year after year less results than other teachers, she should of her own initiative seek assistance from the supervisor of English and try to bring her work up to standard (emphasis his) (Courtis 214-215).

The notion that the Hillegas Scale benefits teachers because it allows them to measure students' progress is quickly turned into a competition and form of supervision and punishment. The need to compare teacher effectiveness under the mask of students' scores is created and as Courtis points out, the use of the scales also allows administrators to compare teachers in their school buildings and across the city. But as Courtis hints, if teachers are not producing appropriate results they need to be assisted by a supervisor who will help to standardize them to get appropriate results.

By requiring his teachers to use the Hillegas Scale, Courtis was able to measure their performance through the increase or decrease of student writing scores. Courtis viewed this type of management as providing freedom for teachers to 'express their

personality, and yet keep them under such control that they development of the children will not be wrecked” (215). Curtis advocated setting a standard on the Hillegas Scale as a graduation requirement or end goal. If teachers could get their students to meet the standard they could teach their students how they wish; however, if the teacher could not get her students to reach that goal she risked losing her job.

To say that no student has completed twelfth-grade English composition until under given conditions he can write as well as quality 75 on the Hillegas scale is to give definiteness to the goal and to set the teacher free. She may work in any way she pleases, use any material and any text. The one requirement is that when she has finished, the desired goal has been attained. On the other hand, if the goal is not attained, both the teacher and the supervisor know it and can act accordingly. Certainly if the teacher’s work does not produce results she needs to change her method, and if after trial she proves incapable of profiting by training she needs to be eliminated as a teacher of English composition (215-216).

Curtis illustrates how administrators saw the potential to use tools such as scales and standard tests to manage their teachers and schools. Using the data produced by the measurement of student writing, Curtis argued that teachers’ efficiency could be measured and decisions rendered about their efficiency could be made by administrators.

Writing scales originally attempted to remove the ‘vague opinions’ from teachers, yet Gunther showed how the supervision of teachers could become more efficient through the implementation of scales and Curtis illustrated how the scales could be used to measure teacher efficiency. A crucial similarity between Curtis and Gunther

illustrate how composition scales were used as tools to manage teachers and schools. Through the collection of students' composition scores, administrators could track the success of each of their teachers and compare their individual teacher and school results to other teachers and schools. From an administrative point of view, the scales defined clear standards of writing and teachers needed to adjust their classrooms and pedagogy to assure student success as defined by the composition scales.

While Curtis and Gunther provide an insight into the mindset of administrators who were interested in managing their schools along more efficient and scientific principles, there was also criticism of the role scales and standard tests played in within schools and the effects these new technologies had on writing, students, and teachers. The main critique toward was that the scales provided a rigid view of writing only concerned with measuring the product, and did not concern itself with issues of pedagogy and student improvement and development.

Fred Scott, in his 1913 President's address to NCTE, titled "Our Problems" discussed his concern of the measurement movement and its encroachment on teaching English. "We must believe that whenever a piece of scientific machinery is allowed to take the place of teaching...the result will be to artificialize the course of instruction" (Scott 4). Scott's concern revealed how measurement for systematic purposes often came at the cost of inquiring into the development of students:

It is surely possible to draw a distinction between a system of measurement which grades a composition for the sake of the grading, that is, for scientific or administrative ends, and a system which evaluates it as a stage in the pupil's progress. To illustrate my point: Suppose that

instead of asking, Is this composition, written by some unknown X, better than that, written by some unknown Y? we ask, Which is the more sincere expression of some growing individuality? Or, which will be more legitimately effective in its appeal to a certain audience? (Scott 5)

In Scott's critique of the systemic approach of writing reform, he asks different questions about student writing than the scales and administrators like Courtis were focused on. Scott's concern resists the administrative need to measure writing for the sake of scientific systemization and reflects a pedagogical concern of considering the rhetorical purpose of writing or development of the writer. Isidore Kayfetz, cited above for his critique of the "science" of writing scales, argued that the measurement of student writing by use of writing scales, "does not deal with the process. It does not concern itself with the question—'how did the child write the composition?'" (Kayfetz 574). Like Scott, Kayfetz offers questions about writing, which the simple measurement of a writing scale or standard test does not address. Kayfetz's concern is better understanding the student's process of writing as well as the thinking and strategies a student uses to compose.

Flora E. Parker, an English teacher at Western High School in Detroit, worked under S.A. Courtis and critiqued the ability of the scales to measure all components of composition. Parker acknowledged that grammar, punctuation, diction, etc. could be measured with a scientific tool; however, composition is not just mechanical but also artistic and the 'intangible graces and beauties which reside in that realm' can not be measured with any form of science" (Parker 204). Parker goes on to argue that scales strip the personality out of writing and in a large school system like Detroit, students need to be able to express their personality. Furthermore, she asserts that students are human

and therefore cannot be treated with “uniformity as the supreme good” (204). Finally, Parker argues teachers grading differently, rather than conforming their judgment to a scale, is a positive experience for students, exposing them to teachers and readers who have different “convictions” about writing.

The voices of teachers and researchers like Parker, Scott, and Kayfetz were in the minority, in terms of the number of articles and pages published in Progressive Era journals. However, the presence of their individual and collective critiques of scales and standard tests as a means by which administrators could efficiently manage their schools demonstrates resistance to these practices. Their concerns over improving students as writers were not the concerns of administrators and efficiency experts who sought to efficiently manage schools and teachers.

Solving the “Writing Problem” with Scientific Management

Solving the “writing problem” through the creation of tools to objectively measure writing so that administrators could managed teachers based on the data gathered from the tools illustrates the way principles of scientific management crept into the discipline of English. As I stated earlier, contemporaries of Taylor and historians have exposed the difficulty of implementing a full systematic overhaul within factories resulting in a fragmented or piecemeal version of scientific management. The mental revolution, Taylor insisted never came to fruition resulting in ideologically charged beliefs and practice around the relationship of measurement and management that can also be seen in the solution to the “writing problem.”

The fragments of scientific management emerge in the solution of the “writing problem” where efficiency experts were consulted or brought into school systems to diagnose problems and provide solutions which would result in a more efficiently run school. In the case of the “writing problem,” the experts attempted to remove rule of thumb methods of grading and marking practices of the teachers through the adoption of scales and then standard tests. These new tools were considered by their proponents more “objective” than teachers judgments because they were constructed along scientific theories and could result in accurate measurement of students’ writing. The new tools also provided “objective” standards for students to be measured against and moreover, worked to define the kinds of writing students should be doing in schools: narrowing writing to the traditional modes which served as the foundation and definition of writing measured on many of the scales and standard tests. However, the new tools were seen by administrators not as tools for improving student writing but as useful for measuring and managing the school system they oversaw. This led administrators to use the scales to measure teacher efficiency and compare the productivity of their schools to others in the city and state.

The ideology of scientific management provided the solution the “writing problem” because embedded in the definition of the writing problem was the solution. By framing the problem under the umbrella of efficiency—as a problem with teacher subjectivity, an economic and mechanical labor issue, and then finally as one of a social productivity—reformers sought a solution that could address all of these issues of efficiency. The only way all four concerns around efficiency could be remedied was through a systematic solution. As Samuel Haber wrote, it was not until Frederick

Taylor's scientific management that all four definitions of efficiency could be reconciled simultaneously (x). Thus, administratively minded educational reformers turned to the principles of scientific management, as many other social and industrial reformers did to create a more efficient system.

The systemization of schools and districts by using standardized tools to solve writing problems is a predominant strategy of reform. In chapter two I explore how a contemporary school district attempted to solve their writing problem through the creation of a standardized writing test and a system of student placement. In presenting this district's effort to reform their "writing problem," I will also trace the residue of Progressive Era reform and the scientific management of writing that still lingers in contemporary practices of accountability and writing assessment.

A District without Schools

Sitting across the table were three district administrators: Mark, the coordinator of the district's Writing Graduation Exam; Mary, the coordinator of High School Language Arts; and Cindy, the coordinator of Middle School Language Arts. I introduced myself to the three, and told them about my background as a high school English teacher and how my teaching experiences informed my work in graduate school. There was no planned agenda for the meeting. I wanted to get some insight into BPS's writing program and the different schools within the district that might be good sites for research. By bringing my interests in writing assessment together with their writing program and schools, I hoped that we might develop some ideas, issues, or themes around which I could build a study for my dissertation.

It was about twenty minutes into the meeting when Mary said something so perplexing, that I couldn't get past it.

"So what do you want to study?" Mary asked.

"Well, I'm interested in how teaching and assessing of writing occurs within different school cultures. I'd like to study at least two different high schools to understand the issues that emerge around teaching and assessing writing within different contexts."

"Are you wanting to conduct a comparative study?" Mary's eyes revealed the correct reply to that question was no.

"Well not in terms of saying one school does something better or worse than the other, that isn't the point, nor is it very interesting to me. What I do want to understand is how different school cultures are formed around teaching and assessing writing, what

issues they run up against, and how teachers in different schools think about the teaching and assessing of writing within their schools,” I said. I saw Mark’s head nod. His eyes squinted before he began to respond, but Mary cut him off.

“Well, something you need to know about BPS is we don’t have schools. We only have teachers and the district,” she said matter-of-factly. “I don’t think looking at the schools would be very useful to you because we want our information whether it be about curriculum or instruction to go directly from here to the teachers. I’m not sure you would find the kinds of differences you think exist at the school level.”

I didn’t know how to respond to her comment. She couldn’t deny school cultures differ between buildings and staffs? She couldn’t really be serious that schools don’t exist here in Butler?

“But not all of your high schools are the same, right?” I paused to see if there would be a response. There wasn’t, and so I continued. “I mean, Marshall High, according to their website, has over twenty languages spoken in the school. I would think that would make a difference in the way the teachers think about how they teach and assess their students compared to schools who don’t have a diverse student body.”

Mary reassured me that several years ago the district introduced a district writing assessment, which they believe is more stringent than the state’s writing assessment. All students who attend BPS must pass the exam in order to graduate. The district administers the exam so all teachers know what they need to teach students and how they will be assessed. She also recommended that if I was committed to examining school cultures that I might work with smaller school districts outside of Butler where there might be a more significant contrast between schools.

The meeting went on for another fifteen minutes or so. When I left the room, Mark walked me to the front door and assured me that we could probably work something out and told me to keep in contact with him. Walking back to my car, I remained floored by the “we don’t have schools” comment. Of course you have schools; there are buildings all over the city where teachers and students meet daily. How could she not acknowledge that there might be different cultures within them and that those cultures might approach writing and assessment differently?

As I drove home, I tried to see it from Mary’s side. In a district with seven high schools with, to a certain extent, open enrollment I could imagine why she would want to convey that all students in the city receive the same education and same opportunities. It also made sense that she would want to present the district as having a centralized goal all teachers work toward. And I could also imagine, from a public relations perspective, why she wouldn’t want a graduate student from the university to conduct a study that might disrupt the notion that all the schools within the district are not the same.

Before my next meeting with BPS, Mary had retired and I continued to work with Mark on the scope of the study. However, after my first round of interviews with teachers, principals and district administrators, I found that in many ways Mary was right. Teachers often thought of themselves as part of the district before they claimed identity as a teacher in a particular school. It was also true that writing assessment was considered a district matter and teachers hardly mentioned their specific school context unless I specifically asked about it. After listening through the first round of interviews, the idea of a district without schools made some sense, but also made me ask more questions. What does it mean to ignore the context in which learning takes place? How

does a systematic approach to teaching and assessing writing impact teachers and students? And how does the creation and implementation of a policy like the Writing Graduation Exam impact the way people within the district think about their work teaching writing?

* * *

Chapter 2

Creating a System of Writing:

The (his)Story and Counter-narratives of the Writing Graduation Exam

We are one of the very few districts that give an exit exam; if you don't meet the graduation exam, you don't graduate. It doesn't matter how much seat time you've put in, how many credit hours you've put in, how many credit hours you have, [or] what your GPA was. If you can't pass this test you simply cannot graduate from Butler Public Schools.

Rebecca, BPS Writing Graduation Exam Coordinator

It's not that I feel ambivalent about the district's WGE testing. I think that there's a value. I just worry about how agendas start to enter in... One thing I appreciate is having a wide range of thoughts, not just one school of thought. I think it keeps the process honest. Does that make sense? To have challenges and embrace them. As an educator, I want room for that kind of talk. Or a dissenting voice to always keep you honest, to keep the process on the up and up.

Carlee, Wilson High School Teacher

Butler Public School District's Writing Graduation Exam (WGE) was formally created as a policy in 1998 and as a practice during the 1999-2000 school year.

Conducting a study seven years later made it difficult to understand the impetus for creating the policy and how the exam was implemented across the district. Moreover, as I talked with teachers, principals and district administrators about the creation of the WGE, depending on their position they offered different stories to explain the origins of the WGE and its impact on the teaching of writing within the district. The stories overlapped and contradicted each other, and not surprisingly pointed to the complex relationship teachers and administrators have with writing and each other.

Rebecca's quote above might best summarize the official story: the narrative told by the district of the problem they had prior to the WGE, the policy which attempted to reform the teaching of writing within the district, the ways the WGE was implemented, and the problems it solved. This is an institutional narrative of reform where a problem was named, a solution created and actions taken to improve the district. On the surface, the official story of the WGE may seem like a coherent narrative of accountability and reform; however, the teachers and administrators who told it also shared counter-narratives of the same WGE system. Carlee's quote above invites an alternative telling of the official story, one that addresses agendas, challenges, and dissent. These counter-narratives not only disrupt and complicate the official story, they also illuminate the fissures within the district, providing a more complex telling of how the system impacts the people working within the district.

The relationship between the official story and counter-narratives may be described as the intended and unintended consequences of implementing a systematic approach to accountability and writing reform. Through a representative telling compiled from interviews with teachers, I will offer four pairings of narratives/counter-narratives intended to mirror what the district thought would happen when enacting the WGE and what teachers actually experience working within the system. Illuminating the fissures existing at various sites and levels of the WGE system provides the possibility of opening spaces for teachers to share counter-narratives—narratives of resistance through which their subjugated knowledges are validated. Prior to exploring the impetus for creating the WGE and the narratives and counter-narratives teachers and administrators told about the WGE, I will overview Michael Foucault's discussion on subjugated knowledges to frame

the counter-narratives and the resistance teachers exhibit when discussing the WGE. The work of juxtaposing the two forms of narratives provides a telling of the WGE that neither fully celebrates nor condemns its implementation. Rather, it illuminates the structures and limitations of implementing a district-wide system of reform based on objectively measuring writing, and managing people within schools based on the data from the tools. While the official story of the WGE demonstrates how the WGE solved some of the district's "writing problems," the counter-narratives force us to address the new "writing problems" which exist in the wake of the WGE's call for accountability.

Subjugated Knowledges

In the first of Foucault's "Two Lectures," in Power/Knowledge, he discusses the "local character of criticism" which serves as an "autonomous, non-centralised kind of theoretical production, one [...] whose validity is not dependent on the approval of the established regimes of thought" (81). By its nature, the local character of criticism works through an insurrection of subjugated knowledges. The "official story" of the WGE represents an "established regime of thought" both in the local narratives told by employees of BPS as well as the ways test-based accountability, the systemization of schools, and tracking of students have become common sense approaches to running schools in America over the past century of school reforms.¹⁸ However, the streamline narratives of efficiency and accountability often come from outside the daily lived experiences of teachers and students inside schools. Therefore, the search for subjugated knowledges, or stories which reveal the knowledge of the people inside the system, is useful for critiquing the system and the consequences of its implementation.

Subjugated knowledges serve as a site of resistance to a hierarchical and/or authoritarian power structure. In an interview, “Power and Strategies,” Foucault argues “there are no relations of power without resistances” (142). In fact, resistance to power should not be viewed as separate from power. For Foucault, power is “‘always already there,’ one is never ‘outside’ it, there are no ‘margins’ for those who break with the system to gambol in” (141). If power is all-encompassing, resistances are effective because “they are formed right at the point where relations of power are exercised...[resistance] exists all the more by being in the same place as power; hence, like power, resistance is multiple and can be integrated in global strategies” (142).

A tool in his genealogical methodology, Foucault explored subjugated knowledges in order to expose the “historical knowledge of struggles” (83). Foucault defines subjugated knowledges as two-fold:

on the one hand, I am referring to the historical contents that have been buried and disguised in a functionalist coherence or formal systemization. [...] On the other hand [I mean] a whole set of knowledges that have been disqualified as inadequate to their task or insufficiently elaborated: naïve knowledges located low down on the hierarchy, beneath the required level of cognition or scientificity. (82)

Foucault advocated listening to the local subjugated knowledges against a unified theory or formal systemization which assigns value and rank based on its relation to disciplinary knowledge, in order to discover the “ruptural effects of conflict and struggle” which are overlooked (and even purposely ignored) by a coherent systemization of knowledge/power. For Foucault, it is through the reappearance of subjugated

knowledges “that criticism performs its work” (82). In this chapter, I am particularly interested in listening to the knowledges of teachers and administrators which have been “insufficiently elaborated” in order to provide a space for elaboration and representing the knowledges of those inside the system which get lost in the “official story” of the WGE. Placing the official story next to the counter-narratives magnifies the fissures apparent in teachers’ experience of working within the WGE system.

The Problem and the Policy

Similar to Progressive Era writing reform, the way Butler Public Schools defined their “writing problem” is indicative of the paradigm through which the solution would be sought. Throughout interviews, it became clear that BPS district officials and school board members were responding to the perceived creep of the accountability agenda on their district. These concerns lead to a top-down approach of test-based school reform, with writing becoming a key component of the tests. While there were “writing problems” occurring within schools and classrooms prior to the adoption of the WGE policy, participants only discussed those in retrospect as a way to illustrate how the WGE helped writing teachers.

When teachers and administrators discussed the impetus behind the WGE, they immediately pointed outside of their classrooms and schools to the district and the school board. “As far as the writing exam, this was not a grassroots effort from teachers saying we need to have a formal writing assessment, but rather [from] the district level. Where exactly, I don’t know,” one teacher pointedly stated. The creation of the WGE policy and examination indicate the relationship and responsibility BPS has to the community.

Participants in the study most often explained the need for the WGE to assure high standards, provide an equal education for all students, and guarantee a student's diploma meant something in relation to their skills and abilities, not merely the amount of seat-time they accumulated. In the name of high standards, equal education, and student readiness for life after public schools, a political promise was offered to the community of Butler by the school board through a policy—their students would be proficient writers when they graduated. BPS's implementation of a standardized test to alleviate concerns around public education mirrors national political discussions where standardized tests serve as accountability mechanisms for public schools.¹⁹ With political dialogues focusing on the use of standardized tests to ensure standards and equity in education, the leaders of BPS, like many large districts across the nation, turned toward graduation exams to accomplish these goals.²⁰

To stay in front of the national curve, BPS district administrators sought to create graduation exams within their district. One teacher explains the movement toward test-based accountability this way:

I think the district people, and that would be from probably [the Assistant Superintendent] on down began to see a trend nationally toward testing, and my understanding is that Butler Public Schools got a jump on that and designed their own teacher-initiated criteria and testing formats. I think that's how it happened; they saw the trend coming and wanted to get ahead of that wave to set up their own things. [...] And then it was up to the curriculum powers that be to begin looking at a formalized testing and assessment program across the district.

Rather than turning to a corporate test or having the state impose a test on the district, BPS wanted control over the test that would be used to measure writing, students, and the district. Feeling the creep of the accountability agenda, district administrators pushed for the district to create assessment tools rather than be forced to use a test that would not reflect the values and standards of Butler Public Schools.

The solution to this problem was formally addressed through the creation of the policy calling for graduation demonstration exams in the skills/disciplines of math, reading and writing. The policy is quoted directly:

INSTRUCTIONAL PROGRAM

Requirements for Graduation, Beginning with Class of 2003

Part 2: Required Areas of Study

In addition to accumulating the required number of credit hours, students will also be required to demonstrate that they have necessary math, reading, and writing skills. Demonstration and standards for successful completion of these skill areas will be developed. Demonstrations may be completed in a particular class and/or independent of a class, depending upon how the demonstration is developed. Demonstrations will not grant credit.

Area	12th Grade Graduation Demonstration
Mathematics	Students must demonstrate the knowledge and skills necessary to apply basic computational skills to problem-solving

situations, numeracy, statistics, correlation and cause/effect, recognize how numbers are used to influence decisions, AND demonstrate the skills of first-year algebra.

Reading

Students must demonstrate the ability to read and comprehend newspapers such as the *Butler Tribune* and the *Valley Register*.

Writing

Students must demonstrate the knowledge and skills necessary to write consistently at a standard still to be determined.

Date of Adoption (or last revision)

8-25-98

Compared to the desired skill sets in math and reading, it is clear the policy creators struggled to describe what they wanted students to do as writers or what would be deemed proficient for writing. The school board policy maintains students “must demonstrate the knowledge and skills to write consistently,” but similar to the Progressive Era “writing problem,” the knowledge and skill of writing are not defined as part of the policy. The lack of a definition of writing—what it is, what it looks like or could look like, what kinds of writing students should produce, and whether writing is a skill or a process—led to an opaque mandate by the school board. In an interview, Mark Thurman, the Language Arts Curriculum Coordinator discussed how the lack of a definition of writing made the creation of a demonstration difficult:

It was a board mandate and the interesting thing about that was when the board did this [they] jumped right on what students should have in terms of math and reading skills...but they were pretty unclear about what students should do with writing. It's just that we want our kids to be good writers, so figure that one out.

The WGE is derived out of an empty policy. Empty because the policy lacks a definition, purpose, goal, or standard for writing, but rather calls for the creation of a “demonstration” (not necessarily a test) to assure students can write “consistently” (undefined) before graduation. Yet eight years after the policy was written, the WGE is much more than a standardized test students must pass to graduate. Rather, the WGE is best explained as a system of accountability permeating the district, managing teachers, students, and writing through the application of objective tools to measure student writing and organize student and teachers within the district according to the results.

In the following sections, I expose four sets of narratives and counter-narratives teachers and administrators told about the WGE to expose the architecture of the district's system and the impacts of its implementation within the district. I have attempted to match these counter-narratives to the official story not just to establish a contradiction or point out two sides to each issue. Rather, coupling the official story with counter-narratives: illuminates the way systematic power works to manage students, teachers and administrators in BPS; uncovers the intentional and unintentional consequences of adopting and implementing the WGE system; and demonstrates teachers pushing back against the predominant narrative of the WGE and the ways they experience the WGE in their schools and classrooms.

The Architecture of the System: Narratives and Counter-Narratives

Resolving the District's "Writing Problems"

As a result of implementing a standardized writing test and having courses designed around students' writing needs, teachers discussed the ways the WGE helped to alleviate "writing problems" that existed within their buildings and in the district. These internal problems (I identify internal problems as those within the schools and district whereas external problems point to larger forces of accountability discussed above) around teaching writing were told through before and after narratives, where teachers highlighted difficulties of teaching writing prior to the district's implementation of the WGE and the ways the WGE had resolved the problems. What is unclear from my interviews is if teachers articulated these "problems" prior to the WGE or if through the implementation of the WGE these "problems" became evident. What is clear, however, is how these "problems" were often discussed as secondary to the external problem of accountability.

Whether actual or imagined, a perception existed across the interviews that some teachers in English departments were not teach writing in courses where writing was to be a key component. Rebecca, the WGE coordinator, explained how the WGE served to remedy this problem. She speaks to this issue from her experience as a classroom teacher prior to taking the position of the WGE coordinator:

I think that one of the things that the WGE did was hold teachers accountable. Because I know for a fact, and I regret to say, there were teachers who were not teaching any writing in 9th and 10th grade courses.

They were more interested in the literature. [...] Furthermore, they didn't like teaching, grading something that they thought as so entirely subjective. They didn't know enough about grammar and mechanics to make assessments and so it was really a quandary. It was easier just to avoid it. Now that is their own assessment of their ability. I'm not saying they weren't able, I think they were perfectly able. I think that teaching composition is a great deal of work; however, it was easier not to do it. I think what the WGE has done is make everyone more accountable because these kids don't graduate without it.

Rebecca's assessment of the writing problem prior to the WGE is two-fold. First, teachers in 9th and 10th grade were spending more time on reading literature, ultimately eliminating writing from the English classroom, a concern shared by several teachers. Second, teachers found great difficulty rendering judgment on student writing, something they saw as "subjective." Several teachers echoed Rebecca's quote; in particular, teachers discussed the difficulty in striking a balance between teaching reading/literature and teaching writing. Most often, teachers cited a lack of time as the reason writing would be pushed out of their instruction. Regardless of the reason why writing was not being taught in English classrooms, a problem existed of how to get teachers teaching writing to all students. Creating an exam that all students must pass puts the responsibility to teach writing on teachers, especially at the 9th and 10th grade levels.

The call for teacher accountability was only one component of BPS's "writing problem." Many teachers pointed to the absence of articulated goals for students and the lack of a shared language around teaching writing that existed in their schools prior to the

WGE. Additionally, teaching writing before the invention of the WGE was often an isolated experience where teachers were unsure of how other teachers taught and talked about writing with their students. Andrea, a veteran English teacher at Marshall High School, captures the insecurity and difficulty she experienced as she taught writing prior to the adoption of the WGE:

You know, you have to look at how we taught [writing] say fifteen or twenty years ago, before we started this process. I remember as a relatively new teacher really having some concerns because there was a teacher that I really admired. In fact, she had taught me when I was a student and she had a reputation as being a really hard teacher. I was always very insecure when I sent students to her classes because I wanted them to be prepared, but I really didn't know what that meant. Prior to [the WGE] you were in your classroom, you may have had a course outline, you had books in the book room, but we didn't necessarily spend a lot of time talking about the content [of writing]. I mean there were just assumptions that everybody was on the same page, [but] we weren't talking about what we expected. We didn't necessarily have common language. We definitely didn't have commonalities in terms of rating.

Andrea's quote is typical of many teachers who addressed the absence of a writing curriculum or program across the grade levels and the district. The lack of a program resulted in teachers feeling isolated in how they approached the teaching of writing, unaware of what their colleagues were doing next door, and unsure how they could prepare their students as they advanced through their high school experience.

Furthermore, a common discourse for teaching writing and the knowledge of how to evaluate writing did not exist in any uniform or systematic way. As Andrea's quote shows, it wasn't until the WGE was put into place that these problems were addressed. Through the construction of the exam, its rubric and exemplars, a standard for writing was defined, bringing with it a shared discourse and common practice for teaching writing.

Without the WGE teacher of BPS lacked a community of practice. Through scoring sessions and teachers becoming acquainted with the tools of the WGE, a culture of teaching writing emerged. Continuing with Andrea's quote, she provides a contrast to the isolation and lack of shared language and content knowledge on writing she and the teachers in the district experienced prior to the implementation of the WGE:

When we first started scoring papers [for the WGE], suddenly we were having discussions about writing and we started developing common language and we started articulating what made for a good piece of writing or what made a specific piece of writing weak. Suddenly we were closer to [being] on the same page and not only building-wise, but also district-wise. So when I came back into my classroom, and I'm teaching students, I have a better sense of what other English teachers will be looking for or what [students] need. And I think it raised the level of concern; it has given us language to discuss the teaching of writing.

Andrea frames the internal writing problem around pedagogical and programmatic issues of writing. Like many of the teachers in the study, Andrea point to the success of the

WGE to provide a common discourse and standard to unify teachers of writing within schools and across the district.

In the creation and implementation of the WGE, several problems both external and internal were solved. In the end, the district had an accountability mechanism to measure students based on minimum standards, and teachers were provided a language for teaching writing and the responsibility to make sure they prepared all students to pass the WGE.

The Problem of Negative Pressure of Teaching

While the “official story” presents the positives of articulating a discourse of teaching writing and creating a form of teacher accountability, the counter-narrative teachers also shared was the pressure they felt to get students to pass the WGE and how that pressure resulted in them narrowing their instruction to narrative and expository writing only. When teachers talk about teaching classes where they are working with students who need to pass the WGE, they most typically talk about the “pressure” they or their students feel to pass the test. Conversely, when teachers talk about elective courses, they discuss the “freedom” they have teaching writing. This pressure/freedom binary became a clear tension as I talked with all the teachers. Typically, teachers discussed an increase in pressure in proportion to the number of students in the class who needed to pass the WGE (see chapter four for a case study of two teachers).

The teachers who work with General English and Composition students were the most vocal about the pressure they feel in their classroom. General English classes are comprised of students who need to pass the WGE in order to graduate, while

Composition courses, depending on the school, might have a mix of students who have or have not passed.²¹ June spoke of the pressure she feels to help her students graduate:

The big bulk of pressure we feel in these classes is not external but internal, because we care so much about the kids. We really, really want them to be able to pass this exam and graduate....And so we work really hard trying to get them ready for this test so that they can pass. We don't want them to fail because we care about them so much. So it's really gut wrenching, those classes are gut wrenching to teach.

The pressure most teachers described was similar to June's – an internal pressure – one of responsibility to the students to make sure they receive their diploma. June points to the internal pressure teachers experience, but the external pressure of time—meeting the standard in a confined amount of time—was especially clear when talking with composition teachers, as many of them work with seniors in their last semester. As Mark discussed in an interview, “a composition teacher who knows they have a senior for second semester who still has not met [the minimum standard] and [the student] meeting the demonstration is resting on the teachers instruction, the student's motivation, and the student's skills has lot of pressure there, and it can be pretty stressful for comp teachers.” In fact, the pressure of getting students up to a passing level, combined with the difficulties of working with resistant students who feel like failures, has lead to teachers in Wilson High School not wanting to teach General English or requesting a year off from teaching that course.

The pressure/freedom binary was not limited to remedial writing courses, but presented itself with English 10 teachers who had classes with a mixture of students who

have passed and have not passed the exam. Andrea, who talked about how the WGE provided a common language for teaching writing above, also discussed the ways she approaches courses differently based on the population of students she is working with. In the quote below, the first class she speaks of is an honors section of English 10 where all of the students have passed the WGE and writing in that course is more “organic.” In the second class, a regular section of English 10 where only half of the students have passed, the writing she assigns is prescriptive and in line with the requirements of the WGE:

I’m very concerned about kids passing that exam and one of the ways that shows up is in the honors class where every kid in the class had already passed that exam. I feel much freer in the kinds of writing assignments that I give. Some of that is [because] they already had those very basic skills so they are capable of more sophisticated writing. But I tend to direct teach less about writing and the writing goes out of the content more. It’s more organic because I’m not worried about not hitting something. In the [regular English 10] classes I’m very aware that there are some students who still haven’t met that requirement. For example, when we were working on the quilt [project], I used it as an opportunity for kids to do narrative writing and expository writing. [...] We did some cause and effect writing and it was really again looking at that five paragraph essay really looking at the basic structure. So I think it is always in the back of my mind. Even the [quilt writing] we did, I tied it back to the two modes that the district has spelled out.

Working with two different courses of sophomore English, Andrea illustrates how the status of her students causes her to think about the writing experiences she provides.

Writing for the honors students comes out of the content of the class and is organic; however, because the exam is “always in the back of [her] mind,” writing opportunities for regular English 10 students are tied directly to the modes of writing on the WGE.

Andrea also discussed the ways being accountable to teach narrative and expository writing took her focus away from other forms of writing with students in English 10. “We do some poetry but not as much as I would have probably done in the beginning of my career. Because there really isn’t a graduation exam in poetry, mostly what we do is stuff that will get them ready for the graduation exams.” A well-documented consequence of standardized tests is how content not tested is pushed to the periphery of teacher’s instruction.²² Teachers across both schools echoed Andrea’s sentiment that narrative and expository writing has become the predominant mode of writing taught in courses where students have not passed the WGE. They also cited this trend in their instruction as a concern and frustration.

There is not enough space to show the variety of examples teachers discussed and classroom activities I observed where teachers enacted ‘freedom’ in classes that were not tied to the WGE. For courses where the WGE was not a factor, like creative writing, journalism and newspaper, American Literature, AP courses, and elective literature courses, teachers discussed the freedom and ability to have students write for real audiences, experiment with new genres, or move beyond a minimum curriculum. Additionally, the freedom allowed teachers to remove themselves from the WGE system of writing while simultaneously not having to be concerned with getting their students to

minimum proficiency. As one teacher said after looking at her roster, “everyone has passed reading and writing. One, two, three, people have not passed the math test but hey, that’s not my problem. (laugh) Yeah they’ve all passed the test which is kind of nice in terms of pressure.” While this teacher celebrates her students passing the graduation exams, thereby minimizing the pressure and responsibility she feels, she also illustrates another counter-narrative of the deferred responsibility within the WGE system which I will discuss in the final counter-narrative.

The WGE solved BPS’s “writing problem” by creating a system of accountability, tools to measure writing, and a shared language around the teaching of writing; however, through its implementation new problems arose, leading teachers to narrow their instruction in courses where students have not passed the WGE, and to feel internal pressure and burn out. The official narrative providing teachers with a clear standard to get their students to and common discourse for teaching writing, becomes complicated when the unintended consequences of the impact of negative pressure on teaching are juxtaposed. These two narratives illuminate the fissure teachers in the study discussed in their classrooms and their experiences working with students to pass the WGE.

A Teacher Created Test

Charged with creating a way for students to “demonstrate the knowledge and skills necessary to write consistently at a standard,” the district appointed a committee of teachers representing each high school to create a standardized test. Heading the committee were teachers with experience scoring Advanced Placement exams and others who were leaders in their buildings or departments. By having teachers construct the

exam, BPS would be able to laud the WGE as teacher-designed, an important component of generating buy-in from teachers across the district who were not a part of the actual committee. The teachers involved in the original committee reflected positively on the creation of the WGE as a useful professional development experience, but more importantly, for the first time in the careers of the teachers on the committee, teachers from different schools within the district were in the same room talking about writing.

The committee was dedicated to having students write as part of the examination, allowing for a more authentic measurement of student writing than just using a multiple-choice test. The Writing Graduation Exam consists of two timed writings where students respond to a prompt in a specific mode, and a thirty question multiple-choice test on grammar and editing.²³ The expository essay prompt asks students to form an opinion, provide adequate reasoning and support the reasons with evidence, “such as examples from literature, history, public life and/or your personal life.” The narrative asks students to write a narrative to be submitted to a faux teen literary magazine. The multiple choice portion focuses on common errors in grammar and usage such as subject/verb agreement, apostrophes, punctuation marks, homonyms, fragments and run-ons and spelling (See appendices for sample prompts, and rubrics). The WGE is given to all students on the same days at the same time and are scored at BPS’s Central Office by English teachers several weeks after they have been written.²⁴

A “Teacher Created” Test

The “official story” touts the WGE as a teacher-created exam, but to equate that statement with teachers supporting the policy or buying into the creation of a

standardized test would overlook the complex relationship between teachers and the district. Sarah, a teacher involved in the early stages of the WGE spoke of her skepticism when working within the “bureaucratic system” of the district. “If the school board asks you to make a test, I always feel like I should know the enemy kind of. And so I will work on that committee to make sure there is something of value in it for teachers and for students.” Referring to the school board as the enemy might have been hyperbole, but it points to the distance between those who make policies and those who work in the realities that are created from policies. For Sarah, participating on the committee to create the WGE provided the possibility to create “something of value” for students and teachers, rather than merely comply with a school board mandate.

Mark, the district’s first WGE coordinator, told a similar story on why he applied for the position of WGE coordinator:

When I applied to be the WGE coordinator it was because I was concerned about the process. I wanted to try to make sure that rather than complaining about what we had to do, I could possibly influence what we had to do and what kids had to do along the way. Now, when you are in a position like this it all changes. You get to see how complicated it actually is. But all in all, one of my goals has been to make sure that control stays in classrooms and that teachers continue to teach students a process approach to writing. That we are not just practicing to take this exam to meet that minimum proficiency but that we are looking at writing as a whole.

Marks sentiment is similar to Sarah's. Rather than allowing an exam to ruin the district's practice of process-based writing pedagogy, something he valued, he applied for the administrative position to have influence in the process of implementing the WGE and to prevent the WGE from turning the district into teaching writing as test-prep.

A common denominator in Sarah and Mark's stories is coming to work on the WGE (in different capacities) in order to assure that the principles they believe in were voiced as part of the enactment of the district's high stakes exam. They both discussed participating in the process with uncertainty, and possibly skepticism in Sarah's case, on what was being asked of them and what they held as important as writing teachers. Though the roles they played were prescribed and served as part of the structure needed to implement the WGE, both discussed their position in relation to keeping some form of autonomy in their classrooms. At stake was the power to decide what was going to happen to writing throughout the district. Knowing what the school board was up to resulted in Sarah's participation. For Mark, a pedagogical commitment to process writing and teachers maintaining control of teaching writing in their classrooms brought him to work in an administrative capacity within the WGE.

However, it is important to note that not all teachers in the district were at the meetings when the test was designed, nor was their input sought in creating the WGE. For some teachers not involved in the WGE process, the assembling of a committee of teachers to create the WGE was an example of the district selecting teachers with similar viewpoints that would do the job the school board required. One such teacher was June. In our first interview it became apparent that she was frustrated with many components of the WGE, the test itself and how it labeled and slotted students into different courses, as

well as who was chosen to create it. When asked about the WGE being teacher-created, June responded:

Well, there are teachers and there are teachers. My gosh, we run the full range just like our students. I've always taught at risk [students]. I've never taught a diff or AP class in my life. All of my reference has always been with struggling students. But I feel that teachers who work with those students are underrepresented in a lot of things. We don't volunteer to do these things, you get selected. My colleagues in the world of at risk have not been selected. Marshall High has the highest mobility rate, the highest poverty rate, the highest number of ELL students in the district.

We should be represented and have a voice when these things are decided.

Teachers from June's school were present on the committee that created the prompts and rubric. However, June's comment points to the perception among several teachers, specifically those who worked with struggling writers and ESL students, that the district's committee lacked teachers who regularly worked with these student populations. Furthermore, the appointment of some teachers over others to design the tools which would define writing and the parameters of the WGE system, led to teachers feeling disenfranchised from the process and hostile toward the WGE.²⁵

To call the WGE teacher-created, despite the fact it was designed by a small group of teachers sitting in a room for several days, overlooks the complex relations these three teachers have to the policy makers who mandated the exam and the district officials who appointed the original committee. The stories presented by Sarah, Mark, and June on their participation or exclusion from working on the WGE resist the clear narrative of

teachers coming together to design a test based on the values of the district (rather than an outside testing company or the state), and highlight the relation of power working within the district. As June said, teachers were selected by the district; they did not volunteer. And though BPS attempted to have representation from across the district, like June, teachers offered counter narratives indicating how certain kinds of teachers were chosen, typically leaving the teachers of ELL and special education students out of the conversation of setting the district's minimum proficiency. Sarah and Mark understood that being part of the early conversations and implementation meant having a voice and the ability to exercise power in the process of reforming writing in BPS. But teachers who were not part of the original committee or who had been hired since the implementation in 1999, often expressed frustration over not knowing why decisions were made or if there was any way to change the test.

Creating Opportunities to Learn

Due to the fact that the WGE is derived from a school board policy, and if student do not pass they are not awarded a diploma, the district is legally responsible for students receiving continual instruction. Ideally, students will have passed the WGE by the end of their sophomore year; however, for the students who have not, the district must assure it provides these struggling writers “opportunities to learn.” To meet the legal responsibility of assuring students opportunities to learn, the district created a series of remedial writing courses—Writing Essentials, General English, and Composition—each with a curriculum to help students meet the minimum standard set by the WGE. By providing opportunities to learn through a set curriculum and placing students who failed

the exam after their sophomore year within a series of classes, the district assures struggling writers are treated fairly and receive multiple opportunities to pass the exam.

Students are placed into the scaffolded courses based on their most recent test score on the WGE and the recommendation of their previous English teachers.²⁶ Students who score a 1 or a 1-2 combination on their WGE are placed in Writing Essentials, a one semester course. These students struggle composing a coherent sentence or write in a simplistic manner often difficult to read. Thus, Writing Essentials focuses on constructing sentences and paragraphs, with the end of the semester goal of students writing multiple paragraphs on a single topic. Students who score a 2 or 2-3 combination on the WGE are enrolled in General English, a year-long course. Students in General English might be able to put a lot down on paper but lack focus, or they might not know how to paragraph or support their ideas. General English focuses on helping students organize what they want to say, typically through the five-paragraph essay.²⁷ Students who score a 3 on the WGE are placed in Composition, which in addition to being a course can serve as an alternative to the WGE. In Composition, students are given more time to write a narrative and expository essay, and can compose it on a computer over the semester without assistance from the teacher. If the teacher feels the piece is a 4 based on the WGE rubric, she can ask another teacher to score the writing, and if the second teacher agrees, the essays are sent to the district office for approval of passing the WGE.

Placing students within a curricular structure serves as an efficient manner to assure students and the community that students are receiving multiple opportunities to learn the required skills necessary to pass the WGE. Students are classified and grouped

by their inability to pass the exam, allowing teachers to work with students of similar ability levels.

The Consequences of “Opportunities to Learn” for Students

Building and district level administrators discussed providing students an “opportunity to learn” as a matter of equity and fairness. Providing “opportunities to learn” was talked about as both a curricular and pedagogical issue the district responded to with the creation of the WGE. Constructing the three remedial writing courses (and the work of English 9 and 10) demonstrates to the public that BPS is accounting for the needs of students who have not passed the test after the completion of their sophomore year. The three courses serve to provide students with the opportunity to learn how to write at a proficient level so that they can pass the test. However, teachers often resisted the “opportunity to learn” narrative for two reasons: first, it assumed that if teachers were not specifically working with students on test prep, students were not getting proper instruction (opportunities); second, the impact of tracking students resulted in classrooms of students who perceived themselves as failures (as per the exam) and thus they resisted writing. While the official narratives discuss the courses as providing support to students, teachers and administrators also spoke of the importance of getting students to pass the WGE quickly so they could move on to other opportunities.

In an interview with Sarah, she articulated her frustration with the phrase “opportunity to learn.” For Sarah, the phrase is part of the bureaucratic discourse around the WGE (and the district as a whole) protecting teachers and administrators from lawsuits. Sarah has taught a wide range of English courses over her career at BPS, and

while not teaching General English at the time of the interview, she spoke back to the district who she feels questions her professional judgment of teaching writing. Sarah explained how district administrators fear a lawsuit and needed to guarantee students have an opportunity to learn:

...and so we need to make sure that there is an opportunity to learn.

Opportunity to learn, I mean that phrase is like. (In a sarcastic voice.) Oh, well, I didn't give them many opportunities to learn in my class. No we were doing poetry, or we were writing in response to images, or we were doing visual response poems. Sorry we didn't work enough on the WGE stuff. (Returns to regular voice) I think that message is very clear that these kids are in General English because they need to pass this exam.

And you darn well better give them the opportunity to learn over and over. I mean that is why they're in the class. The district has made sure that we are giving them opportunity to learn. Ok.

Over the course of the year I observed three different General English courses, and even though they were taught by different teachers in different buildings, the courses were quite similar and predictable. Week in and week out students were given a prompt on a Monday and by Friday they needed to turn in an essay, usually a narrative or expository. The teacher allotted time most every day for writing, revising, or on a rare occasion, some type of group work. When General English students did work in groups on writing, they often used a visual organizer to map the five paragraph essay format for expositories or plotted the narrative to make sure there was a clear beginning, middle, and end.

Sarah used a sarcastic tone to talk about the kinds of writing she might have students do which were not recognized by the phrase “opportunity to learn.” As she stated, the phrase “opportunity to learn” is a message from the district to General English teachers that they need to focus on narrative and expository writing at the expense of teaching other types of writing: “you darn well better give them the opportunity to learn...that is why they’re in that class.” And while Sarah was frustrated with the message sent by the district over the kinds of writing which are acceptable to teach in General English, she also articulated the difficulty of teaching a writing course when the students have been tracked by the system as failures:

Those kids feel dumb, you know, and they’re with other kids who feel dumb and so if you go through an hour and a half of your day feeling dumb and with other kids that feel dumb is this doing you any good? So I don’t know what to do with that. That’s what happens when you track classes though. And [the students] know why they are in there and a lot of them are special education students who feel dumb anyway. So how do we really give them an opportunity to learn? It’s part of my whole problem with the system. You know, I mean, so what do you do with kids, you want to immerse them in a literacy experience, how is General English an immersion in a literacy experience? I mean maybe it shouldn’t be about the WGE yet.

Again, Sarah’s counter-narrative pushed back against the institutional narrative of fairness and equity as she advocated for the students most affected by the WGE, addressing the difference between getting them ready for the test and providing them

with “a literacy experience.” Sarah’s discourse of immersion moves away from the “skills and knowledge” the policy names. Drawing on her knowledge of students and her experiences teaching General English, Sarah offers an alternative: maybe General English students need a different experience, one that is not focused on the skills of the WGE but on a broader notions of literacy, reading and writing. In posing an alternative possibility for intervening with students who have not passed the WGE, Sarah raises the issue of what counts as an “opportunity to learn.” This critique of the official narrative which claims students are receiving such opportunities from the district, could not be made without knowledge from inside the classrooms and the experiences provided in the tracked classes.

What was very clear of the “opportunity to learn” narrative was that while the courses serve the purpose of helping students to meet the WGE requirement, teachers and administrators both discussed the importance of getting students out of those classes quickly so they could move on to other “opportunities.” As Mark said, “we try to provide that motivation of finishing your WGE requirements to, I don’t want to say get out of those courses, but it opens up a lot of possibilities in what else [students] can take.” Maybe the narrowness of opportunities to learn in the remedial writing courses is best summed up by Sarah as she reflected, “it seems weird to say that students have more freedom to write different things once they have passed the exam. That just seems wrong to me.”

Coordinating the WGE

Due to the size of the district and the amount of coordination needed to maintain the WGE, an arrangement of personnel was established to accomplish the various tasks and responsibilities of the WGE. Because of its size, BPS has its own Educational Service Unit (ESU) that provides educational support, and in the case of the WGE, the psychometric expertise for the district.²⁸ The ESU serves as part of BPS and is housed in the district's central office. Mark discussed the importance of the individuals of the ESU within the district in an interview:

We're really lucky to have our assessment and evaluation people as part of our system because they work very closely with our teachers. They're really instrumental in helping us to develop reliable, strong, assessments. They're incredible at developing our assessment knowledge at the district and working with our teachers. They know these areas and in terms of legalities as well as the specifics of what makes for good assessments they are important people in this process.

Mark described the ESU as a collaborator with the district administrators and teachers, able to use their technical expertise to build assessment knowledge as well as maintain the district's assessments such as the WGE.

To assist in coordinating the WGE, the district created a new administrative position, the Writing Graduation Exam Coordinator, who coordinates all aspects of the WGE – from distributing practice exams to every high school English teacher in the district, to running table leader meetings, to training new scorers, to overseeing scoring sessions, to reporting the data to BPS's Assistant Superintendent. The WGE coordinator

also serves as a resource for BPS schools that are focusing on improving writing as a school improvement goal; however, the primary responsibility of this administrator is the coordination of the WGE. The function of these district level employees is to assist teachers and the district as a whole with the laborious task of conducting the WGE every semester.

“Assessment People”

As teachers, principals and district administrators discussed how decisions get made pertaining the WGE and assessment more generally, a narrative played out within the district of the “assessment people.” “Assessment people” are often the people on the hierarchy immediately above the speaker telling the story, who have a certain expertise or authority not held by the narrator. For teachers, “assessment people” are often Mark and Rebecca, for principals “assessment people” are Mark and Rebecca along with the district’s assessment specialists, and for Mark and Rebecca, “assessment people” are the district assessment specialists. The counter-narratives that come out of the “assessment people” myth is two fold: first, there is a constant deferment of authority and responsibility within the system, and second, it promotes a mentality where the narrators do not think of themselves as people who do assessment; rather, they only see themselves as a small part of the process, while the real decisions around writing assessment are made above them in the hierarchy.

In an interview with the principal of Wilson High School, I asked him to explain the role he plays in respect to the teaching of writing. He responded: “My role is to make sure that I get the right people in the right areas, set the goal, and then hold them

accountable.” This principal defined his job in terms of personnel and management: hire quality teachers, make clear the expectations, and then make sure they do what they are supposed to. But when the questions turned to the WGE, his authoritative position shifted away from managing a building to defer to the assessment people at the district level. When asked, “if the WGE place emphasis on learning goals and teaching practices you think are important?” the administrator responded:

I think they do, as much as I specifically know about it. I think they do. I would hope they do. We’re in trouble if they don’t. Sometimes you get in my position and you assume that the testing people and the people who are in charge of all that have done a good job. And I haven’t had an opportunity to take that test and analyze it. I’m not the writing expert. You know, I just hope people who do that are talking about it as they do it. That it works in cooperation with teaching strategies and that type of thing. We’re not teaching to the test but we are giving kids the fundamentals that’ll help them do well on the test.

The principal quickly defers responsibility in regard to teaching writing. “I’m not the writing expert” and “the testing people” are the ones who know if the WGE emphasizes the learning goals the district has for writing. This deferment of responsibility to the writing expert illuminates the roles specific people play within the system. In this way, the counter-narrative of the “assessment people” disrupts the hierarchical organization of schools as efficiently managing schools to illustrate how authority within a school can be displaced outside of it to the district level.

While teachers and principals at the building level name Mark or Rebecca as the people who are in charge of writing assessment, in his interview, Mark defers to the psychometricians of the Educational Service Unit as the people who are in charge of the assessment process. Mark explained how the multiple-choice section of the WGE came to be. “The assessment folks in the district really wanted that multiple choice section because that helps with their reliability and equating from test to test and their much more confident with right/wrong answers than the holistic scoring element of the two pieces of writing.” In the following chapter I will explore reliability and holistic scoring in more detail. For now it is important to see that the teachers who created the tests actually designed it as a writing test without a grammar/editing component. However, due to technical issues of large-scale assessment, the experts of assessment and evaluation dictated a more quantifiable and objective tool to measure writing. Mark discussed the possibility of changing the WGE to a different format, again advocating his goals of incorporating process writing:

I think it would be interesting to see if we could develop something that offers more of the process element along the way, but probably it’s unrealistic at this point in time. I think we’ve talked about doing something that is more based on a portfolio, accumulation, but there again it becomes somewhat unwieldy and our assessment people become uncomfortable with that in terms of assessments and reliability along the way.

As the two quotes from Mark indicate, despite his professional judgment, decisions made about the assessment tools are made by the “assessment people” who have technical

knowledge he lacks. Through teachers' and administrators' talk about "assessment people," what becomes evident is how control over writing assessment gets moved further and further away from the classroom—the site where writing and writing instruction happen—and into the district's central office where the experts reside. Teachers and administrators often position themselves within the system as non-experts lacking expertise in writing or assessment.

The Residue of Reform

In the vignette preceding this chapter, the top-down mentality of Mary's statement of not having schools, "only teachers and a district," became evident in interviews of BPS teachers and observing them in their classrooms. The policy, calling for a "demonstration" of "consistent writing," resulted in the district setting a minimum standard of writing students must meet to be proficient, making teachers accountable to students passing the exam, designing a test to quantify students' writing, placing students into a curriculum based on their score, and positioning experts away from the actual labor of writing in charge of the technical running of the system. This approach to reforming schools and writing more specifically is part of the residue lingering in school reform left over from the Progressive Era. While not a full enactment of Frederick Taylor's scientific management, the mechanisms of achieving efficiency are present in BPS's WGE system as they were in Progressive Era administrators S.A. Curtis and C.C. Curtin's attempts to create systems to manage their teachers and schools through standard tests developed by members of the new field of educational measurement.

Presenting the architecture of the BPS's system of writing assessment and accountability through the official story and counter-narratives provides an opportunity to expose the fissures between the intended and unintended consequences of a systematic approach to reform, while simultaneously providing a space for teachers to speak back to the systematic approach of reform in order to articulate their lived experiences within the WGE system. In the following chapters I move further inside the system to explore fissures within the WGE system. Chapter three focuses on disruption within a scoring session of student exams. In a process where inner-rater reliability is privileged, I contend that paying attention to fissures when teachers question or critique the rubric or their role as scorers presents the possibility of reconsidering issues of test validity. In the fourth chapter I explore the ways the WGE impacts teachers' judgments of writing and the ways teachers draw on their subjugated knowledges to claim agency within their classrooms.

Do You See What I See?

Fans were brought in to help circulate air and cool the room, warmed by a combination of the spring sun shining through the windows and body heat of about forty teachers scoring expository exams. The teachers had returned from lunch over two hours ago, were recalibrated, and currently found themselves pushing through the last few batches of expository papers. I sat at the back of the room observing. The table leader to my right, sitting at the head of the rectangular folding table, stood up and walked over to one of the scorers at her table.

“Would you mind reading this one again?” the table leader asked the scorer.

“Sure,” the scorer said.

“Maybe read it next to the 3 exemplar. See if you think it’s similar.”

“OK.”

The table leader sat back down. A few moments passed. The scorer read the essay a second time, flipped back to the first page and quickly scanned a few lines. She then picked up the three exemplar out of the stack of papers piled to her right and read for twenty seconds. Placing the 3 exemplar down, she looked to the table leader, nodded her head in affirmation and whispered, “Yeah” to which the table leader nodded back.

A few minutes passed and the table leader interrupted the silence in the room with a forced whisper. “Hey guys, we really need to be using the exemplars as our guide. We’re starting to get low on our scores. Go back to the exemplars, especially with the two-three splits.”

Fifteen minutes later at the table to my left, a scorer motioned the table leader over to her seat.

“I think this is a zero,” the scorer said to the table leader.

“Why’s that?”

“It’s only five words long. All the student wrote was, ‘not going to do it.’”

The table leader paused “Well, technically, it’s not blank, and it is in English.

The pronoun ‘it’ references the prompt, so I think it would be too brief to assess.”

Making eye contact with the scorer, “Yeah, it’s a one.”

As the table leader walked back to her seat the scorer rolled her eyes, shook her head from side to side and mouthed “OK.”

* * *

Chapter 3

Disruptions in Scoring Writing:

Possibilities for Conversations on Reliability and Validity

Up to this point I have illustrated how concerns over efficiency and the implementation of a piecemeal version of scientific management resulted in a bureaucratized school system wherein writing is measured and managed. Through an archival study in chapter one and the exploration of institutional narratives and counter-narratives in chapter two, I have illustrated how the ideology of scientific management is still practiced in a disjointed but pervasive way within the teaching and assessment of writing in public schools today. In this chapter, I examine the theoretical pillars of classical test theory—reliability and validity—through observations at Butler Public School’s scoring sessions of the Writing Graduation Exam. Traditionally, reliability has been the renowned pillar, more often discussed in the professional literature of large scale writing assessment. Reliability is part of the residue of reform within our practice of writing assessment today. As Michael Williamson has pointed out in his essay, ‘The Worship of Efficiency,’ education’s historical obsession with efficiency has privileged conversations of reliability over conversations of validity within writing assessment. “This concern for efficiency and fairness,” Williamson writes, “the issues with which the early writing assessment specialists wrestled, has undermined the necessary next steps in

developing approaches to writing assessment with even greater validity. How might a different approach to education lead to a different approach to assessment?" (167).

Williamson's question serves as an exigency for this chapter as I examine moments where disruptions occur within a holistic scoring sessions or table leader meetings and consider how these disruptions might guide us toward a different approach to assessment and education. The disruptions might be viewed as forms of what Michel Foucault refers to as discontinuity, when fissures or ruptures occur within a unified system, in this case a system designed to produce a high inter-rater reliability between scorers. Within these moments of fissure reside possible spaces to ask questions about the WGE, which, as Williamson suggests, could lead to a re-imagining of writing assessment. In these moments I see teachers in their roles as table leaders and scorers asking questions upon which a critique and revisions could be made of the Writing Graduation Exam (WGE) system. It is in these moments that we can begin working toward writing assessment with greater concern and awareness of the ways the WGE impacts students, the ways the test constructs a definition of writing, and the way students are prepared to take the exam: the validity of the WGE. However, in the actual moments of scoring, these questions are passed over due to concerns of inter-rater reliability, efficiency, and time. Therefore, I conclude this chapter outlining a heuristic based on the questions derived from the disruptions in order to help teachers and administrators consider the possibilities of creating systems of writing assessment that re-centers assessment practices on validity.

Classical Test Theory

To help frame the moments of disruption within BPS scoring of student writing, I will quickly overview classical test theory and holistic writing assessment. Large-scale writing assessment has been dictated since the early twentieth century by classical test theory. Founded in the tradition of positivist psychology, classical test theory claims reality can be objectively measured independent of human bias. In chapter one, I explored how Edward Thorndike's maxim "if it exists it can be measured" resulted in writing scales and standard tests to "objectively" measure writing. Positivists looked to isolate skills and abilities and then create devices to measure those skills. Brian Huot articulates the ways positivists attempted to objectively measure writing: "Positivist reality assumes that student ability in writing as in anything else, is a fixed, consistent, and acontextual human trait. Our ability to measure such a trait would need to recognize these consistencies and could be built upon psychometrics, a statistical apparatus devised for use in the social and hard sciences" (83). The creation of an empirical methodology to measure writing would provide the means to discover the natural laws governing writing and student behavior. The statistical apparatuses created out of the notion that measurement would yield a discovery of the natural laws of writing and learning were explored in the first chapter. Despite the fact that both the fields of education and composition have moved beyond the positivist paradigm toward constructivist theories of learning and writing, the residue of positivism and classical test theory lingers in the theoretical concepts of reliability and validity still operative in writing assessment today.

Reliability is driven by "true score," an idealized reflection of what a student knows or a demonstration of what the student can do, and would be consistent with future

performances on the same instrument or test (Huot “Holistic” 203). However, true scores do not actually exist due to instrument or testing error. Because error is inherent in large-scale assessment, an individual’s score is comprised of the true score plus the margin of error. Therefore, traditionally defined, reliability is concerned with the “degree to which test scores are free from error of measurement...[as] measurement errors reduce the reliability (and therefore generalizability) of the score obtained for a person from a single measurement” (Moss 1994 6). Reliability is most often defined in terms of a testing instrument’s generalizability, and consistency (Slomp and Fuite 191). The ability to consistently reproduce a test’s results provides a means for comparisons and generalizations to be made between students, teachers, and institutions because all students had a standardized experience taking the test and each test was scored consistently.

The counterpart to reliability is validity, which refers to the test’s ability to measure what it set out to measure. Validity is not nearly as fixed of a concept as reliability. Tracing conversations within the psychometric community in the second half of the twentieth century demonstrates the ways validity has morphed from a statistical correlation to the examination of actions taken based on the test data. Validity was first concerned with a test’s ability to “correlate with outside criterion” (Huot 48). This correlation became known as criterion validity and served as one part of a three-pronged examination of validity comprised of content, construct, and criterion validity.

- *Construct validity* is the definition or make-up of the skill or performance being measured.

- *Content validity* is the content of the test and its relationship to what the test purports to measure.
- *Criterion validity* is divided into concurrent validity and predictive validity. Concurrent validity is the test's ability to correlate with another independent measure such as grades or another standardized test. Predictive validity is the test's ability to predict the student's future ability on the skill or content measured.

Maurice Scharton illustrates these four perspectives on validity through the following heuristic regarding writing assessment, "How do you define writing ability? What writing curriculum do you assume that the test takers have experienced? Does this test give the same results as other tests? Does performance on this test accurately forecast the test taker's degree of success as a writer?" (55). For a time these three prongs of validity were used independently to validate assessments; however, considering these three forms of validity as "the holy trinity" of measurement or as "separate but equal" criteria to validate an assessment has been an "idea whose time has gone" (Guion in Huot 48, Cronbach 4). More recently, validity has been expanded to consider the social consequences of a test. Pioneering this movement of validity theory has been Lee Cronbach and Samuel Messick. For Messick, "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment (emphasis his) (Messick 5).

Kathleen Blake Yancey historicizes writing assessment using the metaphor of overlapping waves to illustrate how the field has moved chronologically but without fully

displacing the previous wave. In her history, Yancey describes the first wave of writing assessment as comprised of objective multiple-choice tests which were concerned with establishing high reliability. The second wave began to question the validity of the multiple-choice tests and was marked by the rise of holistic writing assessment. In the movement between these first two waves of writing assessment emerges the tension between reliability and validity. Early proponents of holistic writing assessment argued that indirect writing tests (multiple-choice tests) provided high reliability of scores because a student's answer was either right or wrong and human judgment was eliminated from the process of scoring; however, these tests were not a valid way to assess students' writing ability because students did not actually write as part of the assessment (White, Cooper, Breland and Gaynot). Additionally, multiple-choice tests only focused on grammar and editing and those skills were being conflated with the actual performance of writing (Weaver 9-24). Critics argued that indirect writing assessments lacked a "face validity" or the test looking like what it purported to measure. A more valid assessment would allow students to write and be evaluated on their writing.

Holistic scoring of writing samples was viewed by many in the field of composition as a victory for test validity—because the test purportedly measures what it says it measures; however, Brian Huot reminds compositionists that holistic scoring is “a product of the same thinking that produced the indirect tests of grammar, usage and mechanics” and like indirect writing assessment, “holistic scoring was developed to produce reliable scores” (24). The means by which reliable scores are established in holistic scoring is through inter-rater reliability or the training of readers to read essays in a standardized manner according to a rubric and exemplars. Typically, in a holistic

scoring session each essay is read and scored by two readers. Ideally the scorers will give the essay the same score; however, since humans are reading the possibility for disagreement (“error”) exists. Inter-rater reliability is a measurement of the agreement between scores. Perfect reliability would result in a rating of 1.0, a perfect reading indicating all readers read all their essays and scored them the same as the other reader who also read and scored the essay. Typically the rating of .7 is used as an appropriate rate of inter-rater reliability for holistic scoring sessions.

The relationship between reliability and validity is a tenuous one as concerns over reliability have overshadowed questions of validity within holistic scoring and large scale writing assessment. Michael Williamson has traced how writing assessment historically has been concerned with fairness and efficiency, leading the field of writing assessment to favor reliability over validity. Brian Huot has questioned how the field of writing assessment may have taken a step backwards by focusing so much attention on developing procedures to insure consistent scoring when he writes, “It is probably safe to say that the very emphasis on reliability, which made holistic scoring procedures acceptable in the first place, has retarded their scope and stunted their growth” (204). Grant Wiggins has argued that “in performance tests of writing we are too often sacrificing validity for reliability; we sacrifice insight for efficiency; we sacrifice authenticity for ease of scoring” (129). Wiggins also raises how an obsession over reliability at the expense of validity “can easily drive out pedagogical concerns [out of writing assessment] if we are not vigilant” (129). With these concerns about the privileging of reliability over validity in mind, I now return to BPS to examine the adoption of the WGE and the ways they maintain a higher inter-rater reliability.

BPS's Adoption of Holistic Writing Assessment

In chapter two I explained how Butler Public Schools, after passing the policy for a graduation exam in writing, appointed a committee of teachers to construct the WGE. The committee was led by teachers who had experience scoring Advanced Placement exams and was committed to having students write as part of the writing exam. The format created for the WGE was comprised of three components: a narrative essay, an expository essay, and a thirty question multiple-choice test on editing and grammar. For both the narrative and expository essays, students are provided a prompt which they must respond to in forty-five minutes. Both the narrative and expository are scored independently on the district's holistic rubrics. The multiple-choice component was added later as a way to assess grammar and usage as well as aid with the exam's reliability. Together the three scores are averaged and compared to the cut score to determine if a student passes the examination.

In interviews, teachers and administrators explained that the teachers on the original committee had experience with holistic scoring of AP exams and used the AP format as the base for the WGE but modified it to fit BPS. The AP format requires students to write three essays that focus on rhetorical analysis, synthesis, and argumentation. The first modification BPS made to the AP format was switching the essays to two of the traditional modes of writing. The district had a commitment to teaching narrative writing starting in the elementary schools and wanted the test to account for narrative writing. Faculty felt expository writing was a crucial skill for an exit exam because it is a form of writing many students need for their vocations. The second modification came in creating a six point rubric rather than AP's nine point

rubric. While slight revisions were made to the tools of the holistic writing exam to fit the needs of BPS, essentially, BPS merged the format of the AP exam into the context of their district. In doing so, BPS assumed with the exam the theories of reliability and validity that have long guided classical test theory without considering the impact these theories might have on their district, schools, teachers, and students.

The practice of transplanting a test format across contexts and institutions for different purposes is not a new practice in writing assessment. In fact, this is exactly what early advocates of holistic scoring in composition did. Writing Program Administrators like Edward White in the University of California system borrowed ETS's Advanced Placement Programs' new testing technology, holistic writing assessment, and replicated it in their universities.²⁹ Creating prompts to focus student writing, selecting anchor papers and scoring guides to norm the scoring practices of teacher-readers, and calculating acceptable agreement (reliability) allowed White and others to argue they were implementing a more valid and authentic writing assessment because students were actually being measured on their writing. In doing so, movement toward holistic scoring worked "both within and against the psychometric paradigm" of reliability and validity (Yancey 491).

In Teaching and Assessing Writing, Edward White argues that holistic reading and scoring should be viewed as developing an "interpretive community" among readers (214). White argues that the training or calibration of readers to align their reading practices to standards is not "indoctrination into standards determined by those who know best, but rather, the formation of an assessment community that feels a sense of ownership of the standards and the process" (215). The problem with White's assertion

with regards to BPS is the standards, rubrics, prompts, and exemplars were all created or chosen by the district appointed committee and the rest of the teachers were trained to read papers according to the standards and rubrics. The district's top down implementation of the BPS results in scoring being less about community and more about accurate and reliable scoring.

The interest in discussing reliability over validity was evident in interviews and observations of teachers and administrators at BPS. Both district administrators pointed to the high inter-rater reliability of the exam and a few teachers mentioned the improvement in scoring over time. BPS maintains a high inter-rater reliability on their essay scoring. In 2005 the district reported 85% of the expository and 87% of the narratives were read with exact agreement and 99% agreement within adjacent levels on both the expository and narrative essays. Evident in interviews and scoring sessions was the need for the WGE to be reliable and most often this meant having teacher-scorers read in a uniform manner.

Despite the effort for the system to work towards reliability, there were moments in scoring sessions or table leader meetings when teacher disagreement would lead to a disruption in the scoring process. Foucault's notion of discontinuity is useful to explore the ways disruptions occur on a micro level. Below I showcase three disruptions to explore what happens in moments when the system designed to achieve reliability fissures or ruptures due to the fact that teacher-scorers, for a brief moment, do not work toward consensus and reliability. These moments were infrequent but manifested themselves across several observations and provide a small space to level questions about how the system works and how it might work differently toward different ends.

Typically when the disruptions did occur they were quickly pushed aside or corrected by a person in authority such as a table leader or question leader. However, the three reoccurring disruptions represented below raise questions over the purpose of the WGE and the usefulness of holistic scoring, consequences of the WGE on students, and the construct validity of writing as defined by the WGE.

Questioning Holistic Scoring

The first disruption occurred in a table leader meeting. The purpose of table leader meetings is to read through sixty to eighty student exams in order to find exemplars—papers that correspond to the six scores on the rubric—as well as to select papers to be used to train the teacher-scorers prior to the scoring session. The sample of sixty to eighty papers was pulled out of the entire population of students by the question leader prior to the table leader meeting. While the primary purpose of the meeting is to find exemplars, the sample papers are also being scored as live papers by the table leaders. That is, the students whose papers happened to be selected will receive the actual score the table leaders agree upon. In this moment the table leaders were working with the question leader to determine the proper score for paper RRR. A disagreement occurred between the table leaders on how the paper should be scored and in this disagreement a space opened to ask questions about the assessment practices of the WGE.

TL1: I had RRR as a 5.

QL: We ranged from 4- to 5. What keeps it a 4?

TL3: Sentence structure and diction.

- QL: It stayed a 4 for me because I wanted the writer to be more specific.
- TL2: It's not as thoughtful as a 5 (reads from the essay) "She can handle..."
- TL4: There are lots of good things, but at the end with the party, I was like where did this come from?
- TL5: I thought it was a 5. I wasn't bothered by the end.
- TL1: I like that there was humor and she wasn't crude. That shows sophistication.
- QL: Let's compare it to LL, II and M. LL is the exemplar.
- TL3: I'm looking for specificity in RRR and it seemed to lack. I think it's a 4+ like here, "she gained our respect." How did she do that?
- TL1: (continues reading the passage) Here she shows it. I think it's a low 5.
- TL2: But after that she goes on to a new topic and doesn't address it.
- TL4: I kind of wish there was an example of making her feel special. It kind of missed out on support. Fluency I would say a 5 but support a 4.
- QL: Look holistically against the rubric where does it fall? (reads the 5—commendable response from the WGDE rubric. See Appendix B) Overall, the response is thoughtful and effective; it provides a commendable explanation. The writer incorporates

sound logic, reasoning and evidence, controlling the structure and a range of stylistic elements appropriate for exposition.

TL3: This is where I'm like, Can we have a check list please?

TL1: It's weaker than the other 5s

TL3: Some but not all of it is commendable. There is a range of stylistic elements. Mark, what should we do?

TL1: This person is going to pass, let's go on.

QL: We don't need to use this one in training. Let's go on to SSS.

In this short scene, Table Leader 1 began by stating she believed paper RRR to be a five, citing the paper's sophistication and support. Table Leader 3 and 2 scored it a four, arguing the diction, thoughtfulness and support was lacking preventing it from being a five. Table Leader 4 wanted to give RRR two different scores, one for fluency and one for support and the Question Leader continued to bring the table leaders back to the tools by asking the table leaders to compare RRR to the exemplars and reading the rubric out loud. In the end the group could not come to consensus and RRR was not used as an example.

While consensus of a score could not be reached, the conversation offers a space to question the purpose of the WGE and the usefulness of a holistic rubric for the intended purpose. All the table leaders above agree that the students wrote well enough to meet the criteria for a proficient essay: they all scored it at a four or higher. The debate in the scene is if RRR could be used as an exemplar for training teachers and therefore the need to actually place it on a 1-6 scale is important. However, scoring the essay with a whole number raises the question, why is a 1-6 rating necessary for what the district

claims is a minimum writing proficiency test? Recall, the district policy calls for students to “demonstrate the knowledge and skills necessary to write consistently at a standard still to be determined.” Given that mandate, is a six point rubric necessary? Or could a simpler, more minimal approach to assessing writing serve the purpose better?

Critics of holistic scoring like Peter Elbow have argued that holistic scoring can lead to the “dangerous assumption that there is a ‘true score’ for any piece of writing”(85). Elbow advocates for a minimalist approach to holistic scoring that relies on fewer boundaries and even for a Yes/No, Pass/No Pass type of rating (128-9). Thus, rather than score each student essay on a 1-6 scale, a minimal approach to the WGE might form a binary question, did this student meet the minimum standards? However, the district’s explanation for using a six-point scoring rubric is two fold—first, it assists in the placement of students who do not pass the exam, and second, because the WGE consists of three parts which are averaged, the need for a numeric score is necessary to average the three portions. The first reason, to assist in placing students, presents the question of the purpose of the exam. Is the WGE a minimum proficiency exam or a placement exam? While it is serving as both, should it be? Is this a valid use of a single instrument?

If BPS is going to use the WGE for placement of students, and the score is the main data source for placement, then the degree to which a student failed is important since it impacts where they are placed in the curriculum. However, a common criticism of the WGE by teachers, particularly those who taught General English and Composition, was knowing their student scored a 2 (for example), in the case of General English, was not useful information about that student. While it could be argued that the WGE is a

summative assessment, these teachers spoke of wanting feedback about why the student did not pass the exam, rather than only the number. For these teachers it took several weeks to get a handle on each student's writing ability. Since the WGE is given in the middle of the semester, teachers have a very small window to actually help improve students' writing before the exam. Teachers stated that even knowing some of the issues or reasons why each student didn't pass would help them to intervene earlier in the semester with their instruction.

Table Leader 4's comment, that she wanted to give two different scores to the essay, a five for fluency and a four for support, points to the limitation of holistic scoring to describe what a scorer valued in a text or the ability to provide a rationale for the score. No information is cycled back to the teachers about the students who are placed in their class, and simultaneously feedback to the district about their students' writing is truncated by holistic scoring. The question, again under the premise that the WGE should be used for placement, is if holistic scoring is the appropriate tool to use for this type of assessment? A different approach to scoring, like analytical scoring or primary trait scoring, could be a more useful pedagogical tool (albeit less efficient) to determine a student's ability to write at a "consistent level." In fact, the use of analytical scoring on the essays of students who fail the WGE might be particularly useful for the students and teachers as both would know the reasons the student failed the exam. By scoring traits independently or using an analytical rubric, the teachers could start from the first day of the semester knowing the needs of individual students to pass the WGE.

Disruptions like the one above, where scorers would remark on wanting to give a piece two separate scores for two different reasons, were the most frequent disruptions

observed. However, these moments would quickly be passed by due to the need to find consensus in training scorers, or as the scene above illustrated in table leader meetings, if consensus could not be determined, the paper was eliminated as a possible exemplar. Yet, because BPS uses their holistic scoring sessions to measure both proficiency and determine student placement, when questions over numerical scoring arise they provide a space to ask larger questions about the WGE process. Specifically, what is the purpose of the WGE, and what alternative types of assessment tools might provide useful for assessment and instruction?

Consequential Validity

The second disruption builds on the question of placement and the effects the WGE has on students who do not pass the exam. Again, this scene took place in a table leader meeting, where the actual score determined by the table leaders is awarded to the students. In the example above, the table leaders were split on giving the student a four or a five: either way the student passed. However, in this example the student will not pass and the table leaders debate if the paper in question should be awarded a score of a three.³⁰ Table Leader 1 provides the disruption in this scene which leads to questions of consequential validity, or the actions taken based on the scores of the exam.

QL: OK next batch of 3s. CCC, JJJ, OOO, PPP, UUU. What about CCC?

T3: It's kind of all over the place.

T1: I had this as a 2 (reads) "He____. He____. He____." It's kind of listy.

- QL: I think he sees that as support. An effort, but doesn't know how to do it.
- T4: I got stuck and it's simplistic.
- T3: And underdeveloped.
- T1: I look and say I need him in General English and not in Comp. I know I'm not following the rubric, but if he was in my Comp class others would be farther ahead and I don't know if I can get him to a four by the end of the semester.
- T4: I went back to the exemplars and 3 isn't chunky like this one.
- T3: If we are using the rubric and exemplars we have to go 3.
- T4: How's it a 3?
- QL: I think Mark is looking at the big picture, but we need to see it the way the rubric does.³¹
- T2: It's not confusing
- QL: (*Reads from rubric for a 2—limited response*) Overall, the response is undeveloped or confusing; it provides a limited explanation for an opinion or its importance.
- T4: How will a writing facilitator place this student?
- T1: If given a 3 he will be put in Comp. I look at the disconnectedness of the piece.
- T3: They are all things that happen at home.
- T1: But you have to infer that. But you have to go with what the group thinks.

QL: I think you are right. The student will do better in General English than Comp, but we need to score the piece of writing.

T3: I say this is a lower Comp kid. Teach him transitions and sentence variations and he's ready.

Table Leader 1, Mark, begins by asserting that he does not believe paper CCC is a three as he scored it a two. But Mark's justification for his score is not based on his reading of the paper against the rubric as a table leader should, but as a teacher familiar with what happens to students who do not pass the WGE. Mark is unsure if the student would be prepared for Composition, the course students are placed in by scoring a 3 on the WGE, stating that he would have difficulty getting the student to a four level by the end of the semester. Rather, Mark believes this student would be better suited by General English. Mark's disruption points to the issue of consequential validity. His unease is not about accurately placing a numerical score on the student's essay, but with what happens to the student after the test. Mark raises the issue of consequential validity, the action that is taken based on the score the student receives, as his rationale for scoring the paper a 2. While the student has not passed the expository essay portion—his score was lower than a 4—the difference between the three and the two will impact the curriculum, instruction, and overall educational experience the student receives.

The Question Leader, whose role is to lead the table leasers to consensus, first attempts to focus Mark away from his initial concern with the student or the “big picture” by pointing him to back to the rubric. By the end of this scene the Question Leader agrees with Mark that the best placement for this student would be in General English not in Composition, but she says, “We need to score the piece of writing.” It's an important

distinction of what exactly is being scored in this system of high-stakes assessment: the student as writer/learner versus the text. In this context, and despite Mark's concern, scoring the writing as a text divorced from the writer in an objective and scientific manner takes precedence. Mark has scored long enough to know he is supposed to use the rubric, as he admits in his first comments, and by the end concedes that "you have to go with what the group thinks." He disagreed with the group because of his concern on what happens to the student after the exam rather than the question of the "true score" of the text.

Mark's disruption is an excellent example of considering consequential validity, what Samuel Messick referred to as "the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores" (5). Within the system of the WGE, actions are taken to place the student in the class that will best help their writing development. However, the question of where to best place the student is not the task at hand in scoring the student's writing; rather the student's placement is the byproduct of the score. Advocating for the best educational action to be taken to help the student improve as a writer, Mark's disruption offers up the question, what are the social and educational impacts of the assessment on students?

Considering Construct Validity

The final disruption moves away from questioning the use of holistic assessment and the impact of the assessment on students to address issues of construct validity within the testing instrument. This scene took place at a scoring session of the expository component of the WGE. To maintain a high inter-rater reliability level, the Question

Leader and Table Leaders put the scorers through a series of practice scorings in order to help the scores see (and therefore read) for the differences between scores. All of the papers in the training session had been pre-scored by the table leaders, the goal being to assist the scorers in reading the papers in the same way the table leaders did. The training session started with the Question Leader reading the prompt out loud to the entire group of scorers.³² Essentially, the prompt asked students to write an expository essay on someone who has influenced them. The Question Leader then explained to the scorers, “For this prompt kids had to tell stories to support their writing and there is a fine line between too much narrative and not enough exposition. We’ll look at some examples in a little while.” The concern over too much narrative and not enough expository surfaced most frequently over the scores of three and four, an important divide between passing and failing the expository component of the WGE. Below is one example when the question of the use of narrative within an expository occurred:

QL: We are going to go to one more set before we start with live papers.

Go ahead and talk about this next batch

TL: On [paper] L we ranged from 3 to a 4+ and the table leaders saw it as a 3. If it was a narrative it would probably pass.

T3: There is too much narrative?

TL: It doesn’t do what an expository does. This tells a story and doesn’t have support and logic.

T3: This is what screwed me up in the 5s (*referring to the narrative papers that received a score of 5*). That’s what’s going to be hard.

Some are pretty close. Is there too much narrative or are they using it to support?

TL: And that is what I talked about before. We need to be looking for narrative that supports their point not that just tells a story. It (paper L) has a lack of control and it switches point of view.

T3: I saw that as stylistic, but in a narrative sense. I think control is present. The opening paragraph is weak but more stylistically as the voice changes.

TL: If you look holistically can you see that she talks about the person and the influence?

T4: I think it's adequate. Are we looking at writing or sophistication of thought?

QL: We need to read it as her trying to do expository writing. If we read it too narratively we aren't reading for expos.

T1: There is no process in here.

TL: The reader needs to make the connections.

QL: we read 80 papers and only a few were like this (meaning using too much narrative)

T4: I beg to differ

In the practice scoring session the table leaders scored paper L between a three and a high four. The difference in scores reflects the way the scorers interpreted the difference between narrative and expository writing and the use of narrative writing to serve as support in an expository essay. Thinking back to chapter one, the Harvard

Newton writing scale, the first response to the Hillegas Scale, argued that writing scales could not measure too complex of a product and were divided into separate scales according to the traditional modes of writing. Measuring separate and isolated modes of writing was continued throughout much of the twentieth century and is still present in the WGE with its narrative and expository essays. However, the prompt used for this expository test provided problems for students and scorers. In order for student to write about an individual who influenced them they often resorted to telling narratives to show or support their point about the way the person they selected influenced their life. This strategy caused students to blur modes of writing within their essays and provides a space to pose questions about construct validity.

At the end of the scene, the question leader reminds the table leaders and scorers that they should read the piece as the student attempting expository writing, reinforcing the construct validity of the exam, the definition and/or way the performance (writing) is constructed for the exam. The assumption is that expository essays have certain textual features that can be identified and scored by the teacher-readers. Consider the statement of the table leader, “It doesn’t do what an expository does. This tells a story and doesn’t have support and logic.” This brief comment begins to illustrate the different views of narrative and expository writing and the ways the task of writing is defined differently for each by the system of the WGE. In an attempt to have clearly defined modes of writing, a prompt like this one causes a fissure in the training of scorers out of which question on construct validity can be raised.

At issue within this episode of scoring are the traditional modes of writing which have long served as separate and distinct forms of writing within schools. The question

this prompt and scoring session prompts is if the modes are meaningful distinctions along which to divide writing. Or are the modes a false construction of what writers do? In classrooms teachers taught the modes of writing as separate, often defining narratives as linear and expository as giving an opinion on an issue or topic; however, outside of school this construction of writing does not exist. In George Hillocks' study on state writing assessment, he argues the traditional modes of writing truncate thinking due to the fact that students are taught that writing is formulaic rather than a way of thinking (201-203). The tension above over the use of too much versus enough narrative, as if it can be quantified within an expository text points to the way writing in schools and writing assessment has been constructed. Yet outside of schools, writers do not sit and decide on the mode of writing they will use prior to writing. Rather than build assessment systems around modes, might there be a more current construction of writing—such as rhetorical purpose—to define construct validity?

Starting With New Questions

Reliability has maintained a steadfast position in traditional psychometric approaches to large-scale assessment and inter-rater reliability has done the same in holistic scoring. The field of writing assessment has, over the past thirty-five years, focused most on concerns of test reliability and scorers' ability to accurately and consistently read essays. However, this blind faith in pursuing reliability has come at a cost to asking questions about the purposes and uses of assessment. Foucault's concept of discontinuity can be useful as it resists a homogenous reading of a disciplinary system and encourages observations at the micro-level in order to consider the ruptures and

fissures which are continually present. The three scenes above offer moments of rupture when teachers and table leaders momentarily resisted or critiqued the need to read according to the rubric itself. This resistance to the tools or the work of the system, however brief, serves as spaces of possibility out of which questions can be levied about the WGE, the larger system it creates, and the work of writing assessment more generally.

Because BPS replicated the Advanced Placement model of holistic scoring, it “borrowed” notions of reliability and inter-rater reliability founded in classical test theory. To conclude, I pose five questions, based on the scenes above, which together could serve as a larger heuristic for schools and districts to consider as they build a system of writing assessment. I suggest that starting with these questions rather than importing a model from the outside might serve the district in more productive ways.

1) What is the purpose of the assessment?

According to the district’s policy, the purpose of the WGE is to assure all students meet a minimum standard or level of competency. The purpose of the assessment tool is for accountability: to guarantee minimum skills have been met in order to graduate. Yet, these scenes illustrate how the WGE serves at least two purposes: measuring proficiency and placing students who have not yet met the proficiency within a curriculum. The question of purpose is important, as I illustrated how other forms of assessment could be used to determine if the student met the standard. In adopting the Advanced Placement format, the district implemented an assessment tool that was not designed for accountability but for placing students within a college curriculum. In a nutshell, the district repurposed the exam for its own needs of accountability and in doing so found a

way to justify placing students within a sequence of writing courses. In BPS's system the question still remains if the WGE is a valid assessment for both purposes, or if as a single instrument it can only accomplish one of the purposes.

Starting with the question of purpose the questions below begin to fall in line. Knowing the intended goals of doing assessment—accountability, programmatic assessment, classroom assessment—assist in framing questions of definitions of writing and what tools are needed to collect data to answer how well the purpose is being achieved.

2) *What is the definition of writing? What should students do / be able to do?*

Defining writing is a reoccurring problem in writing assessment and points back to the way the “writing problem” is framed. In chapter one, the “writing problem” was framed around teachers’ judgment of writing and their inability to accurately and efficiently grade writing. In chapter two, BPS’s “writing problem” reveals a perceived assessment threat from outside of the district and the “writing problems” that were solved were done so in retrospect. In both cases the definition of a writing problem, where the subject of the problem is writing, was absent. Critics, teachers, and policy makers often claim they want students to improve their writing, but what is meant by “improvement” is not articulated and when it isn’t, the assessments never move beyond grammar and editing or the traditional modes of writing. Furthermore, writing is often treated as an isolated skill separate from thinking (Hillocks 198-204; also note the scene above in which one of the table leaders above asked if they were to be scoring the writing or the thinking).

Therefore, the question of defining writing as both a noun and a verb is necessary when considering framing the writing problem and designing assessments for writing. Writing as a noun would ask, what are students producing? For the past century in large-scale writing assessment the modes of writing determined the product that would be measured. But are there other products of writing or ways to define writing that would allow for new or different products of writing to be created? Second, defining writing as a verb would lead to questions such as, what are students doing when they write? Here questions around thinking, problem solving, and revision could be addressed in classrooms and in assessment tools.

3) What kinds of assessment tools would achieve said purpose and produce data that corresponds to the previously stated definition of writing?

Earlier I explained how the Advanced Placement model of holistic scoring was constructed within the psychometric paradigm of reliability and validity. By importing this assessment tool, BPS also inherited a system dependent on achieving high levels of reliability. In chapter three, I briefly discussed how accountability systems allow for technical tinkering but prevent larger reform. That said, some teachers and administrators at BPS were interested in creating tools that would allow for an expanded definition of writing (n). The portfolio was the most common tool stated to provide a different type of assessment experience. (However, the same dangers of reliability exist within holistic scoring of portfolios; see White, Elbow, Broad all in Black et. al).

Assessment tools are not free from ideology (Turley and Gallagher).

Conversations in writing assessment, mainly at the post-secondary level have begun to splinter off the psychometric paradigm toward a more hermeneutical or interpretive

paradigm of writing assessment (Broad, Moss, Huot). This shift along theoretical lines also has implications for the ideology within the tools of writing assessment.

Considering new theories of writing assessment allows for data to be interpreted in new ways and for new purposes which could more closely align to the purposes and definitions of writing formed from the questions above.

4) What is the relationship between assessment and pedagogy? Does the data from the assessment tool prove useful to schools, teachers, students and parents?

Considering the types of tools to use in an assessment will determine what types of measurements and data will be collected, but how will that information be circulated back to the teachers and students in meaningful ways. While BPS does have a data base where teachers can see the score each of their students received on the WGE, teachers stated time and again that knowing why a students failed would be more useful for their instruction. Again, the WGE is supposed to be about accountability not assessment, and there for a summative approach, one could argue would be appropriate. But the larger question of the relationship between assessment and pedagogy is important to reimagine. Currently in BPS instruction and pedagogy are driven by the assessment, that is teachers do writing and create experiences that help prepare students for the exam. But what would a system look like where assessments are used as informative to teachers, which could help them revise their instruction and pedagogy rather than use it to serve the mandates of an exam?

5) What are the social and educational impacts of the assessment on students?

Mark's interruption above reminds us of the real consequences on students to the decisions that are made in the name of accountability and assessment. It seems necessary to ask this question of ethics and to study, at least informally, the ways systems of assessment and accountability impact students and their education. Recent conversations in validity theory, specifically around consequential validity, are beginning to ask such questions. Questioning the ways assessment tools reinforce social privilege, or track specific populations into remedial courses, need to be considered and addressed. As stated in question three, assessment tools carry with them ideological agendas which have historically and systematically worked to disenfranchise certain students while aiding others in moving seamlessly through an educational institution. Or what Linda McNeil calls the "new form of discrimination" (xxi).

Conclusion

Rather than view holistic scoring as a process solely seeking consensus and reliability, reading transcripts of table leaders and scorers for disruptions illuminates how teacher-scores will resist consensus and even critique the scoring process. The moments of fissure within the holistic scoring of student essays, illustrate the possibility of using teacher knowledge to, as Michael Williamson advocates, "lead to a different approach to assessment" centered on validity (167). The five questions gleaned from observing the BPS teachers in action suggests that determining a clearly defined purpose along with deciphering what tools and definitions of writing are necessary to address the social and educational actions that happen as a result of the decisions made by the system are paramount.

In the next chapter, I move from the district scoring sessions into teachers' classrooms to explore the ways two teachers' instruction complies to and resists prescriptively teaching writing as defined by the WGE rubric and exemplars. In this struggle of knowledge and judgment over the teaching of writing, I trace how teachers take their training as scorers into their classrooms and the ways the training to read according to the districts rubrics works in opposition to their professional knowledge of writing.

Compliment, Question, Critique: Repeat

After attending several district scoring sessions and table leader meetings, I found myself having a reoccurring conversation with teachers who were not part of my study. The first couple of times it happened I didn't pay any attention to the interactions; it wasn't until I read through my field notes that I noticed this pattern. Typically, on a break, over lunch, at the snack table, or in the restroom, a teacher from one of the high schools I was not studying would approach me. Having introduced myself and briefly explained why I was observing at the start of every meeting, I first thought of the interactions as friendly and welcoming, a kind of Midwestern chit-chat. But after the fifth or sixth time having the same conversation, and starting to feel like Bill Murray in Groundhog's Day, I began to wonder if something else might be going on.

Typically the conversations would start with a brief introduction and a hand shake followed by the teacher complimenting Butler Public School's or the Writing Graduation Exam. The teacher would express how BPS has always "been a head of the curve with standardized testing." "We've been doing this for a while now." "Nationally, we're usually on the cutting edge of educational trends." And while this first statement was usually said with a smile the conversation then turned to me. The teacher would ask my thoughts about the WGE, "Is it good a good test?" "Do you think we are assessing right?" "What are your impressions so far?" To which I would politely decline to comment by saying that's what my study will try to figure out. Then they would ask me about school accountability and standardized tests, "so where are you on all of this testing stuff?" "Do you think all these tests will ever go away?" "What do you think

about NCLB?” I would typically say I’m not really for all of it, or that I think there are some real dangers with only using high stakes tests as indicators of school and student success. Finally, as the conversation drew to an end the tenor shifted and the teacher would leave me with a critique of the WGE. “You know I think I was a better teacher before the WGE.” “I used to teach at Valley Public Schools, they didn’t have a district test. I was treated more like a professional there.” “All I do with my low end kids is test prep, that’s not making them better writers.” Following their critique the teachers would usually wish me well, or give a word of support about my study.

It’s hard to know how much to read into these reoccurring conversations, the motives of the teachers who approached me, and even the substance of the conversations. But each time I had the conversation it brought me back to why I started this project and the tension I felt in my teaching over what to do about standardized writing tests.

* * *

Chapter 4

Complexities in Teachers' Judgments of Writing

English 9 Regular English class for freshmen. Assists students in meeting state standards and graduation demonstration exams.

General English The class is designed to give support to students who have not yet passed the Writing Graduation Demonstration Exam by focusing on the writing skills that are essential to passing the WGE.

Composition Through study and guided practice, students focus their reading and writing on narration, exposition, and persuasion...this course may serve as an alternative to the WGE.

From the course offerings booklets at Wilson and Marshall High Schools

“In every state, when teachers have little knowledge of writing, the testing system tends to become the knowledge base for teaching writing”

George Hillocks Jr. The Testing Trap

In chapter three, I examined disruptions within Butler Public School District (BPS) holistic scoring sessions to explore questions about writing assessment and students that are often ignored in an attempt to produce a reliable scoring of student essays. BPS employs their teachers as test scorers once a semester. In this chapter, I turn to explore the impact of the WGE on the classroom level, specifically the ways the WGE shapes teachers' judgments of writing as they work with their students. The prominent placement of the WGE in the course descriptions of English 9, General English, and Composition is one indication of how the WGE filters from the district into classrooms, serving as the authority for teaching writing. In naming the modes of writing and skills needed to pass the exam within the course descriptions, the WGE is reified as the writing curriculum teachers need to teach to ensure students are provided opportunities to pass.

Therefore, it might not be a surprise to observe teachers acquiesce their classrooms to the WGE; however, to describe teacher judgment of writing as simply “teaching to the test” is an oversimplification. In fact, when teachers talk about the kinds of writing they assign, instruct, and assess, fissures arise around their classrooms, pedagogies, and the WGE. These fissures speak to an unexamined but in this study pervasive misalignment between the WGE and the teachers’ professional judgment.

In The Testing Trap, George Hillocks examines large-scale writing assessment systems of five states—Texas, Illinois, New York, Kentucky and Oregon. Though only Texas uses state writing assessments for exit examinations, Hillocks’ findings regarding what happens to writing in schools under state-mandated standardized writing assessments are grim. Hillocks contends that when teachers and administrators lack a solid knowledge base for teaching writing or a clearly defined writing program established in their schools, the state-mandated writing assessments drive writing instruction within schools (102). That is, teacher judgment on what kinds of writing occur in schools and the qualities valued in writing are decided by the standardized tests students are required to take. Moreover, Hillocks argues that the theories of writing and learning underpinning the state tests become the theories of writing teachers carry with them into their classrooms (198).

Due to the high-stakes nature of the WGE and the system of management used to account for student writing proficiency, the driving force of teacher judgment of writing in BPS is more nuanced than the exam itself. I want to extend Hillocks’ argument to the micro-level in order to explore how agency, disciplinary knowledge, and self-disciplining govern teachers’ judgment of writing in their classrooms. Michel Foucault’s work on

power and disciplinary knowledge provides an analytical lens to consider how the technologies of writing assessment manage teachers' decision-making process in the classroom, particularly with regard to the types of writing teachers assign their students, how they instruct students, and how they assess writing in their classrooms.

In Discipline and Punish, Michel Foucault examines the ways power is disseminated within disciplinary systems (schools, prisons, hospitals) not in overt and grand ways but through discreet and micro-physical mechanisms of power. These mechanisms of power rely on the dispersal of bodies within a hierarchy where power is channeled through instruments and interventions used to remove agency from the body and ascribe agency and power to the system the body serves. These mechanisms of power are intended to produce docile bodies, ones that might be subjected, used, transformed, and improved according to and in service of a specific form of disciplinary knowledge (DP 136). Thus, disciplinary knowledge gets enacted through a “mechanics of power” which “define[s] how one may have a hold over others' bodies, not only so that they may do what one wishes, but so that they may operate as one wishes, with the techniques, the speed and the efficiency that one determines” (138).

Systematic discipline results from the manipulation of docile bodies to conform to a prescribed disciplinary knowledge. Foucault summarizes this effect: “In short, [discipline] dissociates power from the body; on the one hand, it turns it into an ‘aptitude’, a ‘capacity’, which it seeks to increase; on the other hand it reverses the course of energy, the power that might result from it, and turns it into a relation of strict subjection” (138). For Foucault, disciplinary knowledge is acquired within a system of power. As an individual is disciplined, s/he gains aptitude or knowledge; however, the

power that comes with the acquisition of the aptitude is restrained to a specific, subordinate role within the hierarchical system.

In this chapter I will begin by illustrating how teachers do “teach to the test” but not through the adoption of a theory of writing as Hillocks claims. While Hillocks maintains the test is the knowledge and theoretical base and for teachers, I argue the foundation of knowledge for teacher judgments is found in smaller technologies that comprise the WGE exam such as rubrics and scoring sessions. I will trace the way BPS’s assessment tools discipline teachers’ judgments as well as students’ writing. While a deterministic reading of Foucault could argue teachers are merely docile bodies, lacking agency within the system, this would not adequately explain the complexity with which teachers talk about teaching writing. Therefore after a reading of teachers’ judgment through the lens of Foucault’s notions of discipline, I provide a second reading which seeks to excavate teachers’ subjugated knowledges of writing. Through interviews, it became clear that teachers were aware of the choices they made with respect to writing and were not always happy with the choices or the results. Conversations with teachers revealed gaps and fissures between their classroom practices and their knowledge of writing pedagogy. Therefore, using Foucault’s methodology of reading discontinuity onto a unified system, the second half of the chapter provides an alternative reading of teacher judgment from one that suggests teacher judgment is simply disciplined by the “knowledge base” of the WGE to explore how teachers’ knowledge of writing and student learning becomes subjugated by the system of the WGE. Through this conflict of knowledge/theory, teachers neither fully submit to the WGE nor do they fully enact

their professional knowledge of writing. Rather, they manage their instruction, in-class assessment, classrooms and students between incongruent theories of writing instruction.

To consider how the technologies of the WGE impact teachers' judgment along with how teachers question and critique their own classrooms and teaching, I present two teachers. Josh teaches English 9, a freshman English class at Wilson High School. I spent a semester and a half in Josh's classroom. Because Wilson is on a block schedule, I was able to see him teach a year's curriculum to one class of students and the equivalent of a semester to a second class. Josh is a third year teacher at Wilson and student-taught at another high school in the district. Laurie has taught English for thirty-one years, the first eleven in another district and the last twenty years at Marshall High School. I observed her over three quarters of the year teaching Composition to mostly seniors, some of them fifth-year, and a few juniors. I chose to represent these teachers because they provide a contrast of context in the age of their students, the schools they teach in, and their experience as teachers; but at the same time they demonstrate how the WGE permeates the way teachers in BPS approach the teaching of narrative and expository writing.

Agency and the Assigning of Writing

In this study, teachers were more rigid in their approaches to teaching narrative and expository writing in courses in which students had not yet passed the WGE. In chapter two, I explored the "pressure/freedom" dynamic: the immense pressure teachers felt in courses in which the main concern was getting students to pass the WGE versus the freedom teachers felt in courses in which students had already passed the test or in

which getting them to pass it was not an immediate concern. Teaching ninth graders, Josh mainly focused on narrative and expository writing but also incorporated writing activities that provided creative outlets and experimentation for students. He was aware of the pressure of the exam on his students, which will be illustrated below, but since his students were first-time takers of the WGE, they still had three years of instruction and practice to complete the graduation requirement. However, teachers such as Laurie, working with students in General English and Composition, felt an increased pressure for their students to pass the exam, resulting in all in-class writing focused on narrative and expository writing.

In our first interview, I asked Josh what kinds of writing students do at Wilson High School. His response is noteworthy for the way he attributes agency to the WGE as shaping how he approaches teaching writing in his classroom:

First let's talk about what we are told we are supposed to teach for writing. Narrative writing is one of the focuses; you know, a story with dialogue, description, and action--those types of elements. And then expository with details, support, organization, and thesis--those are the two main ones that come to mind right away because of the writing graduation exam that they are supposed to take.

The use of pronouns in the quote illustrates the tension Josh experiences in his role as a writing teacher. "We are told" references Josh's compliance to an unnamed authority controlling the decisions about what kinds of writing are to be taught by writing teachers in English 9—narrative and expository. Furthermore "we are told" emphasizes Josh's limited agency as a teacher to change what he teaches due to "the writing graduation

exam that they are suppose to take.” The students (“they”) who take the WGE need preparation for the examination. Josh positions himself between the students who have to take the test and the school board policy that mandates the test. The WGE is both the impetus and definition for teaching writing, Josh serves as the vehicle to teach writing, and the students perform a prescribed form of writing to demonstrate their skill. Josh’s use of pronouns suggests that agency resides in the hierarchical system of the WGE where the roles of people within the school system are defined.

Within the quote Josh not only defines his role in relation to the system, but also outlines what he is “supposed to teach.” Using language from the district’s WGE holistic scoring rubric, Josh defines both narrative and expository writing. The rubric names three textual characteristics for both narrative and expository writing: narrative writing is defined as writing with action, description, and dialogue, while expository writing explains, describes, and informs. Josh describes the WGE as providing “guidelines” or “fram[ing] the choices” he makes when he thinks about teaching writing. In his English 9 class, the major writing projects are a narrative writing project and an expository essay, while the minor writing assignments are used to reinforce the qualities valued in the WGE such as action, description, and dialogue or structural features like organizing an expository essay.

After a semester observing Laurie instruct her Composition students writing narratives and expository essays, I asked her if the Writing Graduation Demonstration Exam impacts her teaching in composition. Her answer:

I think it does tremendously. I just think that is what my whole focus is and I’m actually embarrassed about that because I know how important it

is to these kids. Especially the seniors. Like Jennifer, I've worked with her now, this is the fourth Comp class she's been in [and] all she is worried about is the WGE. [...] Right now all the kids in comp classes are just kids who have not passed. And so I think it influences me tremendously because if you don't concentrate on [the WGE] they might not graduate. So I try to do three expository and three narratives [in a semester]. I think it just totally has taken over my [teaching]. Like you saw yesterday, I talk about the WGE, trying to help them because I want to help them graduate and I know it is important.

Laurie professes her "embarrassment" that the WGE has become her "whole focus." Like Josh, she removes agency from herself to make a professional judgment, ascribing agency instead to the WGE: "I think it just totally has taken over my teaching." But the reason she provides for only teaching narrative and expository writing in Composition focuses on her students rather than the structure of the district. While Josh's quote illustrates the top-down direction of power establishing a hierarchy placing teachers between the district and the students, Laurie uses her student's situation as the rationale for giving over her professional agency and only focusing on expository and narrative writing. As the teacher of Composition, Laurie establishes her role as providing one of the last opportunities for her students to secure their diploma. The student Laurie named in the quote, Jennifer, was placed in Composition after English 10 and for the past two years has remained in Composition unable to pass the WGE or the alternative assessment. A second semester senior in Composition, Jennifer will not receive her diploma if she

does not pass the WGE. Out of an obligation to “help them graduate,” Laurie turns to the modes of writing on the WGE as work she requires of her students’ in-class writing.

The WGE serves as the foundation of knowledge with respect to what Josh and Laurie teach in their classrooms. For Josh, the demands of the WGE, the district, and the school board are forces shaping his judgment of what to teach in his classroom and demonstrates the systematization of power at work within the district and how it impacts him as a teacher. Laurie focuses on her students’ need to graduate and therefore only teaches the modes of writing that the students need to get them their diploma. Evident in both examples is the process of removing agency from themselves and placing agency onto the system that works to control them and their students.

Stopping here would provide an analysis of teachers’ judgment along the lines of Hillocks and others who have argued standardized tests result in teachers teaching to the test. However, Hillocks’ argument does not adequately explain the findings of this study—in fact, it does not go deep enough. While Hillocks interviewed and surveyed teachers he did not spend time in the classrooms with the teachers he interviewed. Spending the school year with BPS teachers provided me the opportunity to compare what happened in the classroom with the ways teachers describe their teaching in the interviews.

Furthermore, in the second round of interviews I was able to ask teachers to reflect more specifically on classroom issues. Through this study I extend Hillock’s argument that the test is simply a knowledge base for teaching writing by illustrating the way power functions systematically within BPS to inform teachers’ judgment of writing in their classrooms; then, I illuminate fissures in Josh and Laurie’s judgments of writing by attending to their subjugated knowledges. Displaying the disciplined knowledge and

subjugated knowledge showcases the ways these knowledges compete within Josh and Laurie's practice of teaching writing in their classrooms.

Disciplinary Knowledge and Writing Instruction

Considering Foucault's discussion of the relationship of power to disciplinary knowledge, I now want to examine how Josh and Laurie take up the disciplinary knowledge of the WGE as part of their writing instruction. Both Josh and Laurie serve as exam scorers for the WGE; twice a year they spend a Saturday using the district rubric and exemplars to judge the quality of student exams. Through the process of scoring Josh and Laurie are able to see how the assessment technologies are read and used to score a paper. The scoring session provides a growth in aptitude or capacity as a writing instructor, gaining knowledge on what to instruct their students on in order for them to pass the WGE. With increased disciplinary knowledge through scoring sessions, Josh and Laurie teach the textual techniques and values promoted by the WGE while simultaneously limiting their professional agency to make decisions about how they teach writing. Various classroom moments demonstrate how Josh and Laurie defer their agency in their classroom based on the knowledge they gained in the scoring sessions of the WGE.

In his first quote above, Josh went beyond saying he taught narrative and expository writing to define narrative according to three descriptors—action, description, and dialogue. The unnamed authority determining what kinds of writing are taught in BPS is also responsible for defining the modes of writing in very specific ways. Yet, the

decision to teach narrative writing as action, description, and dialogue comes from Josh's experience as a scorer of the WGE.

When I'm scoring the exams I really saw that dialogue, description, and action are the things that are really lacking and those things really keep [a paper] from being a four or a five or even a six. So that's the language I use in my 9 English classes. [A] narrative is dialogue, description, and action and we really focus on those things.

Josh explains how scoring has "informed" his classroom instruction:

Being able to look at the rubrics and actually doing the scoring for the district has really helped me. [...] Actually sitting down and reading and looking and scoring [the exams] and getting that practice that has helped inform my classroom instruction and what I look for.

With the knowledge gleaned from scoring sessions, Josh returns to his classroom better aware of the instruction he needs to provide his students as well as what he needs to "look for" in his students writing to help them pass the exam. Through scoring sessions Josh has gained knowledge of what he needs to do to succeed as a teacher within the system of the WGE. Josh presents narrative writing to his students in the same prescriptive manner it was presented to him. Because students are rewarded for exhibiting the textual features the WGE rubric calls for, Josh passes the disciplinary knowledge he received from the district's technologies of writing assessment (rubrics, exemplars, and scoring sessions) to his students.

Below is an excerpt from Josh's English 9 class where he introduced dialogue to his students. Leading up to this particular day, the students brainstormed possible topics,

created brief outlines of the plot, and wrote a draft of their narrative. After students had a first draft, Josh taught a lesson on incorporating description using all five senses into their writing. Josh asked his students to go through their writing and find places where they could describe the scene or the action that was taking place. Next, Josh focused on the final “D” of the ADD acronym—dialogue. On this particular day, Josh demonstrated how to write and properly punctuate dialogue. In class he presented a PowerPoint consisting of four slides, each with a rule for writing dialogue. The four rules of writing dialogue consisted of: 1) Use quotation marks to indicate words which are spoken by characters. 2) Always start a new paragraph when changing speakers. 3) Make sure the reader knows who is speaking. 4) Use correct punctuation, capitalization and spacing. An example was used with each new rule in order to demonstrate what the dialogue should look like:

We were all in the parking lot waiting to go to the store.

“Hurry Up,” Saeed ordered.

“Go away!” Tanesha yelled.

“What’s the problem?” Carlos asked.

Angie replied, “I’m just tired.”

After presenting the brief PowerPoint, Josh informed the class, “I’m going to expect you to use dialogue like this in your narratives. Look at your piece and find places where you can use dialogue in your own story. And it’s not enough just to add two lines. You need to have a conversation.” Josh asked his students to take out their papers, mark in the margin places they might add in some dialogue, and imagine the dialogue they might add. The students followed Josh’s directions and incorporated dialogue into their final drafts.

Josh presented a clear lesson on the rules and punctuation of dialogue, but absent from this lesson was a discussion of why students might use dialogue in a narrative piece, or how dialogue can impact a reader's experience of a narrative or aesthetic qualities of dialogue within a text. Additionally, there was no attempt to connect writing dialogue to class readings, allowing students to explore when and how dialogue was used in the novel they just completed, The Secret Life of Bees. Discussing the purposes of dialogue becomes a moot point due to the demands of the WGE, resulting in the codification of writing where teachers present the values of the rubric as rules and absolutes.

Josh was not the only teacher whose instruction, specifically the teaching of dialogue, was influenced by the experience of scoring. Laurie, too, addressed how her instruction of dialogue was based on its value in scoring sessions. "I've noticed when we score these papers on Saturdays, if the kids don't use dialogue in their papers, sometimes they get scored down. So I do try to teach kids how to do dialogue." The rationale to teach dialogue stems from Josh and Laurie's experience of seeing students rewarded with a high score for using dialogue in their writing on WGE narrative exams; since the students in their classrooms have not yet passed the exam, they teach dialogue.

However, textual features such as dialogue are not the only component of writing being disciplined through the WGE. An important part of expository writing is for the student to "demonstrate control as a writer." Often in scoring sessions, table leaders will ask the readers to consider if the "writer controls the writing or if the writing is in control of the writer." Essentially, the question being asked is, can the writer remain on topic and focused throughout the entire essay? If the writer goes off on a tangent or presents contradicting support to a claim, she or he is not demonstrating control and will not

receive a passing score. Laurie explains how control is valued in the technologies of writing assessment affects her teaching: “I see myself telling them things like, ‘One of the big things is kids will change their mind in an expository.’” To combat students demonstrating a lack of control in their expository writing, as part of her instruction, Laurie continually reminded her students not to change their minds as they wrote their expository essays. The brief scene below was one example of this instruction.

Laurie: Today we begin expository writing. Who can tell me what is it?

Marissa: Narrative tells a story, expository tells opinion.

Laurie: And I’ve told you the biggest mistake you can make on writing an expository essay is--

Nathan: Don’t change your mind.

Laurie: Right, don’t change your opinion half way through.

Camille: What if you can say yes and no to the question?

Laurie: Don’t do it. For the exam they want to see you can answer the question with one answer.

Jasmine: And you need to be careful not to make it too much like a persuasive

Laurie: How do you do this?

Jasmine: Don’t be too pushy and tell the reader they are wrong or try to get them to change their mind just explain what they need to know.

Laurie: Right, Jasmine. To be honest I think expos is easier than narrative.

On the board Laurie outlines the five paragraph essay:

Laurie: It's just five paragraphs. The intro is where you state your opinion and then you list your reasons with support in paragraphs two through four and then in your conclusion you restate your intro. We will have a form to fill out tomorrow to help us with this.

Interestingly, this moment of instruction was complicated by what happened immediately afterwards. Laurie passed out a handout with a series of prompts about gender roles in society. While her students didn't know it at the time, they would be asked to take up one of the prompts for their expository essay. Laurie read a prompt and then invited students to dialogue as a way to brainstorm the ways they might respond to the prompt. The response that sparked the largest debate was to the question, "Can men and women be friends without sexual attraction?" The students came alive. One male asserted that it is impossible. Camille (from above) said that she thought it's easier for girls than guys. After Camille stated her thoughts, Laurie asked her some follow-up questions. The questions pressed on Camille's original answer. Camille then told a story about a friendship she had with a male and how over time, physical attraction became a problem (because she also had a boyfriend). Camille then changed her mind, revising her first answer, saying, "Maybe it is as hard for women and men and that friendships always have some sort of physical attraction." The conversation went on for a few more moments and by the end Camille had again switched her position.

When Laurie instructed students on the rules of writing an expository essay, the students were told to commit to one idea, but in a class discussion Laurie encouraged her

students to think past their initial response. In doing so, Camille reconsidered her position twice and might have again if more time was spent hearing other students' opinions. The message sent is thinking and writing are two separate processes—think first, then write. This issue of changing opinions while writing came up two days later when Laurie was revisiting expository writing.

Laurie: Now I want to switch gears. Take out a piece of paper. OK, exposition. You have to have an opinion and what did I say?

Marcus: Don't change your opinion.

Laurie: Right.

Camille: That happened to me last night. Half way through I started to change and then I was like no and I "X"ed it out.

Laurie: Tell me about that.

Camille: I just knew I couldn't do that. I just started thinking differently, but I knew I couldn't.

Laurie: You know why that happens? Your brain is so powerful it just does that to you subconsciously. We use such a little bit of this big brain but don't let it do that. Keep the same opinion all the way through.

At work in these two classroom moments is another form of disciplinary knowledge being transferred from a scoring sessions to a classroom. Again, writing instruction is presented by the teacher to the student as a rule—do not change your mind. Doing so would result in the production of uncontrolled writing and the failure of the exam. The values of the WGE rubric serve as the basis for Laurie to determine how she

ought to instruct her students in expository writing. Because the WGE does not offer an opportunity to revise, it is important that students commit to an answer and follow through within the 45 minutes they are allotted to write. Therefore the knowledge Laurie gains as a scorer pervades her classroom and the ways she instructs students to think and write expository essays.

Returning to Foucault's mechanics of power, Josh and Laurie's experiences as WGE scorers serves as an aptitude or capacity whereby they are trained to read for prescriptive textual features of writing. Their utility as scorers depends on their ability to read narrative and expository writing according to a prescribed definition set by the technologies of the WGE—the rubrics and exemplars. The acquisition of a discourse for writing from the scoring sessions becomes a component of Josh's and Laurie's professional development—or in Foucaultian terms, Josh and Laurie are being normed and disciplined by the system. Because Josh and Laurie understand how the district rubric is interpreted within scoring sessions, they are able to bring that knowledge back to their students and their classrooms. Josh specifically focuses on dialogue, while Laurie makes sure her students focus on one answer to the prompt. This process of transferring the discourse of the scoring session into their classrooms illustrates how the rubric and scoring sessions serve as instruments and interventions through which disciplinary knowledge is passed from the district to the teachers to the students.

Self-discipline and Classroom Assessment

For Foucault, the acquisition of disciplinary knowledge is only complete when the body takes on the position of a subject internalizing the power and regulating him/herself

within the system. The acquisition of disciplinary knowledge and the process of self-disciplining of a subject are not distinct processes; in fact, they happen simultaneously through the mechanisms of power to continually act on bodies. The very actions, attitudes, and discourses used by the individual replicate and reify the system of power as it is enacted onto the individual and through the individual's actions. Foucault reflects on this internalization of power in an interview with J.-J. Brochier: "...in thinking of the mechanism of power, I am thinking rather of its capillary form of existence, the point where power reaches into the very grain of individuals, touches their bodies and inserts itself into their actions and attitudes, their discourses, learning process and everyday lives" (30). In relation to the teachers of BPS, self-discipline is the internalization of power when disciplinary knowledge filters into the actions and judgments of the individual teachers as a means of self-regulation. Put more concretely, there is not an authority figure in the teachers' classrooms telling them how to teach and assess writing, nor surveilling their classroom practices from the outside; but the teachers have internalized the WGE as the disciplinary knowledge of writing and therefore rely on it to make judgments in their classroom around writing.

Scoring WGE exams is dependent on a teacher's ability to read the rubric and the exemplars in order to make a judgment about a student's paper. For Josh, the act of judgment that occurs within the district scoring session does not remain in the scoring context; rather it is also his classroom assessment practice. Josh elaborates on his classroom assessment practices:

I break down my rubrics into the form they are looking for in the WGE.

So for example [for] narrative, there's a section on dialogue, I know that is

one of the things that we may look for when we are scoring the exam. Or with narrative it might be, “Is there action?” You know, is there detail, is there voice? So in my rubric, I will break it down in those areas. Is there an opening, is there dialogue, is there action, is there description, standard edited English, all of those things we are reading for holistically? [...] I really try to break down those things in hopes that if students can demonstrate some success in each of those areas, when they take the exam, they will do better.

Josh “breaks down” his classroom rubric into seven categories conforming to the major components of the districts rubric “they,” the scorers of the narrative exam look for.

While no one tells Josh he must create a rubric that mirrors the WGE rubric in his classroom, he has internalized the values of the system used for scoring (see Appendix E).

The classroom rubric Josh uses to grade narrative writing focuses on seven qualities, each of which is awarded a set number of points: mechanics 15, opening 10, closing 10, dialogue 15, description 15, action 15, writing process 20, for a total of 100 points. While the process category is worth the most points, the first six categories specifically reference the values on the district rubric. (In the following section I will take up the process section of the rubric as a moment of fissure and evidence of Josh’s subjugated knowledge.) When WGE scorers describe their reading practices for the WGE, teachers most often cited grammar and mechanics, structure and organization, and the textual features of action, dialogue, and description as the qualities they look for in a passing narrative. Josh’s classroom rubric, is a direct reflection of the internalization of

values of the WGE, keeping his classroom assessment, himself and his students subject to the values of the system.

It's important to remember why Josh replicates the district exam in his classroom. In the quote above he mentions his hope that in doing so his students will be better prepared—they will write using the techniques and characteristics of narrative valued by the WGE. The role of the WGE in his classroom assessment illustrates not only the pressure he feels to help his students pass, but also the way Josh has acceded to the values of the system where he is conscious to align his instruction and classroom assessment with the values of the WGE.

Laurie's classroom situation is different from Josh's. BPS has designated Composition courses as the only course where students who have not yet passed the WGE can turn in a portfolio as an alternative way to meet the WGE requirements. The portfolio may be misnamed in that it is not a collection of work, but refers to the fact that students can work and revise a single piece of writing until it reaches the score of a 4 or higher. There are, however, rules that accompany this process. For example, on the pieces of writing a student submits for the alternative assessment, a student may revise as many times as they want to; however, the teacher cannot provide any feedback besides a score that corresponds to the district's holistic rubric. In fact, students in Laurie's course have to sign a "Verification of Unassisted Writing," which states that the student "attests that the writing I am asked to do as 'unassisted' to meet these requirements is my own writing and will not have been responded to or edited by any other person than myself."

The goal of having students' alternative assessments written independently goes back to notions of reliability, or regulating the conditions under which writing is done

that counts for the WGE, as well as validity, ensuring the student's writing ability is measured, not his teacher's teaching. The idea behind the district's alternative assessment practice is that teachers should have their students writing and working on the skills they need to refine, while periodically allowing students to work independently on writing they can submit as their alternative assessment. Laurie's students are struggling writers who take a long time to compose a sentence, let alone an entire essay. The slower pace of her struggling students becomes an issue because she wants her students to do multiple drafts to improve their essays but does not feel there is time for independent writing that can be used for the alternative assessment.

Faced with the pressure of time and need for student success, Laurie has abandoned grades as forms of feedback. Rather, she typically returns students' drafts with a sticky-note on which she has written a score for the essay. "When they write a paper for me in Comp, it's not [out of] 100 points. I use [the district] rubric just so they get use to the one that's going to be used on the final thing, again teaching to the test, but that is how I score comp. In my grade book, it's a 1-6 so that is how I've been assessing it." She will also sometimes have a short list of items for the students to work on (opening, transitions, punctuation, examples, etc). This short list of ideas for revision may violate the policy of not providing feedback to student, but it also raises the question of the role of the teacher in this situation. Using the alternative assessment in her classroom, the WGE defines Laurie's role solely as a scorer not a teacher. The policy's intent is that students will receive feedback and instruction on certain writing projects and an objective holistic score on ones for their portfolio. Constrained by time, Laurie uses the holistic scoring approach on all the essays written in her classroom.

The high-stakes nature of the exam for their students results in both Josh and Laurie giving over their classroom assessment practice to the systematic judgment of the WGE while simultaneously internalizing the values of the system and regulating their classroom assessment against the prescribed values. In doing so, they revert to the scoring sessions and district rubrics at the source of disciplinary knowledge. Josh breaks down the holistic rubric into individual pieces to analyze students' narratives while Laurie imports the holistic rubric and only assesses her students' writing holistically. The self-disciplining nature of power within the system is made evident as the systematic power of the WGE becomes capillary, impacting the judgment of how Josh and Laurie assess writing within their classrooms.

Chris Gallagher and I have written about the problems that attend moving an assessment tool from one context to another. In the example of Josh and Laurie, a holistic rubric designed for large-scale assessment is transported into a pedagogical classroom context. The "law of distal diminishment" claims "any educational tool becomes less instructionally useful, and more potentially damaging to educational integrity, the farther away from the classroom it originates or travels" (Turley and Gallagher). The point is that solid assessment tools perform specific functions within specific contexts. To uproot a tool from one context and transplant it into a different one will not provide for useful for instruction, assessment or the improvement of student writing; in fact, it can even be a dangerous practice to ignore the ideology embedded within rubrics. However, Josh and Laurie illustrate how easy it is to either directly transport a rubric or to create one that replicates the qualities and ideology of a different context.

Up to this point, I have drawn on interviews and classroom observations of Josh and Laurie to illustrate how the technologies of the WGE serve as mechanisms of power which creep into teacher's classrooms and their judgment with respect to teaching writing. But what is also very clear in talking to teachers is they regularly question and critique their classroom practices against what the WGE requires of them. Through my interviews with teachers, it became clear that there are gaps and fissures in the WGE system, surfacing tensions teachers have with the WGE. In fact, as Josh and Laurie talk about their classrooms, they discuss the tensions they have in negotiating their theories of writing and commitments as writing teachers while simultaneously being accountable to the WGE system. Borrowing from Foucault's theory of discontinuity—one he applied to the discipline of history—I will complicate my first reading of agency, power, and self discipline by turning to teachers discussions of their subjugated knowledges of writing.

Genealogy and the Subjugated Knowledges of Teachers' Judgment of Writing

The complexity of studying teacher judgment in relation to writing comes from actual conversations I had with teachers about their teaching of writing, coupled with classroom observations. As an observer in teachers' classrooms, I found the mechanisms of power were clearly present in the ways they assigned, instructed, and assessed writing; however, through interviews and conversations in the hallways I also found that the reading of teachers acquiescing their professional judgment and agency to the system did not fully capture the dynamics of teacher judgment. In fact, teachers often questioned and critiqued their classroom practices, resulting in moments where their questioning and

critiquing the WGE offered moments where they articulated their own professional judgment and agency.

Again, I turn to Foucault to make sense of these moments. In chapter two I took up Foucault's use of subjugated knowledges as a means to elaborate on teachers and administrators knowledges about the ways the WGE system functions from within. Here I will again use Foucault's subjugated knowledges to focus on knowledges teachers have of writing that are not recognized by the disciplinary knowledge of writing discussed above. To review, Foucault defined subjugated knowledges as two-fold:

on the one hand, I am referring to the historical contents that have been buried and disguised in a functionalist coherence or formal systemization. [...] On the other hand [I mean] a whole set of knowledges that have been disqualified as inadequate to their task or insufficiently elaborated: naïve knowledges located low down on the hierarchy, beneath the required level of cognition or scientificity. (82)

It is the second definition of subjugated knowledges that I am interested using here, particularly knowledges that have been omitted or ignored and replaced through a scientific knowledge of measuring writing. Foucault offers the means to study subjugated knowledges through the recovery of the "local character of criticism" that results in "an autonomous, non-centralised kind of theoretical production, [...] whose validity is not dependent on the approval of the established regimes of thought" (81). Or, the knowledges that exist within or between people despite an attempt to devalue or erase it through the construction of an official disciplinary knowledge. What Foucault argues

for, then, is a listening to the local knowledges against a unified theory which assigns value and rank based on its relation to disciplinary knowledge.

As I have argued in the sections above, the WGE serves as a disciplinary system: providing a theory and knowledge base for writing, systematically removing teacher agency and judgment, all the while controlling the production of student writing. Using Foucault's subjugated knowledges to illustrate some of the discontinuities or ruptures helps to complicate Josh and Laurie's judgment of writing. Listening to Josh and Laurie talk about the tensions they have teaching under the WGE reveals subjugated knowledges that are disqualified and ignored by the unified theory of writing established by the system of the WGE.

During interviews with the teachers and district administrators of Butler Public Schools, their commitment to versions of process pedagogy emerged, yet the WGE asks students to write two essays in a single sitting. Knowing their students will not have opportunities to revise on the WGE, teachers who value process writing in their classroom face a dilemma. Josh addresses this tension as he describes two different ways he could teach writing: the first along the lines of his pedagogical judgment and the second as strictly a disciplinary act to prepare the students for the WGE:

I would love a two week period for a student to write a process piece. Ideas, first draft, second draft, peer editing, conference with me, revisions, you know all of that. I would love to take two weeks just to do that...Or I could have kids practice a timed piece of writing multiple times. Here's a narrative prompt, write a narrative in forty-five minutes. OK, let's go back and look at that, what were you able to do in 45 minutes? Is it a 3 or

a 4 or a 5? What did you do well? What can you do better? How do you need to organize that forty-five minutes better? Ok, let's do it again. Here is another narrative prompt write it in forty-five minutes. How did you improve, what did you do differently, same thing for expository. Here's an expository prompt. Write it. Time's up. And I could do that for two weeks. Which one's more valuable? I don't know because they're two different processes of writing, two different methods of writing, for two different purposes. So that's the tension I feel. Because there is a part of me that says, you're teaching to the test, that's not writing. And maybe that is writing and so I try to do both. But there is also a part of me that says if you really want to grow as a writer you need to draft, you need to revise, you need to talk about [writing] with them. And I guess that is where I feel that tension.

By presenting two ways to teach writing and questioning which is “more valuable,” Josh exposes the tension between the established theory of writing prescribed by the WGE and his own subjugated knowledge of the kind of experience a student needs to “really...grow as a writer.” In this case, the subjugated knowledge of process-based pedagogy has been suppressed by the WGE. Because the WGE defines writing as an individual student constructing a text within a confined amount of time, one way for Josh to teach writing is to replicate the conditions of the WGE until students have been conditioned and are able to employ strategies to write at a passing level. But this overly behaviorist rendering of how his class could look under these conditions does not provide students with what his professional judgment believes students need—time to write,

revise, and grow. Josh's professional judgment focuses on process-based theories of writing found in classroom practices of long term writing development, revision, workshopping, conferencing, and peer-editing. Josh pits the theories of writing within the WGE and what he believes as necessary for real growth against each other as incompatible forms of writing instruction serving different purposes.

In contrast to Josh's first quote in this chapter where he articulated "what they were supposed to teach" as being narrowly defined modes of the WGE, when Josh describes his aims for writing in his class he shifts away from talking about writing as modes to discussing writing in terms of activities, thinking, and process. Maintaining a focus on narrative and expository writing, Josh articulates the work of writing he wants students to engage in as exploratory and process-based:

What I would describe happening in my actual classroom are several different types of writings focused toward the narrative and that expository style. The first part is getting students to write about themselves and [to] explore who they are, what they like and what they don't like. And then writing about ideas, what do you think about this? Give me some reasons, explanations, and examples. That's kind of a starting point, a lot of writing about self and a lot of writing about ideas whether it's a in a journal or formal paper, something about a reading or some sort of assignment, creative or structured. Then of course we get into the big huge process writing of the narratives, the personal essay, the expository.

Approaching his classroom, Josh is able to enact his process-based theory of writing. While not always a full enactment of the approach he outlined in the quote, the

major projects Josh engaged his students in often required two or three drafts. In the discussion of the narrative rubric above, I argued six of the seven categories were from the WGE and demonstrated Josh self-disciplining his judgment to the values of the WGE. However, with Josh's discussion of process-based pedagogy in mind, the inclusion of process and revision reveals his claiming and enacting his professional judgment on the teaching of writing. In addition to the criteria for the WGE, Josh's rubric asks if students have demonstrated a "dedication to improve" the piece through multiple drafts containing "evidence of revision/changes." This category is worth the most points of all the categories and rewards student for their process over the duration of the project. As a means for students to meet the criteria, each student had a conference with Josh and for the particular narrative scored on the classroom rubric, the piece was workshopped twice by peers. For conferences and workshops, Josh's students prepared author's notes explaining where they were in the writing process, the strengths and weakness of the narrative, and specific question they want help answering. Through these activities, Josh creates a space for students to have the time and conversations he believes necessary to improve as writers. Incorporating process-pedagogy into his classroom, Josh is able to draw on a theory of writing pedagogy that resists a one-time sit and write experience the WGE demands.

However, this would not be true of most of his colleagues who use a more regimented approach to preparing students for the WGE, particularly within the few weeks before the exam when repeated timed writings were used by teachers who justified it as preparing students for the conditions of the test. The insight provided in the interview points to the agency of Josh's professional judgment. Above, I illustrated how

the disciplinary knowledge of the WGE influenced what Josh taught in his classroom. His instruction accommodates the WGE and he is still able to draw on his subjugated professional knowledge by including process writing within his classroom rubric and instruction. Josh is able to claim agency and make a judgment about his classroom and his instruction that uses his subjugated knowledge, complicating Hillocks' argument that teachers fully assume the theory of the test as their theory of writing instruction.

While Josh's subjugated knowledge of process pedagogy illustrates a pedagogical tension within his classroom practice, Laurie's critique and subjugated knowledge points to an ideological and cultural clash occurring outside of her classroom as part of a national trend of test-based accountability. On the surface, Laurie describes concerns for her teaching in reference to teaching to the test. Remember, in her first quote of the chapter, Laurie said she was "embarrassed" about how the WGE impacted her teaching of writing. Continuing her conversation on teaching to the test, Laurie explained:

I feel like I am just totally teaching to it, trying to give them hints, trying to give them ways and I don't know if that is right. I don't know if that's what the district envisioned for a student, for an outcome. I don't know if they envisioned it in the way that I'm doing it, which might not be right. I don't know I guess because I still have questions about the whole test thing that I just feel like I teach to the test, which I have never done before, so it's an odd thing.

Discussing her Composition course, Laurie is apprehensive about the work she does to prepare her students for the WGE, questioning how she teaches and if the district intended her teaching to serve as preparation for the WGE. Beneath this surface level

description of her teaching, Laure is concerned “about the whole test thing.” Unveiling this hesitation reveals a larger ideological and cultural concern working beneath the questions surrounding her classroom and teaching. The central tension Laurie has with the WGE is it ignores issues of diversity and difference of students and writing. Teaching in the most diverse high school in the district and working with the population of students who are most affected by the WGE causes Laurie to question the politics and ideology of standardized tests. Laurie acknowledges that students’ writing scores over the years have increased, yet conversations about diversity and cultural understanding have diminished.

Laurie explains that since the implementation of the WGE she feels students writing has improved, and that composition teachers are more aware of the standards because of their time together scoring the exams; however:

I believe very strongly that all this testing stuff is a real backlash against multicultural education. Multicultural education which we don’t do at all any more and I think teachers hated multicultural education because there was so much gray... That was our last big movement, multicultural education. And we went right from that to testing across the country. I think [testing] was a reaction against [multicultural education] being too gray, way too individual, too personal and [the teachers] couldn’t do it. [With the WGE] it’s a black and white; either the kid passes or the kid doesn’t and we’ve taken away all of the stuff that we were starting to learn about race factoring in[to learning]. We’re still afraid to talk about race I remember when [the WGE] began all the teachers [were like] ‘wow this is

the best thing ever' because there is a right and a wrong and you can pass them and not pass them and they think everybody is the same.

Laurie locates the removal of multicultural education from the writing classroom as part of a larger national testing movement. Recall in chapter two, teachers and administrators discussed one of the reasons the WGE came into existence was that BPS wanted to stay ahead of the national movement toward testing by creating their own standardized test.

From Laurie's perspective, the adoption of standardized tests ultimately led to the abandonment of multicultural education as a political and ideological trade-off.

Accompanying the WGE are set standards for writing, which are presented as neutral, along with the rubrics used to measure writing. In her quote, Laurie attributes standardized tests to determining "right and wrong" writing while purporting the belief that all of the students are the same. Her commitments to multi-cultural education resist such a read of students and texts. Rather, she wants students to wrestle with cultural difference and issues of diversity. Removing discussions of difference from writing and treating it as an acontextual act simplify and overlook the complexity of literacy Laurie wants her students to engage in.

Laurie also discussed how racial and cultural values can cause conflict for students taking the WGE. Throughout her interviews and our conversations Laurie spoke how the WGE, specifically the narrative, privileges the values of written cultures over oral cultures. Laurie said, "we need to get out some of the biases that are in there with languages and stuff like that. It caters to white kids; you know, their culture is much more a written culture." In another interview, Laurie addressed how non-white students

have more barriers to navigate within the school system and specifically in passing the graduation demonstration exams.

We expect kids to do things the white way all the time and so [the WGE] is another thing. White culture is a writing culture. I mean Native American and African American culture, theirs are oral cultures... You can sit down with a student and say, 'tell me a story.' And they can tell you a story orally, but for the WGE they are forced to do something in a way that's not necessarily part of their culture. So in a way, it's another barrier they have to jump through and they have so many of them.

Several teachers at Marshall High School, including Laurie, raised the narrative as a concern for their non-white students. Narrative as defined by the WGE consists of a linear structure with a clear beginning, middle, and end. However, as Laurie and her colleagues pointed out, many of their non-white and immigrant students struggled constructing a narrative in this format. For some of their students, narratives were circuitous in nature rather than linear or lacked dialogue and description focusing on the moral or lesson rather than dialogue and setting. Keeping with their cultural norms of story telling, students who might be able to write but have not yet acculturated to the particular narrative style which the WGE mandates are penalized. Laurie's insight and knowledge highlights the ways the WGE ignores cultural difference in writing and works as a barrier against students, particularly minority students and ELL students. Laurie's subjugated knowledge of the WGE disputes the notion of standardized tests as a neutral tool; rather she sees the WGE functioning as a cultural and educational gatekeeper of

students while simultaneously perpetuating a belief that writing is an act void of cultural diversity.

In addition to the adoption of a standardized test, and the narrow definition of writing, scoring sessions serve as a site for cultural tensions. Laurie discussed how during scoring sessions she saw scorers marking papers down when African-American students incorporated black dialect in their writing. “You see some of those readers are so strict. If the student slips into a black dialect that they are going to take that out on them.” Again the issue of culture and writing comes up with ELL students who according to Laurie, “are probably hurt the worst.” “ELL kids who struggle with subject verb agreement or they will leave out certain words and some people are really strict with that, so I do think a cultural thing enters into it.” The issues Laurie raises here are often referred in scoring sessions to “control.” The linear narrative structure and the use of standard American English exhibit the writer is in control of the text and as Laurie points out, these values of language and writing are cultural; however, the WGE does not provide flexibility to consider how culture impacts writing and language. Rather, it presents the values of white American written culture as the norm for all students to learn. Or as Laurie summarized, “I do think a cultural thing enters into it... I just think that when you have all white people coming up with the test and grading the test it’s got to enter in.”

Laurie’s commitment to teaching diversity and multiculturalism was evident in interviews as she professed American and multicultural literature as her preferred courses to teach due to the immediacy of issues around race, gender, and class that could be taken up in the readings. Her commitment to social issues also emerged in her Composition

courses though to a lesser degree than the literature courses. One example was her use of writing prompts on gender issues for students to write about. In addition to the prompt discussed above students were asked to consider gender roles and stereotypes to write about, as well as social rituals such that occur on the playground, at slumber parties or in gangs. Through these prompts, Laurie encouraged her students to write about the world they live in and to consider issues, in this example, of how gender impacts students' lives. Bringing issues of diversity and multiculturalism into her class through writing prompts, Laurie is able to have her students write about these issues she feels are a critical part of her students' education. However, using issues of diversity as content to write about does not alter her larger concerns of the way factors of race and culture manifest themselves within institutional structures like the district.

Laurie's discussion of race and culture in relation to writing and standardized assessment illustrate her subjugated knowledge of the role culture and race play into language use and writing. While the district does put the writing prompt and all the technologies through a bias review, it does not account for the concerns Laurie raises in her interviews. Whereas Josh can draw on process pedagogy in his classroom, the nature of Laurie's critique points to larger political and ideological issues that she finds difficult to address or alter through her classroom practice.

The moments when Josh and Laurie provide insight into their knowledge about student writing and teaching that is not valued by the WGE offer moments out of which critique can be built. Or as Foucault said, "it is through the re-appearance of this knowledge, of these local popular knowledges, these disqualified knowledges, that criticism performs its work" (82). Through the act of criticism Josh and Laurie are able

to resist a deterministic reading of the elimination of agency and reclaim their professional judgment, though in veiled or tempered ways.

Conclusion

Thinking back to chapter one and the “writing problem” espoused by administratively-minded reform advocates in Progressive Era educational journals, the largest problem facing the teaching of writing was subjective teachers who lacked standards and scientific tools to accurately measure writing. The result, as I illustrated, was the creation of writing scales as the first attempt to objectively measure writing, moving teachers away from subjective judgments toward positivistic notions of measurement. Following on the heels of writing scales were the first standard tests, an early version of standardized writing assessment. Using Josh and Laurie as examples to illustrate what happens to teachers’ judgment within a standardized assessment system could be read as the consummation of the administrative progressives’ vision of education. Josh and Laurie no longer rely on “rule of thumb” methods of writing; rather what they teach, how they teach, and their classroom assessments are based on a measurable standard of writing and a scientific process of standardized writing assessment. In fact, Laurie’s use of the holistic rubric within her classroom where she assigns numerical scores based off examples that have been calibrated and equated come closest to the most conservative calls for reform of the Progressive Era. Yet, both Laurie and Josh’s ability to make a judgment about writing and writing assessment is more complex than the manifestation of the goals of the administrative progressives.

Furthermore, George Hillocks' argument that the standardized writing test becomes the teachers' knowledge and theory of teaching writing is not entirely true, as I have argued in this chapter. First, Hillocks overlooks the way the mechanisms of power functions within an educational setting to norm teachers' judgment. Teachers do not simply adopt a theory; rather the institution disciplines their knowledge to the point that teachers self discipline their actions and ideas resulting in the values of the test to be codified and present to students as absolutes. However, despite the work of the institution to remove agency from individuals, Josh and Laurie both demonstrate the ways they are able to articulate other knowledges of writing that contradict and critique the theory of writing prescribed by the WGE. Though not always obvious on the surface, these subjugated knowledges demonstrate the ways teachers (Josh and Laurie specifically) work and teach between incongruent knowledges and theories of writing and assessment.

Josh's E-mail: A Response

As I stated in the introduction, writing about teachers is not easy, particularly when issues of structure and agency are at the center of the discussion. Had I stopped with the first half of this chapter and argued teachers are disciplined by the WGE system, which builds capacity to serve the system, that reading would have ignored the knowledge of writing Josh and Laurie carry with them that disrupts and resists the disciplinary knowledge. On the other hand, had I only focused on the disruptions and celebrated teachers' ability to claim agency while ignoring the work of the WGE system on teachers' knowledge and instruction, that reading, too, would be incomplete.

Following in Ellen Cushman's advice, I have attempted to "describe the tension filled place where individuals' acts wrangle with larger structural constraints" (37). While I have done my best to capture these fissures, I recognize that the above chapter is my construction as a researcher. Therefore, as part of my methodology I conducted member checks in order to receive feedback from the participants who shared their professional knowledge, time, and classrooms with me. I have cycled this chapter back to Josh and Laurie, and rather than end with my conclusion, I've decided to end with an e-mail Josh sent me after reading a draft of this chapter.

Hi Eric,

I greatly enjoyed reading chapter 5. Now I want to read the entire book!

First, I should tell you that I think you accurately portrayed me as a teacher in my classroom. As I was reading I kept asking myself the question "is that really me?" or "is that what I really am doing in my classroom?" and the more I thought about it, the more I realized that you were right on.

What you said on page two is incredibly true of me--"through this conflict of knowledge/theory, teachers neither fully submit to the WGE nor do they fully enact their professional knowledge of writing. Rather they manage their instruction, in-class assessment, classrooms and students between incongruent theories of writing instruction."

More than anything, Eric, reading this chapter has challenged me to think deeper about my classroom and the ways in which I teach writing. What really frightened me was reading about "Laurie's" classroom on page 9 when she tells the student not to "change her mind" in the writing because that means the writer was not in control. I see myself expecting the same of my students! But, as a writer myself, my process in writing is never exactly "controlled" but rather more out-of-control. Through writing, I discover, draft, revise, etc. in order to make sense of what I am trying to communicate. Often that is a messy, out-of-control act that ultimately leads to a polished piece of writing.

I don't really know what sense I'm trying to make out of this rambling. Overall, I thought this chapter was thought-provoking and revealing at that the same time. I guess what I want you to know is that I am worried about how I move forward as a writing teacher-- not because of your study or from reading your chapter. But, during our work at Wilson, during Professional Learning Communities (PLCs), we've discussed "essential" learnings for our students. One of those essential learnings for 9th graders has been focused on narrative and expository writings. I argued at great length with the entire English department and my 9th grade group that good writing should not be focused on the WGE, that writing was something that should be living, fluid, and changing. But our move has been that we (9th grade teachers) should focus on teaching kids how to write narrative and expository pieces in 45 minutes. We should do more "prompted" and timed writing in our classrooms because that is essential. I'm so conflicted, torn, and frustrated. I was also the only one in my 9th grade group who was adamant about not teaching a five

paragraph essay as expository writing. I was "told" that was what should happen because it was the most manageable for 9th graders and since I was the sole voice in opposition I had to go along with the group. There have been interesting and valuable discussions too, one being about narrative writing and how we should be having kids do oral story-telling before they write, and that a narrative is not linear but instead circular with embedded smaller sub-stories within a larger story.

I'm starting work on M.A. this summer. I'm eager to begin my own study of schools and instruction because I've become ever so frustrated with my work and my classroom.

I think your conclusion accurately reflects what I see happening in my department's discussion of writing in PLCs: "Teachers do not simply adopt a theory; rather the institution disciplines their knowledge to the point that teachers self discipline their actions and ideas resulting in the values of the test to be codified and presented to students as absolutes."

I hope all that I wrote made sense! Do keep me informed as you finalize your work. I want to read your entire book when it is complete and available. And stay in touch as you move on to bigger and better things!

Josh

Epilogue

I have this huge dilemma. I love Dewey and I love that whole idea of experience and experience creating opportunities to learn. How do you let them have these experiences? How do they happen in a bureaucratic, factory model of education? You know I'm faced with this all the time. I can't think about it too much or I won't come to work anymore. Writing is the same way.

Sarah, Wilson High School

Spending the past year working through a year's worth of data, revisiting the yellowed pages of English Journal, drafting, starting over, writing, going back to the data, I kept thinking about the "factory model" of education we have inherited and wondered, like Sarah, if I keep thinking about this, will I come back to work? A watershed moment occurred in an early draft of chapter three, when I was rereading Foucault's Discipline and Punishment along with my field notes from a scoring session. I was mapping the ways Foucault described the formation of docile bodies within social institutions and the ways BPS trained their teachers to score student exams. At that point in my drafting, I was offering a rendering of the district removing agency from teachers and building a capacity within them. I was frustrated when I came across the moment in the transcript when Mark disrupted the scoring session by considering the consequential validity of the exam and the placement of a particular student. After all, Mark wasn't doing what Foucault said he should have been. While Mark momentarily disrupted the scoring session by not moving his reading to consensus, the practice of inter-rater reliability kept him bound within the larger theories of reliability and classical text theory. He could only advocate for the student to a certain extent due to the larger structural and theoretical systems of power Mark was working within.

About the same time I was thinking about Mark and the scoring sessions, I was sharing some early writing about Josh with my dissertation chair. Again I tried out a similar Foucaultian reading on his teaching of dialogue and the use of his classroom rubric. In early drafts I ignored Josh's category on writing process; however, my dissertation chair wrote in the margins, "How do you account for the process category?" Returning to the interview transcripts I conducted with Josh, I saw the tension Josh discussed of wanting to teach process writing, despite the WGE not including revision as part of the exam. Reviewing my field notes, I had tried to construct Josh as one of the only teachers in my study who did teach process writing, resisting the WGE and its timed writing for a more robust experience for students. However, neither construction of Josh on its own adequately accounted for my observations and interviews. It seemed the more I tried to write about teachers losing their professional judgment and agency within the WGE system, the more I began to see these fissures, disruptions, and counter narratives within my transcripts and field notes. Approaching these fissures, I could not overly romanticize that teachers were radically altering the WGE system or the teaching and assessing of writing within BPS due to the larger system they work under. However, the discovery of these fissures invited me to reexamine my assumptions and readings of the constraints and resistances going on inside the walls of the schools and in the lived experiences of the teachers in my study.

Throughout the last three chapters I illuminated the necessity of resisting a reading of teachers solely being managed by the mandates and processes of accountability. Instead, attending to the disruptions and fissure within the WGE system—either the unintended consequences of the system or through teachers claiming

agency—allows for alternative knowledges of writing, teaching, and schooling to emerge. In moments when the system fissures, the complexity of the lived experience of teachers is exposed and opens a space to formulate questions about the WGE—or any system of assessment and accountability. Pursuing questions from within the system provides possibilities for teachers and administrators to reflect on their professional work and knowledge in order to imagine the possibilities of reforming the district's accountability system, their assessment practices, and their instruction.

The moments of fissure need be captured, documented, and made known to those who create and work as part of educational systems. In the case of this research project, I was able to maintain a vantage point outside the WGE system as I documented the work of teachers through interviews and observations. Returning to the data, I could trace trends of fissures in order to write about them. However, in addition to the work of outside researchers, there is a need for internal documentation of the discontinuity whereby teachers and administrators articulate and reflect on the competing forms of knowledges and theories they work between on a daily basis. Most importantly, the fissures need to be translated into action. If the documentation of discontinuity remains an academic exercise and is not cycled back into the process of reforming schools, then studying fissures within an educational system falls short of its potential to contribute to reforming writing, reforming writing assessment and ultimately reforming schools. Therefore, I want to conclude by posing the possibility of mapping discontinuities in order to reimagine research, policy, writing, teaching and assessment, and reform as they pertain to my dissertation and public schooling more broadly.

Reforming Research

As I have argued across the chapters, the cliché of “teaching to the test” is far more complex than is often attended to in professional literature. While scholars have argued that tests work to manage teachers, they often do not address the discontinuity within accountability systems. This is particularly true due to the methodologies scholars who study standardized tests on a state level such as Hillocks and McNeil have employed. While Hillocks’ study did contain interviews with teachers, they were one-time interviews; since his study lacked observations of teachers in their classrooms and in other professional capacities, he was unable to observe their daily lived experience and the ways teachers both conform to and resist the pressures of a standardized writing test. The critiques of Hillocks and McNeil are often politically popular in educational circles, but there is a danger that these conclusions overlook the effort of teachers working against systematic reform and disciplinary knowledge contained in test-based accountability.

Further research on standardized writing tests ought to offer a voice to teachers’ lived experience and their knowledge of writing, students, and the contexts and communities in which they teach. Rather than view public schools through the theories of academia, research on teachers might look to build new theories that are based on the perspectives and experiences of teachers (Cushman “Rhetorician” 23). That is, teachers and researchers ought to work together in what Chris Gallagher calls “engaged professionalism.” “The very purpose of [engaged] professionalism, [...] would be to redeploy expertise as a tool of *expanding*, rather than constricting, access to public conversations and institutional practices” (Gallagher “We” 89). The documentation of the

fissures and discontinuity within teachers' professional lives and within school systems serves as an exigency to expand the dialogue and illuminate the institutional practices within schools. Through the collaborative work of researchers and teachers engaged in knowledge-making, moments of discontinuity can be used to question and critique current institutional practices. By making these fissures known to other teachers, administrators, policy makers, parents, and the general public, the discourse and practice of teaching, assessing and learning to write can be revised from a simple accountability mandate into a robust discussion on the impact of such policies on the experiences of students and teachers who work daily within schools.

Reforming Writing

Rather than defining writing as a technical skill for employment or functional literacy, writing in schools needs to be expanded to a meaning-making activity. As James Berlin wrote, "In teaching writing, we are not simply offering training in a useful technical skill that is meant as a simple complement to the more important studies of other areas. We are teaching a way of experiencing the world, a way of ordering and making sense of it" (246). Exploring the fissure across the last three chapters illuminates the ways teachers in interviews, scoring sessions, and in their classrooms articulate writing as more than a simple set of skills, but as a creative act embedded in socio-cultural contexts. Rather than squelching conversations on the use of narrative and expository techniques, the role of revision and process, or the impact a student's race and culture has on composing, these fissures open up the possibility to define writing as something more than neutral skills students must acquire before graduation. Instead,

these fissures point to the ways writing and writing pedagogy has shifted over the twentieth century away from current-traditional practices focused on modes of discourse toward process-based and culturally located acts of meaning making.

Though teachers addressed contemporary theories of writing and writing pedagogy in various places of this study, the tools used to measure writing on the WGE are similar to the writing scales and standard tests used a century earlier. Embedded within the tools of contemporary writing assessment are the positivist ideas of writing and measurement which lead to students producing short essays based on the traditional modes of writing, essays easily scored according to form and conventions. A clear lesson from the early pages of English Journal as well as Butler Public School's WGE policy is composition teachers need to assist in framing "writing problems." When left in the hands of measurement specialists and policy makers, the work of writing is reduced to basic skills to be easily measured or scored. However, teachers and researchers can draw on teachers' subjugated knowledges of writing to reframe "writing problems" and the act of writing as a socially dynamic process of meaning making. Accompanying this shift to align writing in schools with contemporary theories of writing and writing pedagogy, new assessment tools and processes will need to be designed to reflect the theories underpinning the reframing of the "writing problem."

Reforming Teaching and Assessment

In BPS's current use of the WGE, the purpose of the assessment tool is two-fold. Its primary function is to serve as an accountability mechanism followed by being used as a placement tool. Beyond providing the district's teachers with a common language for

teaching writing, the relationship between assessment practices and instruction is tenuous, especially in courses where students have not yet passed the WGE. Specifically, teachers were interested in more analytic feedback about why students did not pass the WGE in order to inform their instruction. On a larger scale, some teachers voiced the possibility of conducting alternative forms of assessment such as portfolios, as a means to measure a student's proficiency in writing over their entire high school career. Through these critiques and questionings, teachers articulate the importance of a reciprocal relationship between assessment and instruction to support the work of teachers and students in classrooms. The question then becomes how to strengthen the relationship between teaching and assessing writing so that teachers are better able to help students learn.

The fissures in chapter three's scoring sessions, frame questions of writing assessment around student learning and provide an opportunity to shift the purpose of assessing students' writing. Rather than starting with concerns of technical design and control when approaching assessment, working backward from learning and instruction—from the classroom rather than the district office—allows for teachers to collectively build goals and outcomes for their students.³³ Grounding assessment in conversations of students learning, instructional practice, theories of writing and pedagogy, and the curriculum students experience, assessment tools can be designed to produce information and data about students, teaching, and schools that can be cycled back to further improve students' learning and teachers' instruction. This is not to say technical concerns should be ignored in doing writing assessment. Rather, the technical components of assessment ought to compliment and support the work of teachers

teaching writing in classrooms, not constrict the act of writing or truncate the feedback teachers and administrators receive due to narrow conceptions of reliability and validity.

Reforming Policy

In Composition in the Twenty-first Century, Edward White posited that the problem for writing assessment in the twenty-first century “will be to bring together the goals, instruments, and methods the most informed writing teachers use in their classrooms with those used by the assessment community to shape public policy” (“Will” 101). Contemporary political discourse around schools and educational policy focuses on standards and outcomes, drawing primarily on the assessment community to provide metric systems to fulfill policy mandates. However, policy makers must begin to confront the consequence (both intended and unintended) of this discourse the ways these policies are implemented in schools. Addressing the discontinuity between the official story and the counter-narratives (chapter 2) or the incongruent theories of writing instruction (chapter 4) provides an opportunity to bring teachers’ subjugated knowledges to bear on policy conversations. More specifically, by focusing dialogue on how policies impact classrooms, the site of teaching and learning, there is potential to shift the discourse of policy making from standards and management to learning and teaching. That is, to flip the current top-down practice of policy making to consider and move toward an inside-out version where the experiences of teachers and students are valued and deliberated on as part of the process of making and revising policy decisions.

Enacting Reform

Schools are social institutions, a product of history and rhetoric, and therefore can be revised and reformed. As Tyack and Cuban in their study of school reform illustrate, reform occurs through groups of people posing problems and presenting solutions which bring about the types of changes they seek. Over time, successful acts of reform become part of the structures and practices of schools—what “real schools” do (88). The residual effects of past reform efforts have been embedded in schools today, and while this demonstrates the power of successful efforts of school reform, it also illustrates that institutions can be altered by people. As James E. Porter, Patricia Sullivan, Stuart Blythe, Jeffery T. Grabill, and Libby Miles wrote in CCC, “Though institutions are certainly powerful, they are not monoliths; they are rhetorically constructed human designs (whose power is reinforced by buildings, laws, traditions, and knowledge-making practices) and so are changeable” (611). Porter et al. advocate for a methodology of institutional critique in order to “expose and interrogate possibilities for institutional change” (631).

Throughout this dissertation I have drawn on both historical and contemporary sites in order to explore the ways writing assessment has been used as a tool for institutional reform. In the archival study, the ideology of efficiency shaped the debate on the “writing problem,” and in the study of Butler Public Schools, a response to accountability trends led to the creation of the WGE. In both cases, standardized tools were created to measure the output and ability of students and teachers within schools. The two studies cover a time span just shy of a century in American education and reveal how standardized tools used to measure writing have become part of the “grammar” and structures of schools today (Tyack and Cuban 88). Nevertheless, both historical and

contemporary studies also illuminate teachers' resistance to the simplistic and narrow definitions of writing and the "science" used to objectively measure student writing.

The work that remains is making the subjugated knowledge of teachers a part of larger discussions on school reform and writing assessment. As I have outlined above, the moments of fissure open spaces to consider expanding definitions and practices of writing, the relationship of instruction and assessment, and the overall impact of policy making on the lived experiences of teachers and students. In order to change schools and the way writing has been practiced within schools, the recovery of subjugated knowledges must move beyond the walls of classrooms and schools. It would be naïve to argue that reforming schools and writing will be an easy task. In fact, it will be most difficult. However, the questions and concerns found in the moments of fissure provide a foundation to establish conversations aimed at initiating reform of writing and writing assessment. Working with various stakeholders of public schools—students, teachers, administrators, parents, and policy makers—is an ongoing act of dialogue and an enactment of democracy. At the very least, engaging stakeholders in conversations centered on the subjugated knowledges of teachers and administrators provides an opportunity to focus public dialogue on the goals of teaching writing and the purposes of assessing writing within public schools—a relevant conversation to sustain.

Appendix A

WGE Expository Writing Prompt

Writing Graduation Exam

Day One Expository Writing Assignment

Directions:

1. Using a blue or black ink pen, write a "1" on the cover of your answer book.
2. Write any planning notes or outlines on the inside cover of your answer book.
3. If you need to make a correction, draw a single line through the mistake. Write your corrections after or above the word(s) you cross out.
4. Complete the writing assignment by the end of the time provided.
5. Read today's writing assignment printed below and begin.

First Writing Assignment

Starting in seventh grade, students in the Butler Public Schools enroll in both required and elective courses.

An adult asks, "Which class would you recommend students take before graduating from high school?" Tell why you think a current course or a new course you create would be a valuable experience.

After selecting the course:

- write an essay sharing your opinion.
- provide adequate reasons for your position.
- support and explain your reasons with sufficient evidence such as examples from literature, history, public life, and/or your personal life.

Appendix B

WGE Expository Rubric

Writing Graduation Exam Holistic Scoring Rubric: Expository

6--Exceptional Response

Overall, the response is perceptive and sophisticated; it provides an exceptional explanation, orienting the reader to an opinion and contextualizing its importance. The writer incorporates compelling logic, reasoning, and evidence, controlling the structure and a wide range of stylistic elements appropriate for exposition (point of view, word choice, syntax, tone, figurative language, or figures of speech). Though the writing is not without flaws, the writer effectively controls usage, grammar, mechanics, and spelling, as appropriate to the content.

5--Commendable Response

Overall, the response is thoughtful and effective; it provides a commendable explanation, orienting the reader to an opinion and contextualizing its importance. The writer incorporates sound logic, reasoning, and evidence, controlling the structure and a range of stylistic elements appropriate for exposition (point of view, word choice, syntax, tone, figurative language, or figures of speech). The writer controls usage, grammar, mechanics, and spelling with occasional lapses.

4--Adequate Response

Overall, the response is plausible and functional; it provides an adequate explanation, orienting the reader to an opinion and contextualizing its importance. The writer incorporates sufficient logic, reasoning, and evidence, controlling the structure and a restricted range of stylistic elements appropriate for exposition (point of view, word choice, syntax, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may occur, but do not confuse meaning.

3--Inadequate Response

Overall, the response is vague, underdeveloped, or simplistic; it provides an inadequate explanation, offering little orientation to an opinion and its importance. The writer may control the structure, but uses unclear reasoning, flawed logic, or inadequate evidence, and applies few stylistic elements appropriate for exposition (point of view, word choice, syntax, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may cause distraction and some confusion.

2--Limited Response

Overall, the response is undeveloped or confusing; it provides a limited explanation for an opinion or its importance. The writer presents an opinion with little or no logic, reasoning, or evidence, and may neglect structural and stylistic elements appropriate for exposition (word choice, syntax, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may confuse meaning.

1--Minimal Response

Overall, the response is unfocused, incoherent, or too brief to assess expository skills; it may only provide a minimal reference to an opinion. The writing lacks structure, and the stylistic elements are limited to basic word choices and fragmented or simple sentence patterns. Errors in usage, grammar, mechanics, and spelling may lead to incoherence.

0--Unscorable (Consult TL) The response is off topic, not written in English, illegible, or blank

Appendix C

WGE Narrative Writing Prompt

Writing Graduation Exam

Day Two Narrative Writing Assignment

Directions:

1. Using a blue or black ink pen, write a “2” on the cover of your answer book.
2. Write any planning notes or outlines on the inside cover of your answer book.
3. If you need to make a correction, draw a single line through the mistake. Write your corrections after or above the word(s) you cross out.
4. Complete the writing assignment by the end of the time provided.
5. Read today’s writing assignment printed below and begin.

Second Writing Assignment

A publication called TeensWrite: A Collection of the Stories of Adolescents is interested in publishing narratives written by teens based on their own experiences.

The subject for the upcoming issue is: **a memorable event or memorable events.**

Choose a memorable event or memorable events. Write a narrative for possible submission to TeensWrite in which you recreate this experience for the reader

Appendix D

WGE Narrative Rubric

Writing Graduation Exam Holistic Scoring Rubric: Narrative

6-- Exceptional Response

Overall, the response is captivating, sophisticated, and engaging; it provides an exceptional account of an experience, orienting the reader and contextualizing the event's significance. The writer incorporates narrative techniques (action, dialogue, description) and controls a wide range of stylistic elements appropriate for narration (word choice, syntax, voice, tone, figurative language, or figures of speech). Though the writing is not without flaws, the writer effectively controls usage, grammar, mechanics, and spelling, as appropriate to the content.

5--Commendable Response

Overall, the response is genuine, effective, and fluent; it provides a commendable account of an experience, orienting the reader and contextualizing the event's significance. The writer incorporates narrative techniques (action, dialogue, description) and controls a range of stylistic elements appropriate for narration (word choice, syntax, voice, tone, figurative language, or figures of speech). The writer controls usage, grammar, mechanics, and spelling with occasional lapses.

4--Adequate Response

Overall, the response is plausible, sufficient, and functional; it provides an adequate account of an experience, orienting the reader and contextualizing the event's significance. The writer incorporates narrative techniques (action, dialogue, description) and controls a restricted range of stylistic elements appropriate for narration (word choice, syntax, voice, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may occur, but do not confuse meaning.

3--Inadequate Response

Overall, the response is underdeveloped, simplistic, or insufficient; it provides an inadequate account of an experience, offering little orientation for the reader and/or meager contextualization of the event's significance. The writer lacks control of narrative techniques (action, dialogue, description) and uses limited stylistic elements appropriate for narration (word choice, syntax, voice, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may cause distraction and some confusion.

2--Limited Response

Overall, the response is inconsistent, undeveloped, or confusing; it provides a limited account of an experience, neglecting to orient the reader or contextualize the event's significance. The writer lacks narrative techniques (action, dialogue, description) and neglects stylistic elements appropriate for narration (word choice, syntax, voice, tone, figurative language, or figures of speech). Errors in usage, grammar, mechanics, and spelling may confuse meaning.

1--Minimal Response

Overall, the response is unfocused, incoherent, or too brief to assess narrative skills; it provides a minimal reference to an experience, neither orienting the reader nor contextualizing the event's significance. The stylistic elements are limited to basic word choices and fragmented or simple sentence patterns. Errors in usage, grammar, mechanics, and spelling may lead to incoherence.

0--Unscorable (Consult TL) The response is off topic, not written in English, illegible, or blank.

Appendix E

Josh's Classroom Rubric

NARRATIVE RUBRIC	Points Available	Points Earned
MECHANICS Is standard edited English used? <input type="checkbox"/> grammar <input type="checkbox"/> punctuation <input type="checkbox"/> sentence structure <input type="checkbox"/> spelling <input type="checkbox"/> easy to read and understand	15	
OPENING Is the beginning attention getting? <input type="checkbox"/> provides introduction of the characters and setting <input type="checkbox"/> sets up the story in a creative & interesting way	10	
CLOSING Does the ending give a sense of completeness? <input type="checkbox"/> the significance/importance of the story is clear <input type="checkbox"/> there is some resolution that ends the story	10	
DIALOGUE Is there dialogue included? <input type="checkbox"/> the dialogue carries and adds to the story <input type="checkbox"/> there is enough dialogue included <input type="checkbox"/> it is clear who is speaking <input type="checkbox"/> it is punctuated correctly	15	
DESCRIPTION Are there vivid descriptions of the characters, settings, and the event? <input type="checkbox"/> creative words are used <input type="checkbox"/> the five senses are used: sight, touch, taste, sound, smell <input type="checkbox"/> a combination of simile, metaphor, compare and contrast is used	15	
ACTION Is the action re-created? <input type="checkbox"/> showing AND telling are both used <input type="checkbox"/> detail of the events is included <input type="checkbox"/> exaggeration/overkill is used to create a picture	15	
WRITING PROCES Has there been dedication to improvements? <input type="checkbox"/> there are several drafts of the story <input type="checkbox"/> each draft there is evidence of revision/changes <input type="checkbox"/> the author has been dedicated to additions/improvements <input type="checkbox"/> class time was used to be focused and working	20	
TOTAL	100	

Works Cited

- Anderson, D.A. "The Efficiency Expert in Education." Educational Administration and Supervision 2.8 (1916): 477-482.
- Ballou, Frank Washington. "Scales for the Measurement of English Compositions." Harvard-Newton Bulletins II (1914): 3-93.
- Barden, Carrie. "More Experiments with Scales." English Journal 13 (1924): 648-653.
- Berlin, James A. "Contemporary Composition: The Major Pedagogical Theories." Cross-Talk in Comp Theory: A Reader. Ed. Victor Villanueva Jr. Urbana: NCTE, 1997. 233-250.
- Black, Laurel, Donald A. Daiker, Jeffrey Sommers, Gail Stygall, Eds. New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large-Scale Scoring. Portsmouth: Boynton/Cook, 1994.
- Bloom, Lynn Z., Donald A. Daiker, Edward M. White. Eds. Composition in the Twenty-First Century: Crisis and Change. Carbondale: SIUP, 1996.
- Breland, H. and Gaynot, J. "Comparison of Direct and Indirect Assessments of Writing Skills." Journal of Educational Measurement, 16.2 (1979) 119-127.
- Briggs, Thomas. "English Composition Scales In Use." Teachers College Record 23 (November 1922): 423-452.
- Broad, Bob. "'Portfolio Scoring': A Contradiction in Terms" New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large-Scale Scoring. Ed. Laurel, Donald A. Daiker, Jeffrey Sommers, Gail Stygall. Portsmouth: Boynton/Cook, 1994. 263-276.
- . What We Really Value: Beyond Rubrics in Teaching and Assessing Writing.

Logan: Utah State UP, 2003.

Brossell, Gordon, "Writing Assessment in Florida: A Reminiscence" Assessment of Writing: Politics, Policies, Practices. Ed. Edward M. White, William D. Lutz, and Sandra Kamusikiri. New York: The Modern Language Association of America, 1996. 25-32.

Brown, Marian D., and Haggerty M.E. "Measurement of the Improvement in English Composition." English Journal 6 (1917): 515-530.

Callahan, Raymond. Education and the Cult of Efficiency. Chicago: U of Chicago P, 1962.

Certain, C.C. "Are Your Pupils Up to Standard in Composition." English Journal 12 (1923): 365-377.

---. "By What Standards are High-School Pupils Promoted in English Composition." English Journal 10 (1921): 305-315.

---. "Why Not Include Standard Tests in Your Teaching Program this Term." English Journal. 12 (1923): 463-480.

Charmaz, Kathy. "Grounded Theory: Objectivist and Constructivist Methods." Handbook of Qualitative Research. Ed. N.K. Denzin and Y.S. Lincoln. Thousand Oaks: Sage, 2000. 509-535.

Clapp, John M. "The Better Business Letters Movement." English Journal 7 (1918): 237-44.

Cody, Sherwin. "The Ideal Course in English for Vocational Students (Continued)." English Journal 3 (1914): 371-380.

Cohen, Ronald D. and Raymond A. Mohl. The Paradox of Progressive Education: The

- Gary Plan and Urban Schooling. Port Washington, NY: Kennikat, 1979.
- Cooper, C.R. "Holistic evaluation in writing." Evaluation Writing: Measuring, Describing, Judging. Eds. C.R. Cooper and L. Odell. Urbana, IL: National Council of Teachers of English. 1977. 3-32.
- Cooper, C. R. and Odell, L. Evaluating writing: Measuring, Describing, Judging. Urbana, IL: National Council of Teachers of English, 1977.
- Courtis, S.A. "The Value of Measurements: The Uses of the Hillegas Scale." English Journal. 8 (1919): 208-217.
- Cremin, Lawrence. The Transformation of the School: Progressivism in American Education 1876-1957. New York: Vintage, 1961.
- Cronbach, Lee J. "Five Perspectives on Validity Argument." Test Validity. Ed H. Wainer. Hillside: Lawrence Erlbaum, 1988. 3-17.
- Cushman, Ellen. "The Rhetorician as an Agent of Social Change." CCC 47.1 (Feb. 1996): 7-28.
- . The Struggle and the Tools: Oral and Literate Strategies in and Inner City Community. Albany: SUNY P, 1998.
- David, Jane L. and Larry Cuban. Cutting Through the Hype: A Taxpayer's Guide to School Reforms. Mt. Morris: Education Week Press, 2006.
- Dolch, Edward William. "More Accurate Use of Composition Scales." English Journal 11 (1922): 536.
- Elbow, Peter. "Will the Virtues of Portfolios Blind Us to Their Potential Dangers?" New

- Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large-Scale Scoring. Ed. Laurel, Donald A. Daiker, Jeffrey Sommers, Gail Stygall. Portsmouth: Boynton/Cook, 1994. 40-55.
- . "Writing Assessment in the 21st Century: A Utopian View." Composition in the Twenty-First Century: Crisis and Change. Eds. Lynn Z. Bloom, Donald A. Daiker, Edward M. White. Carbondale: SIUP, 1996. 83-100.
- . "Writing Assessment do It Better, Do It Less." Assessment of Writing: Politics, Policies, Practices. ed. Edward M. White, William D. Lutz, and Sandra Kamusikiri. New York: The Modern Language Association of America, 1996. 120-134.
- Faust, Wirt G. "An Effort to Standardize Descriptive Theme-writing for the Senior Year of the High School." English Journal 5 (1916): 257-271.
- Foucault, Michel. Discipline and Punish: The Birth of the Prison. Trans. Alan Sheridan. New York: Vintage, 1995.
- . Power/Knowledge: Selected Interviews and Other Writings, 1972-1977. Ed. C. Gordon. New York: Pantheon, 1980.
- Gallagher, Chris W. Radical Departures: Composition and Progressive Pedagogy. Urbana: National Council of Teachers of English, 2002.
- . Reclaiming Assessment: A Better Alternative to the Accountability Agenda. Portsmouth: Heinemann, 2007.
- . "We Compositionists: Toward Engaged Professionalism." JAC 25.1 (2005): 75-99.
- Gere, Anne, Leila Christenbury, and Kelly Sassi. Writing on Demand: Best Practices and Strategies for Success. Portsmouth: Heinemann, 2005.

- . A Student's Guide to Writing on Demand: Strategies for High Scoring Essays.
Portsmouth: Heinemann, 2006.
- Gibbony, Richard A. "Intelligence by Design: Thorndike versus Dewey." Phi Delta Kappan (Oct 2006): 170-172.
- Gunther, Charles. "My Experience with the Hillegas Scale." English Journal 8 (1919): 535-542.
- Haber, Samuel. Efficiency and Uplift: scientific management in the Progressive Era 1890-1920. Chicago: U of Chicago P 1964.
- Hatch, Roger Conant. "A Standard of Measurement in English Composition." English Journal 9 (1920): 338-344.
- Hillegas, Milo B. A Scale for the Measurement of Quality in English Composition by Young People. Teachers College: New York City, 1913. Reprinted from Teacher College Record 33.4 (1912): 1-55.
- Hillocks, George, Jr. The Testing Trap: How State Writing Assessments Control Learning. New York: Teachers College Press, 2002.
- Hopkins, Edwin M. "Can Good Composition Teaching Be Done Under Present Conditions?" English Journal 1 (1912): 1-8.
- . The Labor and Cost of the Teaching of English in Colleges and Secondary Schools with Especial Reference to English Composition. 16th ed. Urbana, IL: NCTE, 1923
- Hout, Brian. (Re)articulating Writing Assessment for Teaching and Learning. Logan: Utah State UP, 2002.
- . "Reliability, Validity, and Holistic Scoring: What We Know and What We Need to

- Know.” *College Composition and Communication* 41.2 (May 1990): 201-213.
- Hudelson, Earl. “The Development and Comparative Values of Composition Scales”
English Journal 12 (1923): 163-168.
- . “Some Achievements in the Establishment of a Standard for the Measurement of
English Composition in the Bloomington, Indiana, Schools.” *English Journal* 5
(1916): 590-597.
- Johnston, Joseph Henry. “A Brief Tabular History of the Movement Toward
Standardization by Means of Scales and Tests of Educational Achievement in the
Elementary School Subjects.” *Educational Administration and Supervision*. 2.8
(Oct. 1916): 483-492.
- Kayfetz, Isidore. “A Critical Study of the Hillegas Composition Scale.” *Pedagogical
Seminary* 21 (Dec. 1914): 551-570.
- . “A Critical Study of the Harvard-Newton Composition Scale.” *Pedagogical Seminary*
23 (Sep. 1916): 335-352.
- Kliebard, Herbert. *Schooled to Work: Vocationalism and the American Curriculum,
1876-1946*. New York: Teachers College Press, 1999.
- Kohn, A. *The Case Against Standardized Testing*. Portsmouth: Heinemann, 2000.
- Lagemann, Ellen Condliffe. “The Plural Worlds of Educational Research.” *History of
Education Quarterly* 29.2 (Summer 1989): 185-214.
- Leonard, S.A. “Building a Scale of Purely Composition Quality.” *English Journal* 14
(1925): 760-775.
- . “On Setting Up Composition Targets.” *English Journal* 8 (1919): 401-411.
- Love, Jon F. “The use of Composition Scales for the Measurement of Quality in English

Composition (1903-1963).” MA. Kansas State U, 1962.

Lynne, Patricia. Coming to Terms: Theorizing Writing Assessment in Composition Studies. Logan: Utah State UP, 2004.

McNeil, L. Contradictions of Reform: The Educational Costs of Standardized Testing. New York: Routledge, 2000.

Messick, Samuel. “Meaning and Values in Test Validation: the Science and Ethics of Assessment.” Educational Researcher 18.2 (Mar. 1989): 5-11.

Moss, Pamela. “Can There Be Validity without Reliability?” Educational Researcher 23.2 (Mar. 1994): 5-12.

---. “Validity in High Stakes Writing Assessment: Problems and Possibilities.” Assessing Writing 1.1 (1994) 109-128.

Nelson, Daniel. Managers and Workers: Origins of the New Factory System in the United States 1880-1920. Madison: U of Wisconsin P, 1975.

Noyes, Ernest C. “Progress in Standardizing the Measurement of Composition” English Journal 1 (1912): 532-536.

Ohanian, Susan. One Size Fits Few: The Folly of Educational Standards. Portsmouth: Heinemann, 1999.

Parker, Flora E. “The Value of Measurements: the Measurement of Composition in the English Classes.” English Journal. 8 (1919): 203-208.

Pike-Baky, Meredith and Gerald Fleming. Prompted to Write: Building On-Demand Writing Skills. San Francisco: Jossey-Bass, 2005.

Porter, James E., Patricia Sullivan, Stuart Blythe, Jeffery T. Grabill, and Libby Miles.

“Institutional Critique: A Rhetorical Methodology for Change.” CCC 51.4 (June 2000): 610-641.

Resnick, L. B., and Resnick, D. “Assessing the Thinking Curriculum: New Tools for Educational Reform.” Cognitive Approaches to Assessment. Eds. B. Gifford and M.C. O’Connor. Boston: Kluwer-Nijhoff, 1992.

Rice, J.M. Scientific Management in Education: Its Men, Ideas, and Institutions. New York: Arno Press, 1969.

Scharton, Maurice. “The Politics of Validity.” Assessment of Writing: Politics, Policies, Practices. Ed. Edward M. White, William D. Lutz, and Sandra Kamusikiri. New York: The Modern Language Association of America, 1996. 53-75.

Scott, Fred Newton. “Our Problems.” English Journal 2 (1913): 1-10.

Slomp, Davis H. and Jim Fuite. “Following Phaedrus: Alternative Choices in Surmounting the Reliability/Validity Dilemma.” Assessing Writing 9 (2005): 190-207.

Smith, M.L., “Put to the test: The effects of external testing on teachers.” Educational Researcher 20 (1991): 8-11.

Spring, Joel. The American School: 1642-1990. 2nd ed. New York: Longman, 1990.

Taylor, Frederick. The Principles of Scientific Management. 1911. New York: Harper & Brothers; Mineola: Dover, 1998.

---. Scientific Management. Westport: Greenwood Press, 1972.

---. Shop Management. 1903. in Scientific Management Westport: Greenwood Press, 1972.

---. Testimony Before the Special House Committee. Scientific Management Westport:

Greenwood Press, 1972.

Thorndike, Edward. "Notes on the Significance and Use of the Hillegas Scale for Measuring the Quality of English Composition" English Journal 2 (1913): 551-561.

Thorndike, Edward L. "Measurement in Education." Teachers College Record 22.5 (1921): 371-379.

Tomlinson, Stephen "Edward Thorndike and John Dewey on the Science of Education." Oxford Review of Education 23.3 (1997): 365-83.

Turley, Eric D., and Chris W. Gallagher. "On the *Uses* of Rubrics." English Journal 97 (2008): 87-92.

Tyack, David. The One Best System: A History of American Urban Education. Cambridge: Harvard UP, 1974.

Tyack, David and Larry Cuban. Tinkering Toward Utopia: A Century of Public School Reform. Cambridge: Harvard UP, 1995.

Villanueva, Victor Jr., ed. Cross-Talk in Comp Theory: A Reader. Urbana: NCTE, 1997.

Weaver, Constance. Teaching Grammar in Context. Portsmouth: Boynton/Cook, 1996.

White, Edward. "The Opening of the Modern Era of Writing Assessment: A Narrative." College English 63.3 (2001): 306-20.

---. "Portfolios as an Assessment Concept." New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large-Scale Scoring. Ed. Laurel, Donald A. Daiker, Jeffrey Sommers, Gail Stygall. Portsmouth: Boynton/Cook, 1994.25-39.

---. "Power and Agenda Setting in Writing Assessment" in Assessment of Writing:

- Politics, Policies, and Practices. Eds. Edward M. White, William D. Lutz, and Sandra Kamusikiri. New York: The Modern Language Association of America. 1996.
- . Teaching and Assessing Writing: Recent Advances in Understanding, Evaluating, and Improving Student Performance. 2nd ed. San Francisco: Jossey-Bass, 1994.
- . "Writing Assessment Beyond the Classroom: Will Writing Teachers Play a Role?" Composition in the Twenty-First Century: Crisis and Change. Eds. Lynn Z. Bloom, Donald A. Daiker, Edward M. White. Carbondale: SIUP, 1996. 101-111.
- ., William D. Lutz and Sandra Kamusikiri, eds. Assessment of Writing: Politics, Policies, Practices. New York: Modern Language Association, 1996.
- . and Thomas, L. "Racial Minorities and Writing Skills Assessment in the California State University and Colleges." College English 43.3 (1981): 276-283.
- Wiggins, Grant. "The Constant Danger of Sacrificing Validity to Reliability: Making Writing Assessment Serve Writers." Assessing Writing 1.1 (1994): 129-139.
- Williamson, Michael. "The Worship of Efficiency: Untangling Theoretical and Practical Considerations in Writing Assessment." Assessing Writing 1.2 (1994): 147-173.
- Willing, Mathew H. "The Measurement of Written Composition in Grades IV to VII." English Journal 7 (1918): 193-202.
- Wilson, Maja. Rethinking Rubrics in Writing Assessment. Portsmouth: Heinemann, 2006.
- Yancey, Kathleen Blake. "Looking Back as We Look Forward: Historicizing Writing Assessment." College Composition and Communication 50 (1999): 483-503.

END NOTES

¹ Publications such as Anne Ruggles Gere, Leila Christenbury and Kelli Sassi's Writing on Demand: Best Practices and Strategies for Success and A Student's Guide to Writing on Demand: Strategies for High Scoring Essays, as well as Meredith Pike-Baky and Gerald Fleming's Prompted to Write: Building On-Demand Writing Skills have become a mainstay in book catalogues and are marketed as professional resources for teachers.

² Pseudonyms have been given to all administrators and teachers from the qualitative study.

³ It would be an over simplification of history to say the administrative progressives won the day over pedagogical progressives—though some have, see Lagemann, Ellen “The Plural Worlds of Educational Research” *History of Education Quarterly* 29:2 1989; Cohen and Mohl, *The Paradox of Progressive Education: The Gary Plan and Urban Schooling*;

⁴ Percentages of women teachers in the United States grew from 59 percent in 1870 to 70 percent in 1900 to 86 percent in 1920. Female teachers were often brought into the public schools because they were seen as subordinate to the male managers who ran the system.

⁵ See Raymond Callahan's history of the growth in the “new profession” of school administrators and superintendents in his chapter, “A New Profession Takes Form” in Education and the Cult of Efficiency.

⁶ See Chris Gallagher's *Radical Departures* 22-24 for a more in depth read of the Hopkin's report

⁷ Noyes worked at the Fifth Avenue High School in Pittsburgh, PA. I am not certain if he was a teacher or an administrator.

⁸ See Noyes “Progress,” Hudelson “Some” and “Development,” Brown and Haggerty “Measurement,” Willing “Measurement,” Parker “Value,” Courtis “Uses,” Gunther “My Experience,” Hatch “Standard,” Certain “By What” and “Are Your,” Dolch “More,” Barden “More,” Leonard “On Setting” and “Building,” Faust “An Effort,”

⁹ Taylor's most popular experiments with scientific management occurred at the Bethlehem Steel Company. Under the old labor practices, the loading gang at the Bethlehem Steel Company loaded 12½ long tons of pig-iron per man per day. The work consisted of picking up a pig of iron weighing approximately 92 pounds, carrying it up an inclined plank onto a railroad car, and placing the pig of iron down. After studying the materials and labor, Taylor concluded that the men should be able to handle 47 tons per man per day rather than current 12½ (Taylor 18-19). Rather than viewing the loading of

pig iron as one job, Taylor broke it down into five measurable units: picking up the pig from the ground or pile; walking with it on a level; walking with it up an incline to car; throwing or laying it on a pile; and returning back empty to get a load (*Shop* 148). Through the atomization of a job each movement was studied until a standard time was established for each movement and the job as a whole.

The second step was the selection and training of the men. At the Bethlehem Steel Company there were 75 pig-iron handlers; Taylor estimated one in eight was physically capable of handling 47 tons per day. The men who were selected and worked under the new system were referred to as “first-class men” rather than workers or laborers because “first-class men are not only willing but glad to work at their maximum speed providing they are paid from 30 to 100 percent more than the average of their trade” (*Shop* 25). Taylor’s creation of “first-class men” created a new industrial class system whereby laborers could significantly improve their wages by working at a scientifically determined rate.

But not all men were capable of laboring under the time requirements Taylor established which forced many out of a job. Always chasing efficiency, Taylor addressed this concern:

With most readers great sympathy will be aroused because seven out of eight of these pig-iron handlers were thrown out of a job. The sympathy is entirely wasted, because almost all of them were immediately given other jobs with the Bethlehem Steel Company. And indeed it should be understood that the removal of these men from pig-iron handling, for which they were unfit, was really a kindness to themselves, because it was the first step toward finding them work for which they were peculiarly fitted, and at which, after receiving proper training, they could permanently and legitimately earn higher wages. (Taylor 31)

Taylor’s language invokes a social Darwinian idea that men were fitted for a certain type of work and therefore a certain amount of pay. In an example with bricklayers, laborers who could not lay bricks at the correct speed were demoted to stacking bricks or mixing mortar both jobs even if performed at the level of a first-class man did not make the near the wage of a bricklayer. The result was less workers producing more labor or in Taylor’s eyes a more efficient system.

¹⁰ Four of the foreman—the gang boss, speed boss, inspector, and repair boss—worked on the floor with the men teaching, training, timing, and inspecting the product and machinery. The Planning department located away from the physical labor consisted of four foreman—the order of work and route clerk, instructional clerk, time and cost clerk, and shop disciplinarian—who oversaw the planning and running of the factory.

¹¹ See D. A. Anderson’s “The Efficiency Expert in Education” for a celebration of the efficiency expert as a necessary part of a progressively run school system.

¹² See Joel Spring’s discussion of Thronkide’s dream of “turning teaching into a scientific profession” as well as the use of tests to achieve his “social vision” (248).

¹³ To find the fifteen samples that would become standard markers on the scale, Hillegas and an assistant read seven thousand compositions by young people and rated them into ten classes. From the ten classes, 75 samples were selected and eight samples were artificially included ranging from examples that were purposely written poorly by adults to writings of Jane Austen and the Brontes when they were in their youth to compositions of college freshman. Hillegas then sent the samples to readers (teachers, authors, and literacy workers) and asked them to rank the samples in order in hopes of finding agreement in the quality of the samples. Due to the large number of samples, Hillegas replicated his experiment using only 27 samples which he sent to one hundred readers. Upon receiving seventy-five responses, Hillegas mapped the responses using a consensus method and assigned a rank to each of the compositions.

¹⁴ The values of the scale were “not as accurat[e] as millimeters express the length of lines” however, “a difference of two hundred in this scale is equal to twice a difference of one hundred taking in any part of the scale. The sample with a value of 369 is a little more than twice as good an English composition as the sample with a value of 183 (Hillegas Scale 13).

¹⁵ It is important to note that the zero point was actually an artificial sample created by an adult who attempted to write very poorly. The first three of the fifteen samples are all artificially created.

¹⁶ For a more in-depth comparison, Jon F. Love’s thesis spends more time with each of these scales.

¹⁷ Not only did students take standard tests under a controlled environment, but Certain sets out the importance of having teachers score compositions in a specific manner. Some of his directions include: subdividing the papers into good and poor piles; then subdivide the good papers as a set, and the poor papers as set. Once the range has been established teachers needed to match the student composition with specimens on the Nassau County Supplement to the Hillegas Scale and assign each text a value. Teachers were not to pay attention to the age or training of the student nor the speed at which they completed the composition (“Why” 473-4).

¹⁸ See David Tyack’s *The One Best System* for a history of how schools adopted a bureaucratic model of education.

¹⁹ On a national scale, reforming education through the adoption of standardized test has taken prominence within the political arena starting with the 1983 *A Nation at Risk* report, to George H. Bush’s 1989 call for a national education summit, to Bill Clinton’s 1994 Educate America Act, to the current No Child Left Behind Act (Gallagher 19-23).

²⁰ Starting in the 1970’s, in states like Florida, California and Texas, there were moves toward using test-based as exit exams and so the educational trend was growing (see Brossell for Florida). During the 1990’s a national trend of test-based accountability

particularly around the use of exit or graduation exams to ensure student, teacher and school accountability was on the rise. A logical follow up to standard-based reform efforts, of the 1980's and the movement of states to create "standards-based systems" of education by the mid 1990's, resulted in a boom in the standardized test industry. The educational testing industry grew from a \$263 million business in 1997 to \$924 million industry in 2004 (David and Cuban 18).

²¹ Marshall High school only allows students into Composition who have not passed the exam. This allows for smaller classes where teachers can give individual attention to students. Wilson high school mixes their students however they have far less students enrolled in composition who need to pass the exam.

²² See Smith, Resnick and Resnick, and Kohn for examples of the ways curriculum is narrowed by standardized tests.

²³ The multiple choice section on grammar and usage was created based on common errors found in the pilot version of the WGE.

²⁴ A majority of the scorers are high school English teachers who regularly teach courses where students have not yet passed the WGE. Middle school teachers are also used to score exams to help expose them to what is expected of their students when they come to high school.

²⁵ To support June's statement, in other district situations such as table leader meetings and scoring sessions it became clear that many of the same teachers were chosen to participate repeatedly while others were not invited to participate. This was particularly true when it came to scoring, as teachers who were "problem readers"—those who did not read according to the rubric, or were unwilling to norm their reading to the rubric and exemplars—were not asked back. The system is built around consent and agreement; teachers who dissent too much or raise the ire of those in charge of the system are not included.

²⁶ Based on my research the student's score is the primary means for placement.

²⁷ A component of General English, in lieu of a semester exam, the students retest by writing an expository and narrative essay. The multiple choice section of the exam is dropped and the students are given ninety minutes instead of forty-five minutes. Additionally, teachers are encouraged to let students test on computers in order to assist them with spelling, grammar, and revising. The students' essays from General English are scored as a separate and much smaller scoring session than the twice a year all district scoring sessions.

²⁸ There are nineteen Educational Service Units across the state. Most serve multiple districts however due to the largeness of BPS, the state has allocated the resources of one ESU solely to BPS.

²⁹ In Teaching and Assessing Writing Edward White names the New Jersey Basic skills testing program as well as the City University of New York as modeling their writing assessment on the model California State University implemented (197).

³⁰ Each of the scenes has been taken from different meetings. Though I refer to them as Table leader 1-5, they are different teachers and administrators in each scene.

³¹ I have purposely named Mark in this scene and in others. This is the same Mark Thurman who appeared in chapter two.

³² I have agreed to not reproduce any of the prompt topics verbatim because BPS may reuse these prompts in the future. Instead, I have permission to summarize the topic of the prompt and reprint the district's template.

³³ See Bob Broad's work designing the dynamic criterion map as an example of building up from teachers' knowledge to create an assessment tool for a specific educational context.