

THE RHETORICAL DEPLOYMENTS AND THEORETICAL ASSUMPTIONS
OF QUANTIFICATION IN EDUCATIONAL LAYPERSON TEXTS

By

Justin Neal Thorpe

A DISSERTATION

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

Curriculum, Instruction, and Teacher Education – Doctor of Philosophy

2013

ABSTRACT

THE RHETORICAL DEPLOYMENTS AND THEORETICAL ASSUMPTIONS OF QUANTIFICATION IN EDUCATIONAL LAYPERSON TEXTS

By

Justin Neal Thorpe

The following dissertation considers the relationships of rhetorical arguments and quantifications, particularly how quantification has become a rhetorical trope in educational research and writing. This dissertation seeks to contribute to how generalist texts argue to general audiences about the conditions and recommended changes in education through inclusion of and reliance upon quantified data. This dissertation looks at three generalist texts as objects of study to consider the rhetorical deployments of quantification. These texts are: Diane Ravitch's *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*, Abigail and Stephan Thernstrom's *No Excuses: Closing the Racial Gap in Learning*, and Tony Wagner's *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills Our Children Need—And What We Can Do About It*.

This dissertation begins with an experience that I had during the beginning semester of my graduate education in the Department of Teacher Education after changing from the Department of Statistics and Probability, both at Michigan State University. The introduction considers some of the influences in my personal journey and also influences in shaping this dissertation, including the National Research Council's text *Scientific Research in Education*. From this experience, the dissertation considers the language of quantification and the roles of quantification in changing how research is done in different areas; quantification has become common as accepted evidence in argumentation. This dissertation considers this proliferated acceptance as a potential space for quantitative illiteracies and shortcomings, with a potential to misuse the quantified data in making educational arguments.

The dissertation, written for an introductory educational quantitative research course, considers conditions of education through the lens of goal steering, particularly how predetermined educational outcomes are steering the curriculum, the teaching, and the learning. The dissertation considers the influence of the business concept of auditing and how the current educational trend is focused by an audit culture. The auditing of educational practice and performance and the goal steering associated with it come through measurements, particularly through standardized test data, and the applications of quantifications.

The dissertation provides two case-study chapters based on the three objects of study listed in the opening of this abstract. The first of these two chapters (Chapter Three) focuses on the rhetorical use of quantification to make predictions and to generalize. The second case-study chapter (Chapter Four) considers three rhetorical deployments of descriptive quantification in educational texts: the rhetoric of comparison, the rhetoric of transparency, and the rhetoric of the jeremiad.

The second to last chapter explores some of the assumptions associated with quantification, hoping to examine more the theoretical conditions that have led to a strong rhetorical presence of quantification in generalist educational arguments. This chapter considers how the rhetorical influences of quantification have been impacted by positivist notions: including the assumptions that it is possible to convert quality into quantity, through a discussion about qualia; the assumptions that qualities apply beyond the sample; the notion that the future will be like the past, i.e. that history cycles; the conflation of probability as a certainty; the views that humans can be studied like the natural world; and problems with predicting and forecasting.

I conclude the dissertation with a question of what types of evidence are required for what types of educational reforms. I offer a sibling type of evidence that might be used in educational argumentation. This chapter asks the question posited by Dutch educationalist Gert Biesta “Are we measuring what we value or valuing what we can measure?”

Copyright by
JUSTIN NEAL THORPE
2013

I would like to dedicate this dissertation
to my parents Charles and Audra Thorpe,
my best friend and sweet wife Alicia,
and to my family and friends.

ACKNOWLEDGEMENTS

There have been many people who have helped and supported me during this time of my life. I am especially grateful for my parents and family. They consistently asked how my life and work was going. It was wonderful to receive cards and letters telling how much I was loved and missed. The packages of cookies were especially helpful! I appreciate the love and support from a distance.

My friends have supported this work in many different ways, including an encouraging text or a friendly email.

I am grateful for the work of my advisor, Dr. Lynn Fendler. I do not think this work would have taken shape if it were not for her insights and commitments. She not only has played a pivotal role in helping shape the study but has also helped shape me as a scholar and human. I am grateful for her friendship and support. I appreciate her help when I began thinking about educational philosophy and engaging the world critically. I appreciate her thoughtful comments and willingness to let me explore as I grew as a scholar. Thank you, Lynn!

I am also grateful for the other members of my committee. Julie Lindquist helped me to consider more what writing could be and the enjoyment that comes from language; from her I also learned about teaching in ways that invite contribution. Vince Melfi helped me to consider more the statistical roots that I have; he has taught me how to be considerate of others' ideas, even if they do not align with my own personal thoughts. Thank you, Vince, for helping me through my time in statistics and in education, seeing me through a practicum, coursework, a dissertation proposal, and the insights into the revisions of this dissertation. Steve Weiland has helped me to think about the contributions of my individual work; he has helped me to think about the value of the humanities and rhetorical scholarship.

Sandra Crespo has not only been a member of my committee but also a principal investigator on the project that I worked for during my graduate studies. Not only has she helped my work in this dissertation but also has affected me as a scholar. I appreciate her willingness to let me learn about researching in education, including helping design protocols, conducting interviews, and making visits to the classrooms. I am especially grateful for her willingness to let me explore in my own way the data we were collecting. I think it was in her data that I really began to explore rhetorical scholarship. Thank you! I will always appreciate the trips made and the times spent together in collecting data, writing field notes, and eating great food. Long live the key lime pie!

The other members of the PIR team have been an amazing help in my scholarship and growth. Thank you, Aaron Brakoniecki, for the many great lunches and conversations. Thank you, Ann Lawrence, for helping me learn and develop the craft of writing. Thank you, Leslie Dietiker, for your thoughtfulness and openness. Thank you, Joy Osland and Curtis Lewis, for showing me that it is possible to complete this work with a smile. To the Honorary PIR member Sharon Mills, I owe so much as she helped me to navigate the red tape of Erickson Hall and was always there with a smile and so much encouragement.

Thanks to Adam Greteman who read many drafts of this document and helped me to think beyond myself.

Finally, I wish to express my deepest gratitude for my wife Alicia. I recognize that she has joined this process later than others, but I appreciate her support and love through the revisions and formatting. She has helped me to carry-on and finish this dissertation. I am immensely grateful for her smile and support and her love during the long nights. She helped me with the organization of the Table of Contents, which would have been tedious without her. I love you, Alicia!

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
A PERSONAL INTRODUCTION	1
<i>Introduction to Educational Research</i>	<i>1</i>
<i>Educational Philosophies.....</i>	<i>4</i>
<i>The Painter or the King (or Someone Else).....</i>	<i>6</i>
<i>The Language of Quantity</i>	<i>8</i>
 CHAPTER ONE	
THE LANGUAGE OF QUANTITY	11
<i>Moneyball</i>	<i>14</i>
Quantitative Il/Literacies	15
Rhetoric	18
<i>A Rhetorical Trope of Quantitative Literacy.....</i>	<i>20</i>
Three Objects of Study	22
Summaries of the Three Books	24
<i>The Death and Life of the Great American School System.....</i>	<i>24</i>
<i>No Excuses</i>	<i>25</i>
<i>The Global Achievement Gap</i>	<i>26</i>
Outline of This Dissertation.....	30
 CHAPTER TWO	
AUDITING EDUCATIONAL OUTCOMES: QUANTIFICATION'S RELATIONSHIP TO THE AUDIT CULTURE IN EDUCATION	33
Goal Steering: A Product of the Relationship Between Quantification and Outcomes- Based Education.....	33
Goal Steering through Measurement.....	35
Measurement & Calibration.....	43
<i>Human Qualities and Quantities</i>	<i>44</i>
<i>Qualia</i>	<i>45</i>
We Live in an Audit Culture.....	46
On the Epigraph.....	56
 CHAPTER THREE	
RHETORICAL CONSIDERATIONS OF PREDICTION AND GENERALIZATIONS	59
Inferences	65
Rhetorical Future Predictions: Predicting the Conditions of School Systems	69
<i>The Death and Life of the Great American School System.....</i>	<i>69</i>
<i>No Excuses</i>	<i>71</i>
<i>Critical Considerations of Predicting</i>	<i>73</i>
<i>Accuracy, Precision, and Construct Validity.....</i>	<i>77</i>

Arguing through Generalizations: Generalizing School Systems	80
<i>No Excuses</i>	80
<i>The Global Achievement Gap</i>	81
Beyond the Sample?	83
<i>Critical Issues of Generalizing</i>	84
Conclusion.....	89
 CHAPTER FOUR	
DESCRIPTIVE RHETORICS OF	
COMPARABILITIES, REVEALINGS, AND JEREMIADS	91
Different Deployments of Descriptive Quantification in the Three Books	93
A Rhetoric of Descriptive Comparabilities	94
<i>Thernstrom & Thernstrom</i>	94
<i>Ravitch</i>	97
<i>Wagner</i>	98
<i>Comments and Considerations</i>	100
A Rhetoric of Descriptive Transparency	103
<i>Arguing for School Proficiency Transparency</i>	103
<i>Arguing about the Transparency of Choice</i>	106
<i>Arguing for Transparency of Global Knowledge</i>	109
<i>Arguing with the Transparency of Test Scores</i>	113
<i>Comments and Considerations II</i>	116
A Rhetoric of a Jeremiad	117
<i>How Did the Home Team Do?</i>	118
<i>A Shamefully Ignored Issue</i>	121
<i>Comments and Considerations III</i>	122
Rhetorical Descriptive Quantifications.....	123
 CHAPTER FIVE	
YUCK: A CHAPTER ON THE ASSUMPTIONS OF QUANTIFICATION.....	126
An Experience	127
Quantifications and Statistics	128
Descriptive Quantifications.....	131
<i>Converting Qualities into Quantities</i>	131
Inferential Leaps.....	135
<i>Assuming Qualities beyond the Sample</i>	136
<i>Assuming the Future will be Like the Past</i>	140
<i>Conflating Probability as Certainty</i>	143
<i>Positivistic Views that Humans can be Studied Like Natural World</i>	146
<i>The Problems of Predicting and Forecasting</i>	149
Governance in Education.....	151
 CHAPTER SIX	
WHAT KINDS OF EVIDENCE FOR WHAT KINDS OF REFORMS?	157
An Age of Educational Reform	159
Measuring What?.....	162
A Sibling Rhetoric	169

APPENDIX	176
REFERENCES.....	181

LIST OF TABLES

Table 1. Support or Opposition to Standardized Assessments	114
Table 2. Comparison of Quantification and Statistics.....	129

LIST OF FIGURES

Figure 1. Mean NAEP Scores Demonstrating Four Year Racial Gap.....	95
--	----

A PERSONAL INTRODUCTION

When I first started my graduate studies at Michigan State University, I was enrolled in the Department of Statistics and Probability. I had completed work as an undergraduate in mathematics and statistics, and considered getting a Ph.D. in statistics and probability to be a logical next step. However, as I worked in the department, I realized I was thinking about statistics education and implications in educational research instead of the theories of statistics and probability. This prompted a change of academic venue to education. Although much more could be said about my change, it is of little consequence for understanding what lies ahead. However, I was able to gain admission to the Ph.D. program in the Department of Teacher Education at Michigan State University.

In my first semester in the teacher education program I took courses that met some of the requirements specified by the college and the department. There was really no explicit reason for taking the courses at that time, but I believe the mixture of those courses has had profound impact on my development in as an educational researcher. In that semester I took two research method courses (a course in qualitative research and an introductory course in educational research) and a course in the philosophy of education. Although I learned a great deal from the qualitative course, I only mention briefly experiences from the other two.

Introduction to Educational Research

In MSUs College of Education, all Ph. D. students across the college must take this course that introduces the concepts of educational inquiry and research. Although different instructors focus on different aspects and qualities, the semester in which I was enrolled had a mixture of qualitative and quantitative studies and experiences, although not to have us complete studies using

those methods but to introduce us to different methods in order for us to begin thinking about future work. In that course, I was required to read a report published by the National Research Council on scientific research in education, a report that suggested what counted as rigorous research regarding US education.¹

This report haunts me.

Why might I suggest that this report haunts me? I would think that anyone who was coming from a background in mathematical statistics, a background that considered the empirical nature of research as part of the search of knowledge. I didn't know at the time, but I was struggling in part to understand my own considerations of how knowledge is understood and accrued and how there are different methods for understanding. I can see now, after five years in a graduate program that has taken me through diverse courses and experiences, that the term haunting is appropriate for this work by the National Research Council. It seems to me that this report holds to empirical studies, studies that consider the accumulation of knowledge through six principles: (1) Pose questions that can be considered empirically, (2) Link research to relevant theory, (3) Use methods that permit direct investigation, (4) Provide a coherent and explicit chain of reasoning, (5) Replicate and generalize across studies, and (6) Disclose research to encourage professional scrutiny and critique.²

I do not think that my response to this document is solely based in the definition of what counts as scientific. This document can, of course, claim what it considers to be scientific. The haunting nature of the report is that it claims *what works* in educational research, limiting the research only to that which can be directly investigated and replicated to create a generalization across contexts and conditions. This is not to suggest that educational research cannot be informed by

¹ Committee on Scientific Principles for Education Research, *Scientific Research in Education*, ed. Richard J. Shavelson and Lisa Towne (Washington, D.C.: The National Academies Press, 2002).

² Ibid., 3–5.

studies that would be classified by the National Research Council as scientific; however, educational studies are also informed by research that may not hold to one of the criterion outlined above. I believe that contributions to educational research can indeed come through studies that are not considered by this list to be scientific. This dissertation will draw on that assumption. Most would not consider this a “scientific” study.³ I embrace that claim and suggest that this study contributes to the understanding of educational writing and research, particularly through inclusion of quantified data. This is not a study of what works in writing in education with quantification; instead, this is a work that explores educational conditions through the writing that is used to portray it.

The NRC report was published in 2002, when the United States government was implementing the first stages of *No Child Left Behind* and it seemed to me that the standards of what counted in education and in educational research was that which could be measured empirically and then generalized to other situations and studies. Yes, I can accept that as a way that knowledge about education is considered, but I could not help but think that it limited what questions could be asked and the methods that could be used to find answers.

The NRC report outlines that there are scientific principles that guide the work of educational research. These principles are commonalities of scientific research, whether that research was taking place in biology or physics, anthropology or economics, etc. The purpose of these principles in this report, moreover, is to suggest that education can also adopt/adapt these principles in effort to determine what scientific research in education is, in attempts to further allow consensus among teachers, policymakers, advocates, among others. Educational research was being portrayed

³ AERA, “Standards for Reporting on Humanities-Oriented Research in AERA Publications,” *Educational Researcher* 38, no. 6 (September 2009): 481–486; AERA, “Standards for Reporting on Empirical Social Science Research in AERA Publications,” *Educational Researcher* 35, no. 6 (September 2006): 33–40.

as a science, although by academic communities, its status as a science would remain debatable, if debated at all.

Also in this introduction to research class I was exposed to some research that employed quantitative methods for researching educational issues. I had come from an intensive background in statistics and probability. Most of my undergraduate courses were in some form of statistics or probability. I had taken courses that were beyond just introductory courses, including courses in the theoretical underpinnings of how statistics were computed. I came to this course having some understanding of what statistical tests were available and appropriate. I came to this course have some understanding of the assumptions that needed to be met in order to run those statistical tests and to make statistical claims. I had learned about the nature of probability and how it relates to the tests that were being run, only to be shown examples in the form of articles and studies that assumed conditions were adequate to run the study or make sweeping claims from samples that were insufficient or did not meet the requirements for use of the probability models. From my perspective as a student of statistics, the examples of quantitative educational research were poorly done.

Educational Philosophies

During this same first semester in The Department of Teacher Education, I was taking a course in the philosophy of education. In this course I was introduced/reminded of thoughts and ideas that shaped how I read and think about the world and particularly how education is viewed and fits within that world. In that course I read philosophical works, like Aristotle's *Rhetoric*, a book that has been a mainstay in my approaches to understanding not only the nature of argument and rhetoric but in also understanding the nature of language and communication, Descartes' *Meditations*, and Rousseau's *Emile*. In this course I was exposed to postmodern ideas of Foucault and the post-

linguistic considerations and use of language. I was introduced to educational researchers like Gert Biesta who work in writing about the qualities of education instead of just the measured quantities of education.

I was also introduced to new ways of thinking, particularly about systems, relationships, and the production of knowledge. In doing so, I found that I was able to think about how issues are portrayed and accepted. I began to consider issues that relate to ethics, particularly as ethics and values relate to the systems of education. I began to look into what is accepted as knowledge and how those acceptances can be challenged and accounted.

I was taking a course that required me to consider the methods of educational research and a course that challenged me to think beyond my normal considerations. I was beginning to recognize the different ways and different purposes in which education is portrayed, argued, and received. Education became more than simply learning of core subjects like mathematics and language arts. I began to consider more aspects of educational research that were not contained within the walls of a classroom, filled with students and a teacher trying (micro)manage. I began to see that educational research was more than what happened in schools but also included critique of the structures and systems of education, which for me included—and still includes—the language and rhetoric of education.

Educational research, for me, is a complexity of more than simply the acts of teaching and the work of educating future teachers. It seems in teacher education that the focuses of educational research become imbued with searching for the best ways, or put in the language of business—the best practices, to teach a mathematical algorithm or to improve a teacher’s ability to explain how to use context clues. There is potential for the focus to become one of educational practice, considering what practices work “the best” in attempts to understand how to replicate in other situations and contexts Educational research, in this practice-driven plane, becomes a sight for only

looking at the ground level, only considering the next step to meet standards or to improve measures or to find a new method of teaching and learning (with a potential for new marketization) with the ultimate goal of improving test scores and maintaining demonstrated growth and improvement. Yes, I believe that there is much to consider in looking at the ground level, but educational research cannot forget that there are other ways to consider the same picture or scene.

The Painter or the King (or Someone Else)

In his *The Order of Things*, Michel Foucault opens by beautifully considering the different perspectives available in understanding Velasquez's painting *Las Meninas*. In Foucault's critique of this painting, he considers the different representations and figures that are presented, and how the different figures change focus and perception of the painting. Foucault highlights the figure of the painter in the left and what it is he is painting; Foucault considers the mirror in the center which reflects the subject of the painting being done by the painter, the king and queen. There is the man in the background who is interrupting the painting, drawing the outside world in; there is the cluster surrounding the little princess and her entourage who draws attention away from the painter again to that of the king and queen.⁴

Within the completeness of this painting, the people who are there and the careful constructions of lines and motion that are portrayed, Foucault suggests that there is a "vacancy" which can never represent all that is occurring in the picture. Foucault concludes his analysis by suggesting

It can never be present without some residuum, even in a representation that offers itself as a spectacle. In the depth that traverses the picture, hollowing it into a fictitious recess and projecting it forward in front of itself, it is not possible for the pure felicity of the image ever

⁴ Michel Foucault, *The Order of Things: An Archaeology of the Human Sciences* (Routledge, 1970).

to present in a full light both the master who is representing and the sovereign who is being represented.⁵

Although Velasquez tries to include himself in his work as well as the portrait of the king and queen who is not physically in the picture but represented through their portrait being painted and their reflection in a small mirror, it is impossible, according to Foucault, for the painting to present both completely and accurately. The king and queen are not presented in full light and they cannot be with the painting focusing not on them. I am beginning to understand this idea of presenting in full light in educational research and writing. Who is the master portraying education and who is the sovereign of education being portrayed? It is not possible for education to be presented in full light. It is even more impossible to portray education in full if the only perspective that is considered to work is the empirical study of education. Education is a complexity, a complexity with diverse qualities and quantities.

In the introduction-to-research course I mentioned earlier, research was broadly defined as quantitative or qualitative.⁶ However in the philosophy of education seminar, I was exposed to humanities-oriented work. All three existing in the field; all three offering insight into the portraits and receptions of education. However, I began to notice in my courses and readings that much of the speaking in and about education came from numbers. It appeared to me that the communications of education were pivoted toward quantifying the qualities in the classrooms so

⁵ Ibid., 16.

⁶ These terms are often used in educational research to denote two types of empirical data collection. I write these terms from the standards published by AERA, where quantitative work relates to descriptive or inferential statistics and qualitative work relates to work that investigates through interviews, observations, and written responses. I recognize that there are many who use qualitative methods while counting, such as when a report claims that 3 out of 5 interviewees said X. I do not consider this a quantitative study, although counting has been done. For the this dissertation, quantitative work concerns research questions that are investigated through descriptive or inferential statistics.

that results and practices might be generalized into other settings. In fact, in Michigan State University's College of Education, all PhD students are required to take a course in quantitative methods while qualitative methods courses were optional and humanities-oriented methods course was not offered until an experimental course was offered my fifth year.

This dissertation considers (and embraces) education as a complexity, although it only focuses on limited aspects of this complexity. In the writing that follows, this dissertation will consider the language of quantity and the rhetorical trust that has come from the use of quantification. I do not think that there is a problem with rhetorical use of quantification; the difficulty is when it becomes the standard of believability in educational arguments that important perspectives are lost.

The Language of Quantity

In the preface to *Trust in Numbers*, the question “What is special about the language of quantity?” is asked by Theodore Porter to understand how quantities have become a major currency in communication economies.⁷ Porter considers the history of quantification and statistics and how quantification is useful in transcending boundaries of communities and localities. His works not only consider how numbers have become a source of trust for those seeking answers, but he considers how society has gained trust in numbers through a seemingly reverse direction by considering the political, economic, and business influences and acceptance of quantification instead of the influences of the natural sciences on society's acceptance of quantification.

⁷ Theodore Porter, *Trust in Numbers* (Princeton, NJ: Princeton University Press, 1995), ix.

Porter provides a key understanding for my interest in pursuing this dissertation. He considers “numbers, graphs, and formulas first of all as strategies of communication.”⁸ Porter contends that these numbers and relating items are not simply quantities but ways to communicate within communities that have accepted bounds and practices. I agree that numbers have communicative properties, having abilities to aid in the persuasive appeals within those communities. However, I am interested in how those quantities are used in shaping arguments, particularly how they are used to shape educational argumentation.

This dissertation is a personal fulfillment, a project that has allowed me to look into areas that I find important in considering education. Although I did not continue my studies in statistics and probability, I am still concerned with how statistics are used, particularly in maintaining the integrity of the statistics through their proper use. One of the difficulties that I have had in educational classes is when assumptions that relate to the appropriateness of the test or probability model to be used in analysis are overlooked for the sake of ease or audience interpretation. I maintain that quantification is a vital part of educational research and that can provide insight when properly used. I must also recognize that I am concerned with the human condition and the questions that can only be considered through humanities-oriented research. I think that there are questions about education that are meant to explore human existence and make shifts in human desires. This dissertation is allowing me to mingle both worlds.

In this dissertation, I consider a question that has followed me from my initial readings in an introductory research course that I took in my first semester coupled with the broadening of ideas and considerations in a philosophy course. This dissertation is a consideration about how it is possible to measure qualities through calibrations in attempts to allow for comparison and

⁸ Ibid., viii.

evaluation and ranking.⁹ I see this as an ethical question, a question that often gets overlooked in attempts to understand education in replicable and scientific ways. For me, understanding statistics helps me be a better philosopher, and understanding philosophy helps me be a better statistician. What follows is some of the thinking that I have done in this regard. I recognize that it cannot be all of my thinking, but I consider a dissertation not an ending but a passageway into future and further questions and thoughts and interpretations.

⁹ Calibration is a common theme of this dissertation. I will discuss this term in more detail in Chapter 2. However, for the purposes of this preface, I consider calibration in relation to how the uncountable become countable (and counted) through the use of standardizing and comparing against this standard.

CHAPTER ONE

THE LANGUAGE OF QUANTITY

My approach here is to regard numbers, graphs, and formulas first of all as strategies of communication. They are intimately bound up with forms of community, and hence also with the social identity of the researchers. To argue this way does not imply that they have no validity in relation to the objects they describe, or that science could do just as well without them....What is special about the language of quantity?

- Theodore Porter¹

Education affects more than just the students and teachers in the classrooms; it is considered and discussed by everyone. Pardon my hyperbole in the previous sentence. Yes, I recognize that education is not considered by *everyone*, as there are many who pay no thought to the condition of education. However, education is beyond just an academic discipline but is considered by parents who desire certain outcomes for their children and by politicians who consider economic and monetary educational products and by businesses who want to invest in how education functions for a future return in the investment and by others. One might consider the purposes of education that David Labaree highlights in his work on how education can be seen as a public good as well as a private good.² Because education has garnered opinions from diverse sources, there are many depictions written to explore, exemplify, or condemn activities associated with education, often written toward a (fractured) truth of what education *is* or what it *should be*. These writings range from practitioner research to the opinions pages in local newspapers and magazines to books to popular cinema.

Opinion pieces in newspapers and Internet sites are devoted to ideas of what is wrong and right with education in the United States. Educational topics continue to be discussed because of the constant need for hope in the future. Education became and continues to be a topic that is

¹ Porter, *Trust in Numbers*, ix.

² David F. Labaree, "Public Goods, Private Goods: The American Struggle Over Education," *American Educational Research Journal* 34, no. 1 (1997): 39–81.

considered by the lay public and the specialist. However, it is not a field in which immediate agreement about the *hows* and *whys* can (if ever) be reached. There are many stakeholders that have interest in how education functions and the outcomes that are produced. These opinions are developed and shaped through the intersections of these varying ideas and suggestions about education. It is within these intersections I read educational texts as sources that shape the perceptions of education.

It is among vast differences in stakeholders, vast differences in purposes, and vast differences of opinions about educational success that this dissertation takes shape. Among these differences are arguments about what education might become and how those who participate in education might get there. I do not consider argument to be a negative word, although it is sometimes connected with fighting. I consider argument to relate to the skills of persuasion, which is about engaging with others in an attempt to change, whether to change opinions or to change conditions or to change interactions. Argument clarifies where persuasion changes.

I consider the persuasive arguments of educational change to be fertile ground for rhetorical consideration and comparison. Educational change often gets labeled under the term “educational reform.” Texts on educational reform create spaces where an author considers the wrongs of the current educational system (and for the purposes of this dissertation, I will consider the educational system of the United States) and engages in argument to persuade changes to and in that system. This is not the only purpose of educational text, writ large, but the purpose of educational reform texts are to challenge the current notions of educational practices or policy in attempts to make a better system, well better at least in the eyes and opinions of the reformer.

I am interested in persuasion. I believe that texts are interpreted with purposes, although impossible and impractical to limit to one interpretation of that purpose. I believe that this interpretation is also part of educational texts. This dissertation is about these texts, although there

are many different foci within this broad categorization of educational texts. There are texts that are written about the acts of teaching/managing; there are texts that are written about the growth and development of the child; there are texts that are written about the current state educational policy; there are texts that are written about educational research; there are texts written about...and so forth. I could go on. It is sufficient for me to suggest that these texts about education are forms of persuasion of *what* or *how* education might be. I am concerned with the persuasive rhetoric of these texts about education.

There are many lenses of rhetorical analysis that could be used to analyze educational reform texts, such as considering through the lens of post-structural feminism or through an analysis of the metaphors and similes used. I could have considered the arguments through the logos, ethos, and pathos or the appeals to policymakers. In this analysis, however, I consider the rhetorical use of quantification, particularly as quantification as a rhetorical trope. In doing so, I am exploring how quantification is deployed in educational reform texts. I do not suggest that this work is a complete listing of the history or use of quantification, but a consideration of the ways quantification functions in persuasive texts. This dissertation will consider two roles. First, I will consider the rhetorical implications of generalized quantification (Chapter 3) and descriptive quantification (Chapter 4) and some of the theoretical issues that result from using quantification as a trope in educational arguments (chapters 2, 5, and 6). I do so because in order to consider the rhetorical weight of quantified evidence, I believe it is important to consider the theoretical and historical underpinnings that have led to such acceptance.

Moneyball

It seems that the actor Brad Pitt is a drawing card for viewing audiences. He did after all create interest in the Trojan War and showed audiences a possible world of a man who ages in reverse. However appealing these movies are, I mention Pitt in his connection to the 2011 film *Moneyball* based on the non-fiction work of the same title by Michael Lewis.³ In this film, which is based on actual events, Pitt portrays a baseball manager who uses statistical analytics, often called sabermetrics, to change the poor, losing Oakland Athletics into a winning team.

The story focuses on Oakland's general manager, Billy Beane, and how his team used different statistical analyses than simply considering the number of stolen bases or a batter's batting average. Beane and his staff considered ways to successfully compete against richer teams in Major League Baseball while maintaining the limited budget that was available. Through rigorous statistical analysis, the leadership was able to determine that things like the on-base percentage and slugging percentage were better indicators for offense players, allowing Oakland to consider cheaper players than the more expensive players who were valued because of their speed or throwing abilities. The results of these complex analyses went against the conventional wisdom and experiences of baseball "insiders" and experts. This change of wisdom provided Oakland with a competitive edge against teams who spent more in acquiring and training players.

The events portrayed in *Moneyball* did not necessarily introduce quantification into the game of baseball; the difference is that the sabermetrics introduced by the Oakland Athletics transformed baseball from considering players who were fast or threw at certain speeds into a sport that considers certain indicators through more aggregate (and possibly complex) statistical measures.

³ *Moneyball*, directed by Bennett Miller (2011; Burbank, CA: Sony Pictures), Film; Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (W. W. Norton & Company, 2004).

Baseball teams began to consider how certain indicators relate and to establish a team for the cheapest price possible. I guess even Brad Pitt can make quantified data and analyses seem sexy.

Moneyball suggests that quantification has impact in changing how things are conceived and done in practice. The game of baseball changed because of the inclusion of statistical analysis and computation. Baseball now is influenced by the inclusion of such quantifications and analyses. Baseball is not the only area of study that has been influenced by the integration of quantification and statistical analysis. Quantification has gained acceptance in diverse fields, promoting a confidence and faith in the numbers, and possibly, as *Moneyball* highlights, a change in how things are done because of what the numbers suggest. However, with the faith that comes in numbers there is also the potential to rhetorically use the numbers unintentionally. Theodore Porter suggests “standard statistical methods promote confidence where personal knowledge is lacking. They are also used to train and discipline outsiders, such as students and uncredentialed assistants.”⁴

Quantitative II/Literacies

I admit that this dissertation is being written around the time of the 2012 election cycle. It seems as though everything revolved around the election and which candidate was leading among which demographic or gaining electoral edge in what swing state (as far as the presidential election between Barak Obama and Mitt Romney was concerned). It seemed as though each different poll had its own reading and interpretation of polling numbers and what they meant for the future of the country.

Polling interests me because my background as an undergraduate student, and the beginnings of my graduate work, was in statistics. As part of this background I had taken statistical theory and analysis classes that shaped in part my thinking about how to quantify and measure. Polls

⁴ Porter, *Trust in Numbers*, 200.

seem to measure the attitudes of the potential electorate: what is being valued and what is not. Polls can be found accurate at the end of an election, but can also be off as exemplified in *Literary Digest's* poll in 1936 that suggested President Franklin Roosevelt would lose to the Kansas Republican Alfred Landon.⁵

One of the most interesting things about polls for me is the misuse that accompanies them. I enjoy sitting in conversations and hearing how people's opinions are swayed by polls and their interpretations of what these polls mean. It almost suggests a quantitative illiteracy. In order to consider this term more, it might help to consider what quantitative literacy entails to consider more what it is to be illiterate. In doing so, I recognize that I am not a literacy scholar, so I will draw from concepts of Jon Star, Sharon Strickland, and Amanda Hawkins.⁶ They describe mathematical literacy, and I wish to adapt their concepts to consider more quantitative literacies.

Star, Strickland, and Hawkins consider two different thoughts about content-area literacies. The first is content-area *literacies*, where the focus is on the literacy. They suggest that this is a notion of developing literacy concepts within different content areas, such as reading mathematical texts in a math class to develop literacy skills and not necessarily mathematical ones. The second is *content-area* literacies, where the emphasis is on being able to understand and interpret the content, in this case mathematics. In an analysis of the term mathematical literacy, Star, Strickland, and Hawkins found

that *mathematical* literacy can hold several different but interrelated meanings. First, about three-quarters of the articles indicated or implied that *mathematical* literacy is synonymous

⁵ Peverill Squire, "Why the 1936 Literary Digest Poll Failed," *Public Opinion Quarterly* 52, no. 1 (1988): 125–133.

⁶ Jon R. Star, Sharon Strickland, and Amanda Hawkins, "What Is Mathematical Literacy? Exploring the Relationship Between Content-Area Literacy and Content Learning in Middle and High School Mathematics," in *Meeting the Challenge of Adolescent Literacy: Research We Have, Research We Need*, ed. Mark W. Conley et al. (New York: Guilford Press, 2008), 104–111.

with mathematical *understanding*, including knowledge of content and the ability to approach mathematical problems (such as those seen in mathematics texts) logically, analytically, and thoughtfully. Second, about half the articles indicate that mathematical literacy included an *appreciation* of mathematics, including the ability to recognize when and how mathematics is used in the real world. For example, being mathematically literate could include noticing how mathematics is used in stores, restaurants, and newspapers. Third, about one-third of the articles suggested that mathematical literacy involves the *application* of mathematics to real-world problems, including calculating tips in restaurants, working with budgets, and reading graphs in newspapers. Fourth, about one-fourth of the uses of mathematical literacy were related to the ability to *reason* mathematically, including mathematical communication.⁷

Mathematical literacy thus could be conflated with any part or all of four considerations. Star, Strickland, and Hawkins suggest that there is a sense of understanding the mathematical content (such as being able to understand the formulas or algorithms). I suggest, analogously, that a quantitative literacy would also be used to understand the workings of quantification, being able to do the work of quantification. The second point for Star, Strickland, and Hawkins is that there is an appreciation of mathematical ideas in real life, such as when it should be used. Analogously, quantitative literacy would be a literacy that invokes appreciation of when quantifications are used in real settings. Third, there is an application of mathematics with real world problems. I suggest that in a quantitative literacy would require considerations of when to use quantification in the real world such as reading graphics and statistics in reports. Finally, Star, Strickland, and Hawkins suggest that mathematical literacy includes being able to reason mathematically. For quantitative literacy, I consider this as being able to think critically about quantitative deployments.

⁷ Ibid., 110.

It is in this ability to understand, apply, appreciate, and reason quantitatively that this dissertation considers the rhetoric of quantification in educational reform texts. I suggest that considering the rhetorical nature of quantification is a type of quantitative literacy. This dissertation considers the rhetorical deployment of educational texts.

Rhetoric

Educational reform texts are filled with different appeals and rhetorical devices of persuasion. Rhetoric, as considered by the ancients, was an art of making decisions and persuading others. Rhetorical studies were integral parts of Greek and Roman learning and instruction, finding the rhetorical teachers mixing with the young men in spaces often associated with military and physical training.⁸ It was the ancient rhetor Aristotle who considered rhetoric to be “the power of observing the means of persuasion on almost any subject presented to us.”⁹ Aristotle’s treatise on rhetoric is for me the foundations of rhetorical influence in Europe and the United States. This art of persuasion is not lost in the contemporary deployment of language, although the terms associated with it may get maligned due to the nature of changing opinions and the negativity that gets associated with it. Although rooted in the skills of the ancients, the rhetoric of the United States is influenced by the culture and contexts of the United States. It is not possible to separate the contexts from rhetorical deployments. “Rhetoric functions within a culture,” Sacvan Bercovitch suggests. “It reflects and affects a set of particular psychic, social, and historical needs.”¹⁰ Educational reform rhetoric also functions within a particular culture, one of comparison and change. There are within educational reform rhetoric complex interactions of what it means to learn,

⁸ Debra Hawhee, *Bodily Arts: Rhetoric and Athletics in Ancient Greece* (University of Texas Press, 2004).

⁹ Aristotle, *Rhetoric*, trans. W. Rhys Roberts, 1954, 1355b.

¹⁰ Sacvan Bercovitch, *The American Jeremiad* (Univ of Wisconsin Press, 1980), xi.

how that learning is demonstrated, and the historical influences of what those mean. In writing about educational rhetoric, I do not diminish these psychic, social, and historical needs; instead I write through recognizing their influence in the rhetoric.

Rhetoric, unfortunately, has become almost a negative term in the public discourse. I say unfortunately because of the great potential I see in the study of rhetoric for analyzing argumentation and potentials within rhetorical analysis to push the boundaries of knowledge in productive ways. Rhetoric has become a synonym of flowery language and deception, language that is often associated with political spin and maneuvering. The connotations of rhetoric almost seem to suggest a separation of language from action—a distancing from doing. So much so that an introductory textbook on ancient rhetoric suggests: “Rhetoric is characterized as ‘empty words’ or as fancy language used to distort the truth or tell lies.”¹¹ It is a shame to me that this art of persuasion and argumentation has been diminished to a characterization of just fancy language that is used for (purposeful) deception. This has not always been the case.

Rhetorical analysis has seen resurgence in the twentieth and twenty-first centuries, allowing for considerations of texts through the arguments presented. In considering rhetoric, there is a recognition of humanity, particularly the public nature of humanity which is understood through complex contextual interactions. Rhetoric exists in more than in communication or English departments within the university. As rhetoric is about persuading, and rhetoric takes place in texts across academic fields and disciplines. Rhetoric is more than sermons found in the local church or the political speeches of a candidate. Michael Leff suggests

Rhetorical discourse occurs in contexts where judgments must be rendered about specific matters of communal interest. Such judgments normally invoke the general principles that

¹¹ Sharon Crowley and Debra Hawhee, *Ancient Rhetorics for Contemporary Students*, 4th ed. (White Plains, NY: Longman, 2008), 1.

categorize and direct our response to public events. Yet the application of the principles is open to question, and, in their abstract state, they are insufficient to allow for an adequate decision in any given case. Moreover, these principles themselves are subject to revision in light of our concrete experience. Consequently, rhetorical judgment cannot suffer reduction to strictly formal or methodical procedures.¹²

Leff is suggesting that rhetoric is found in spaces where judgments are made, particularly about communal interests. These decisions are based on certain principles and norms that are common to the community; however, these norms for which judgment is made are not automatic without the inclusion of argument. These decisions are, importantly, able to be modified and revised depending on contexts and outside considerations.

This dissertation concerns the rhetoric of educational reform texts, a space of communal interest where there must be judgments about conditions and recommended changes. Educational reform texts are written for the purposes of creating and espousing change in the current (and future) educational conditions. These judgments about educational reform are not resolved due to the norms of the society or laws of the country. Recalling that Leff suggests that Rhetoric exists in spaces where judgments are to be made without being made by the common norms of the community, educational reform texts are such a space. This dissertation considers this rhetorical space through the deployment of quantification as a rhetorical trope.

A Rhetorical Trope of Quantitative Literacy

Recent work on quantification in educational research has addressed its affordances and constraints, issues relating to generalizability and validity, and the debated value of quantification

¹² Michael C. Leff, "The Habitation of Rhetoric," in *Contemporary Rhetorical Theory: A Reader* (New York: The Guilford Press, 1999), 61.

when compared to qualitative methods, assuming that educational research is able to become as rigorous as sciences.¹³ Regardless of these prior debates, quantification continues to be a trope in educational texts, especially concerning issues that some label educational reform. The research on quantification and education has provided important questions regarding the acts of education, however the emphasis of the quantification debates has been studied through epistemological lenses, causing divide in the appropriateness of diverse methods of research in educational scholarship. Because of this type of scholarship and its transmissions to the public, perceptions about education are influenced by such narratives. In an age of measurement, educational perceptions are influenced by the arguments of educational texts, shaped by stylings and ornamentations, such as rhetorical quantification. How might quantification be used within arguments to shape the layperson's views and perceptions of education? This question is the central theme I explore further in this dissertation.

Arguments that use quantification assume some familiarity with quantification from the audience, a type of quantitative literacy. This dissertation provides a different consideration of quantitative literacy. I recognize that I have limited using statistics in this prospectus for the broader category of quantification; however, I purposefully use it momentarily. Statistical literacy is “an individual's or group's ability to understand statistics. Statistical literacy is necessary for citizens to

¹³ See Cleo H. Cherryholmes, “Construct Validity and the Discourses of Research,” *American Journal of Education* 96, no. 3 (May 1, 1988): 421–457, doi:10.2307/1084999; Cleo H. Cherryholmes, “Theory and Practice: On the Role of Empirically Based Theory for Critical Practice,” *American Journal of Education* 94, no. 1 (November 1, 1985): 39–70, doi:10.2307/1085291; David F. Labaree, “The Lure of Statistics for Educational Researchers,” in *Educational Research: The Ethics and Aesthetics of Statistics*, ed. Paul Smeyers and Marc Depaepe, vol. 5 (Dordrecht: Springer, 2010), 13–25; Paul Smeyers and Marc Depaepe, “Representation or Hard Evidence? The Use of Statistics in Education and Educational Research,” in *Educational Research: The Ethics and Aesthetics of Statistics*, ed. Paul Smeyers and Marc Depaepe, vol. 5, Educational Research (Dordrecht: Springer, 2010); Patti Lather, *Engaging Science Policy: From the Side of the Messy* (New York: Peter Lang Publishing, Inc., 2010).

understand material presented in publications.”¹⁴ To consider statistical literacy in light of quantification, I offer a version of quantitative literacy that might be viewed similarly as the ability to understand quantification, which is necessary in understanding presented material. Instead of only being able to understand quantification used in publications, such as in the objects for study, this dissertation considers a different type of literacy, namely a rhetorical literacy which includes abilities to interpret how quantification ornaments educational arguments. This dissertation offers a different view of quantitative literacy allowing for interpretations of argument, not only in understanding what the quantification means but also how quantification functions persuasively in the argument.

Three Objects of Study

I mentioned earlier that rhetorical analysis exists where judgments are to be made through the uses of language and argument. As such I consider texts about educational reform to be an appropriate space for judgments about changes. I have selected three texts that deal with educational reforms to explore the rhetoric that is involved in these texts. These texts have been written for general audiences that address different aspects of educational reforms. All three of the texts are written for general audiences, and can be purchased in a local bookstore in a section called “Education.” The three texts chosen are:

- *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education* by Diane Ravitch¹⁵
- *No Excuses: Closing the Racial Gap in Learning* by Abigail Thernstrom and Stephan Thernstrom¹⁶

¹⁴ http://en.wikipedia.org/wiki/Statistical_literacy

¹⁵ Diane Ravitch, *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education* (New York: Basic Books, 2010).

- *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills*

Our Children Need—and What We Can Do About It by Tony Wagner¹⁷

I have selected these three texts because of the broad approach to educational reform that are taken in each of the texts, although all three rely on quantification in their arguments in some form.

Ravitch writes about her views on schools of choice and vouchers and how her opinions of these matters have changed during her forty years of educational policy examination. The Thernstroms write about the achievement gap (as measured on standardized tests) between racial groups. Wagner writes about schools that he considers capable of preparing students to engage in a global knowledge economy.

All three texts address diverse educational reform efforts. As such, there will be little cross-comparison throughout this dissertation. I am not writing about the arguments of the individual books or the affordances and constraints of the suggestions and recommendations within each book. I am not trying to evaluate the merits of the proposed reforms. That is not the point of this dissertation. In using these three books, I am writing about their rhetorical use of quantification, or put a little differently, how do the three books use quantification in arguing for the specific reforms that are being advocated.

I recognize that not everyone will have read these books, and possibly will have little interest in doing so. However, as this dissertation concerns rhetorical argumentation, I provide a brief summary of the arguments of each of the three books. This summary will structure the reading of chapters three and four of this dissertation.

¹⁶ Abigail Thernstrom and Stephan Thernstrom, *No Excuses: Closing the Racial Gap in Learning* (New York: Simon & Schuster, 2003).

¹⁷ Tony Wagner, *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills Our Children Need—and What We Can Do About It* (New York: Basic Books, 2010).

Summaries of the Three Books

The Death and Life of the Great American School System

Some might suggest that Ravitch's book was written within an apologetic genre, because in it she reverses her previous stance on testing and charter schools. This book was written after a well-published and lengthy career in educational history, policy, and national educational standards, which reflects on current trends and current conditions while considering the role the author played in championing for some of the causes while decrying others. The argument comes through reflection and desires for clarifying previous statements and standpoints. In the opening chapter, the text positions this apologia through the memory of Ravitch having her office painted. In the course of boxing up items and looking at her archived past, Ravitch asks why she "was lacking confidence in the reforms."¹⁸ She concludes that she had the right to change her mind but continued to focus on why she would change her mind.

The argument of the book then takes the reader through the reasons for her change of stance, including space for Ravitch to make recommendations for the future based on these reasons (which will be discussed at greater length in the next chapter). Ravitch suggests that she had been lured by the promises of "quick fix[es] to intractable problems."¹⁹ There was a hope that came from accountability and freedom of markets in application to education. It would be a way to empower, enable, and elevate. This hope suggested ways to close gaps in achievement. "Testing would shine a spotlight on low-performing schools and choice would create opportunities for poor kids to leave for better schools. All of this seemed to make sense, but there was little empirical evidence."²⁰

¹⁸ Ravitch, *The Death and Life of the Great American School System*, 2.

¹⁹ Ibid., 3.

²⁰ Ibid., 3–4.

Ravitch's text reflects on the hope of choice and accountability in schools. The book narrates the failure of that hope. The title of the book suggests that these reforms have led to the death of the American school system. Ravitch argues that her text explains why schools have failed "most of the reform strategies that school districts, state officials, the Congress, and federal officials are pursuing, that mega-rich foundations are supporting, and that editorial boards are applauding are mistaken."²¹ Thus, the argument begins that educational policies and practices are currently being corrupted through the reform practices of choice and accountability.

A key to understanding the arguments of this book is that Ravitch has reversed her positions about school of choice and school vouchers. Throughout her career as an educational researcher and policy advisor she had been an advocate for change that involves allowing for a US educational system of "school choice" that included use of tax-payer vouchers for private schools and an accountability system of high-stakes testing. *Death and Life* reverses her previous stance.

No Excuses

Thernstrom and Thernstrom are long-time civil rights advocates. They argue in this text that, "racial inequality is America's great unfinished business, the wound that remains unhealed."²² This book argues that although schools are no longer segregated by race in the United States, the disparities of graduation rates and scores on standardized testing between Blacks and Whites continue to be significantly different. The text considers how schooling is "the key to racial equality."²³ Thernstrom and Thernstrom draw from their experiences in civil rights, education, and history to explore and argue that there are really no excuses for the disparities that exist in the

²¹ Ibid., 14.

²² Thernstrom and Thernstrom, *No Excuses*, 1.

²³ Ibid., 2.

United States. The text considers the problems of racial achievement as it relates to *No Child Left Behind* and the National Assessment of Educational Progress (NAEP).

Through this text, the argument follows what role teaching (and qualities of “great teaching”) do for diminishing the racial achievement gap, including looking at cultures of different racial groups—excluding the cultural influence of Whites. The Thernstroms also consider the traditional wisdom in resolving educational achievement gaps, which include sending money, racial isolation, and improving teacher quality among those who do not score as well on standardized testing. The argument concludes with a set of conclusions that highlight the challenges of changing US education and the implications offered by the Thernstroms.

The book has the feeling of a research report. The problem for study and connections to current literature are highlighted immediately at the beginning of the work. The argument moves through a presentation of large-scale data, with data figures, providing the discussion of what is being presented in the figure. The text also presents statements from interviews that support the need for educational change. Of course, there is a sprinkling of quotations by scholars in the field that support what is being argued. The work then argues through a discussion of what the data mean in regards to the issue at hand—the racial achievement gap. In conclusion, the text draws from the data to make recommendations for necessary future change. The purpose of the book becomes one that might “bring an end to that heartbreaking story” of racial inequality in schools.²⁴

The Global Achievement Gap

Imagine sitting on a plane next to a large corporation’s CEO and the topic turns to the educational preparedness of the youth in the United States. The CEO suggests that these youth are not prepared to be workers because they do not have problem-solving and social skills that are

²⁴ Ibid., 8.

necessary to compete in this global market. Wagner begins his text precisely with this narrative.²⁵

This narrative continues to reappear throughout the text, offering a backbone to the structure of the argument. The United States is not preparing the future business leaders and workers—it is not living up to its economic potential. This prompts Wagner to suggest that there are two gaps in education. The first is the well documented and discussed gaps that exist between rich and poor or urban and non-urban. The second is the focus of his text. “The second one is the *global* achievement gap, as I’ve come to call it—the gap between what even our *best* suburban, urban, and rural public schools are teaching and testing versus what *all* students will need to succeed as learners, workers, and citizens in today’s global knowledge economy.”²⁶

Wagner suggests that even schools that score well on standardized tests, which according to Wagner is the current focus of education, do not have the skills necessary to compete and that matter most in this current market. He claims that the current educational system—including the curricula, the teacher education, and assessments—were written during a different age for a different age. The text argues that there are seven qualities that will provide success in the twenty-first century. The argument becomes one similar to a sales pitch, supporting the proposed survival skills and implementing the methods of educational change that are pitched by Wagner, who is the co-director of the Change Leadership Group at Harvard’s Graduate School of Education.

Within the text, Wagner suggests seven skills that schools teach in order to make US students more prepared and marketable in the global economy. He suggests that schools that consider these skills to be in position to help make graduates more prepared for college and the global market. He suggests that in considering these seven skills, educational analysts “can also observe how much better prepared the graduates of these schools truly are for college and

²⁵ Wagner, *The Global Achievement Gap*, 1–3.

²⁶ Ibid., 8, emphasis in the original.

careers.”²⁷ Wagner suggests that the following seven skills are essential in educational change to prepare for future global influence: (1) Critical thinking and problem solving, (2) Collaboration across networks and leading by influence, (3) Agility and adaptability, (4) Initiative and entrepreneurialism, (5) Effective oral and written communication, (6) Accessing and analyzing information, and (7) Curiosity and imagination.²⁸

The text argues for these changes by considering the old methods espoused by US education, including the focus on assessing what school-aged youth know. The exploration of assessing knowledge is later considered as Wagner delves into the current trend to have standardized testing to determine knowledge, both in the implementation of state standards testing and Advanced Placement testing. Wagner suggests that the educational systems in the United States need to be remodeled in ways that support motivating students to do work instead of just memorize. Concluding his argument, Wagner provides three examples of schools that are deploying in some form his seven survival skills as testimony to the effectiveness of these skills.

Education is a field where opinions are developed and judgments made. I could have chosen works that were all written solely by education scholars or by non-education scholars. I purposefully chose to mix both. Wagner taught high school English and was later a principal. Ravitch has been an educational historian and policy analyst for her career. Abigail Thernstrom is a political scientist interested in issues of race in society whereas Stephan Thernstrom is a Harvard history professor. The authors of these texts span differences in interests in education and policy. From a former practicing teacher to a political scientist, the three texts were purposefully chosen to not come from the same field to support the arguments that quantification has become a common rhetorical tool addressing educational issues across diverse domains and fields.

²⁷ Ibid., 252.

²⁸ See *ibid.*, 14–42.

As objects of study for this dissertation, I could have chosen peer-reviewed journal articles or articles written in popularist media like *The New York Times* or the *New Yorker*. Do I consider edited volumes or not? Because of the nature of educational writing, the choices were not limited. Since my study considers a type of quantitative literacy, I chose works that were written for non-educational specialists and especially for those who may not have extensive backgrounds in quantitative reasoning and interpretations. These are “big” books by prominent theorists, and collectively they cover a wide range of frameworks for educational reform. I see choosing works of this nature important if we are to understand how education is portrayed to those not within this field. If these works have an influence on how educational practice and policy are perceived, analysis from these works can help shape/challenge discussions within education discourse about the assumptions and conditions.

I could have chosen books about teacher education, such as the works by Gloria Ladson-Billings or Deborah Britzman. Although interesting to consider how quantification influences teacher education literature, I was more concerned with how quantification was deployed toward the public in efforts to reform educational practices and perceptions. I could have also chosen legal documents to analyze, such as *No Child Left Behind* or *Race to the Top*. Although of importance for how education is practiced, that would limit the appeals to only those who consider educational policy. I have nothing against writing with respect to educational policy; however, the choice of objects for this dissertation was an attempt to engage in the rhetoric for the general public, which will come more to play as I consider the types of rhetoric of comparison in Chapter Four and the uses of generalization in Chapter Three.

All three books contribute to educational portrayals in different ways. Ravitch’s work discusses educational trends within the United States concerning schools of choice and the evidence that is mobilized to change the educational system in favor of or in opposition to choice. The

Thernstroms' text considers closing the racial achievement gap, an issue that is often discussed, especially among policymakers. Wagner, similarly, considers the achievement gap but through the eyes of the deficiencies of the United States in comparison to global education.

I could have chosen to compare the arguments of each of these texts, contrasting what is being said about education by the three authors. That would have shaped this dissertation in a different direction than what I was interested in. I am interested in what arguments are being said in the different books; however, this study focuses specifically on how they use quantification as rhetorical devices of persuasion. This dissertation considers how the texts construct and use quantification. There will be similarities and there will be differences in how the three texts deploy quantification. In writing the chapters that follow, I provide some of these points of similarity and some of the points of difference. In some sections there is much more grouping of the three texts under theoretical commonalities, while others will consider the texts independently. This was done purposefully as the generalizations of the texts are distinct and warrant individual attention.

Outline of This Dissertation

Following this chapter, this dissertation has four analysis chapters and one conclusion chapter. I consider the four analysis chapters as consisting of two parts. The first part is a consideration of how quantification relates to education and some of the theories involved in the acceptance of quantification educational reform rhetoric. The second part is a pair of case-study chapters considering how the three objects of study deploy quantification in their arguments. The first part is more about the theories of quantification and education while the second part considers rhetorical examples from the three texts.

Chapter Two considers the role of quantification in educational rhetoric, particularly as quantification is an analytic tool in the auditing culture that has consumed education. This chapter

considers the concept of goal steering and how education is being steered by outcomes, which are numerical proxies for qualities such as learning and comprehension. I draw from the work of Thomas Popkewitz to consider how education is “goal steered” through predefined standards that are being used against teachers and students as methods of accounting standards have been met in effective and efficient ways.

I then move into a consideration of the three books, by Ravitch, Thernstroms, and Wagner. Chapter Three continues the work of analysis across the three books; however, this chapter considers the three texts through inferences and generalization. The three books make generalized statements toward how to change the educational system in the United States. These generalizations are based on quantification, although the explicit expression of these generalizations does not always include the quantifications or the inferential analysis that is the assumed basis for making such claims. I consider the generalizations in the texts and offer my commentary about the difficulties of generalizing human qualities based on calibrated quantities.

In Chapter Four I consider the rhetorical deployment of descriptive statistics in the three books, considering three different ways that these descriptive statistics are used in the three books. I consider in this chapter how descriptive statistics are used as a rhetorical trope of comparison, a rhetorical trope of transparency, and finally a rhetorical trope of the jeremiad. In this chapter each of the three texts will be considered through the three tropes, providing that the three tropes have space in three different works, by different authors, with different educational reform ends.

Chapter Five moves into a more theoretical look at ways educational rhetoric conflates aspects of quantification based on underpinning assumptions of quantification. In this chapter I explore how quantification and statistics are used as synonyms, suggesting that there are differences between the terms and the associations with these terms. I then consider how educational rhetoric conflates the concepts of descriptive statistics and inferential statistics, using descriptive statistics as

a basis to infer happenings in the future. I begin by consider the problems of converting human qualities into quantities in order to describe the past. I then move to issues of inference, considering how inferring in educational reform texts assumes that qualities of sample can apply beyond the sample to the entire population, how inference assumes that the future can be modeled by the past, how inferences conflate randomness with certainty, and how inferences assume that humans can be measured akin to the measurements of the natural world.

I conclude the dissertation with Chapter Six by returning to the concepts of goal steering in the United States, looking at how the three texts assume that educational reform can be steered through the deployment(s) of quantification as rhetorical tools. In considering this goal steering, I consider what good education is in this age of measurement, drawing on the works of Gert Biesta. I conclude by looking at educational reforms through a critical lens, offering a vision of a type of education that is steered by influences other than outcomes and evaluation.

I structured this dissertation to write to an audience of graduate students who might be taking a first research methods course in education. I have tried to structure this document in a way that approaches the current conditions of quantification in educational research and writing, highlighting the connection of quantification to goal steering in the United States' educational reform. I then suggest two chapters that look at cases of how quantification is used in texts that might be read in as part of a graduate seminar or for general interest. I then offer some critiques of quantification and the theoretical assumptions that are associated with quantification. Throughout this document, there will be times when the statistician is informing the rhetorician in me and other times when the rhetorician is informing the statistician. I speak from both experiences. I conclude, then, by returning to education to ask the question of what is being measured in educational research and is it what we as educational researchers desire.

CHAPTER TWO

AUDITING EDUCATIONAL OUTCOMES: QUANTIFICATION'S RELATIONSHIP TO THE AUDIT CULTURE IN EDUCATION

In many ways, statistical analysis is compellingly attractive to us...in education. It is a magnet for grant money, since policymakers are eager for the kind of apparently objective data that they think they can trust....The path of least resistance is to continue in the quantitative vein, looking around for new issues you can address with these methods. When you are holding a hammer, everything looks like a nail.

- David Labaree¹

Goal Steering: A Product of the Relationship Between Quantification and Outcomes-Based Education

In the United States and in other countries educational achievement is measured through outcomes that empirically measure performances on standardized tests whose items were written before the school year had even started. Outcomes-based education, as currently deployed in the United States, requires some form of learning measurement against a calibrated, predetermined standard. This demonstration is a demonstration of learned skills that were pre-determined, often by a source or institution outside of the classroom or school. A claim in favor of outcomes-based education is that all students, regardless of race or socioeconomic status, should be expected to meet rigorously defined national, state, and district standards, which are demonstrated through performances on standardized testing, tests which have become uniform through the sameness of conditions and tested content. Outcomes-based education supposedly allows for comparison of content mastery and predetermined performances.

How do these demonstrations look in the current educational context? In international comparisons, standardized tests like the Programme for International Student Assessment (PISA) or Trends in International Mathematics and Science Study (TIMSS) are commonly used to assess the demonstrated learning of a country's children and youth, allowing for comparisons across countries.

¹ Labaree, "The Lure of Statistics for Educational Researchers," 21.

Although not required by international law or moral guideline, these tests are positioned as methods for determining learning. This trend is furthered in the United States. The National Assessment of Educational Progress (NAEP) is now required as part of the 2001 legislation *No Child Left Behind*, although the mandate does not use the scores of the NAEP to determine federal funding, simply requiring states who wish for federal funding to administer this exam.²

Of particular interest for this dissertation are the comparisons that are portrayed in the books serving as objects of study as high-stakes, those tests which do have meaningful impact on the lives of the tested. These impacts might be considered in tests that allow for high school graduation, such as the Texas Assessment of Knowledge and Skills (TAKS) or the Massachusetts Comprehensive Assessment System (MCAS) or test that measure growth from year-to-year, such as (considering this dissertation is being written for Michigan State University) the Michigan Educational Assessment Program (MEAP).

However, this also includes testing that determines if a child or youth is proficient in certain content areas, such as mathematics and literacy. I take this concept of proficiency from the requirements of *No Child Left Behind*, which states that each state will have 100 percent of the children and youth (starting with children in third grade) proficient in mathematics and literacy by the year 2014. Toward that goal, states must determine a state-wide standardized examination protocol with accompanying classifications of proficiency. Schools must make adequate progress toward having all students at the proficiency level determined by the state.

² See Thomas S. Popkewitz, "PISA: Numbers, Standardizing Conduct, and the Alchemy of School Subjects," in *Pisa Under Examination*, ed. Miguel A. Pereyra et al., vol. 11, Comparative and International Education (SensePublishers, 2011), 31–46, <http://www.springerlink.com/content/pn502k54w313q435/abstract/>; Sotiria Grek, "Governing by Numbers: The PISA 'effect' in Europe," *Journal of Education Policy* 24, no. 1 (2009): 23–37, doi:10.1080/02680930802412669; Stefan Thomas Hopmann, "No Child, No School, No State Left Behind: Schooling in the Age of Accountability 1," *Journal of Curriculum Studies* 40, no. 4 (2008): 417–456, doi:10.1080/00220270801989818.

In this current educational understanding, the benefits of education are measured by demonstrating learning through scores on standardized tests which suggest the learning of pre-defined state and national standards. Instead of measuring the inputs for the students (the curriculum and teaching methods they receive), educational evaluation has become focused on demonstrated outputs on standardized examinations (the test results). I believe that in evaluating learning and performance, outcomes are better indicators than inputs. However, the problem is that in current educational policy, outcomes have been pre-determined by politicians and policy makers. There are particular curricular consequences when outcomes are pre-determined, especially by people who are not in the classrooms.

A central concept of this chapter is auditing. In my experiences as a statistician, I have come to recognize that the ideas of auditing have become major influences in establishing standards (whether those are quality standards in industry or educational standards of learning) and marks of whether those standards have been adequately met. Thus, I begin by writing about the audit culture and how this culture has become a mainstream consideration in educational reform discourse. I then consider how the audit culture influences education through a specific type of outcomes-based education, an educational evaluation based on predetermined standards and the measurement of completion of those standards. I conclude this chapter by considering an alternative form of outcomes-based evaluation.

Goal Steering through Measurement

I mention goal steering in this chapter because outcomes-based educational evaluation is a technology of goal steering in education. In my reading about goal steering, I consider goal steering to be directing education from afar through (pre)determined standards, where the standard is determined by outsiders before the teaching begins. For me, outcomes-based education is an

instrument of goal steering, where the difference is that the goal steering does not require the use of standardized testing (and as I will argue in this chapter, comparing through quantification), but in this case, standardized testing is used as an instrument of goal steering. Much of this development of my ideas about goal steering have come through the works on educational reform and international educational comparison by Thomas Popkewitz.³

Goal steering, as written by Popkewitz, is an educational reform that accomplishes both centering and decentering, both hoping to stabilize the reforms while ultimately destabilizing education. He suggests that goal steering stipulates what should be covered in education, suggesting the directions that one must go. In this relationship, goals and directions determine *a priori*. The connection between goal steering and the current discourse of outcomes-based education, for me, is that educational effectiveness/achievement is directed by these prior determined goals. Outcomes-based education, which is in part manifested by applications of measuring student learning, provides a technology in which education can be steered by the content that has been previously defined. In these relationships, education is steered toward certain measurable and auditable goals and outcomes. Said differently, the goals of learning are determined by an outside source prior to the learning events, and stated in terms of things that can be quantified. Goal steering is opposed to an approach in which the goals of education are determined by the child and/or the teacher who is doing the learning/teaching.

I contend that evaluating education through quantified outcomes-based educational reform (or policy) is an instrument of goal steering because the educational work and direction is determined beforehand and from outside sources, sources who are not in the classroom and do not

³ Thomas S. Popkewitz, *Educational Knowledge: Changing Relationships Between the State, Civil Society, and the Educational Community* (SUNY Press, 2000); Thomas S. Popkewitz, *Changing Patterns of Power: Social Regulation and Teacher Education Reform* (SUNY Press, 1993); Sverker Lindblad and Thomas S. Popkewitz, *Educational Restructuring: International Perspectives On Traveling Policies* (Charlotte, NC: Information Age Publishing, 2004).

know the child, but instead have some governing or funding role. In the current discourse there is quantification and outcomes-based education. They would not have to occur together; however, I contend that in this current educational reform context, they do. Consider some of the reforms and policy decisions which have shaped the historical relationships of quantification and outcomes-based education. A major component of outcomes-based education has become called standards-based education, where reforms are shaped through what children and youth should know and be able to demonstrate by the conclusion of an academic time period. The portrayed purposes of standards-based reforms is that all children and youth, regardless of economic status, race, or gender should be held to the same standards, namely those that can be carefully measured through standardized examinations.

This use of steering education through tests that provide quantified data is not a new system that only resulted because of the passage of mandates and laws in the twenty-first century. In examining this relationship between data and goal steering, I have become aware of a few moments where the directions of education and learning have been steered through the inclusion of goals determined prior to the learning scene. In part, these examples highlight the reliance on probability models for inference about typical human characteristics. For further consideration of the history of quantification, I suggest Ian Hacking's *The Taming of Chance* and Theodore Porter's *Trust in Numbers*.⁴ I am more concerned about the use of these statistical models to steer teaching and learning in the United States.

In the early 1900s, the concepts of testing and learning became more related, such as the US Army's support of the Army Alpha and Army Beta exams for officer candidacy during World War I. Testing gained further support as a measure of intelligence as the Stanford-Binet IQ tests become

⁴ Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1999); Porter, *Trust in Numbers*.

more recognized as a source of measuring normal (used here in reference to the Gaussian probability model) intelligences. Stephen Jay Gould has provided an excellent summary of this in his *The Mismeasure of Man*.⁵ This results in a type of educational sorting, based on supposed intelligence.

Reiterating Gould's argument, Lynn Fendler and Irfan Muzaffar also suggest

Sorting students by ability is one among several competing purposes of education. Educators who promote sorting often justify those practices on the basis that a bell curve represents the normal distribution of things in the natural world. Social Darwinism and normal distribution patterns have provided justification for norm-referenced standardized tests, grading on a curve, detection of "at risk" students, and the whole concept of an intelligence quotient. Insofar as the normal curve is held to be a model of natural distribution, the normal curve is regarded as an objective basis for *sorting people*. The bell-curve model of distribution has been taken for granted in education because it is generally accepted to be a fact of life.⁶

From Fendler and Muzaffar, I consider the implications of tracking and labeling through carefully calibrating tests to the norm. Thus, through the normal distribution of the natural world, it becomes easier to classify (or in their words--to sort) people, thus steering education through the quantified appeals to the normal distribution.

In presenting this short history of quantification in education, I present evidence from court cases about IQ testing and some of the policies and recommendations about education. This notion of educational reform in the United States might have been recognized when the courts ruled about testing as a tool for placement and tracking, and ultimately as determining the directions of

⁵ Stephen Jay Gould, *The Mismeasure of Man* (W. W. Norton & Company, 1996).

⁶ Lynn Fendler and Irfan Muzaffar, "The History of the Bell Curve: Sorting and Idea of Normal," *Educational Theory* 58, no. 1 (2008): 64 emphasis mine.

education based solely on the achievement score of standardized (and pre-determined) tests. The 1967 District Court ruling on *Hobson v. Hansen* ruled that IQ testing was culturally biased against black students; and the 1971 *Larry P. v. Wilson Riles* ruled that Black children could not be placed in special education classes solely on the score of an IQ test. In these two cases, the courts ruled that IQ testing was biased, and that comparisons of children based on these tests were not sufficient to qualify them for special education services. I mention these two court cases as reforms in how the testing is perceived in educational discourses. I am not stating that these court cases shut down testing as a method of educational evaluation. Quite to the contrary, these cases shifted the focus from that of racial profiling and bias to one of educational evaluation and comparison through the use of quantification and testing. Prior to these court rulings, educational goals, particularly those of minority students, were determined through the scoring on IQ tests. These tests functioned to steer the educational goals and future potential for minority students through the positioning and tracking of these children based on a quantified measure.

At the same time, major policy reforms in the United States that have helped to shape the notion of quantifying outcomes have not been limited to court cases which challenge racial boundaries and biases. A brief listing of reform policies might help to explain the current relationship between quantification and outcomes-based education. In the 1980, the Reagan Administration released *A Nation at Risk* which highlighted the dire situation of US education when compared to foreign countries.⁷ *A Nation at Risk* suggests that the United States cannot function competitively in an international educational market. This set of recommendations is based on “indicators of risk” which establish the potential risk through international numeric comparisons.

⁷ National Commission on Excellence in Education, *A Nation at Risk*, April 1983, <http://www2.ed.gov/pubs/NatAtRisk/risk.html>.

Although *A Nation at Risk* was not a mandate, instead only a recommendation from the Department of Education, the connections to goal steering and the purposes of this chapter should be considered. In the rhetoric of *A Nation at Risk* I suggest that there is call for those in education to consider more notions of standards and how those standards can be determined and proficiency measured before children enter a classroom. The nation was falling behind in the international comparisons of these predefined measures, and these measures were probabilistic in inferring future conditions.

I claim that in the wake of this rhetoric of fear the United States experiences establishing a loss of educational (and possibly economic) prominence. Within the United States there had been furthered connections between measurements of outcomes and using those measurements to steer the content being taught in the classrooms. Within the wake, there were suggestions that there are ways that measurements could be used to standardize what is being learned and ultimately allow for comparisons among demographics to allow for steering those educational practices among those demographics.

Where *A Nation at Risk* was only a recommendation from the Department of Education, during the early 2000s the United States government established policy that would help each child not be left behind in academic achievement; this policy, *The No Child Left behind Act of 2001* (NCLB), establishes connections between federal funding and student achievement particularly through student achievement on state selected standardized tests. In part this act requires all students to be “proficient” by the year 2014; schools must be making adequate yearly progress (AYP) for all students becoming proficient by that year. If schools fail to meet adequate yearly progress, among other things, the schools and districts could lose federal funding. This federal funding is tied directly to student achievement and the auditability and the ability to quantify student learning and achievement.

However, in 2009 as part of the *American Recovery and Reinvestment Act of 2009*, the Department of Education along with Secretary of Education Arne Duncan established reforms that have since been called Race to the Top. Although containing different layers of particular interest to this dissertation, I find that the rewarding of states meeting performance-based standards to be particularly indicative of goal steering. The images of leading a horse with a carrot on a stick might be an appropriate analogy. The horse is being steered in the directions the master wishes because the horse desires the tasty carrot while expecting that the reward will ultimately be given. In exploring educational goal steering, the coveted funding will go those states that go in the directions outlined by the government, which include measuring success through performances. I note that as a part of these performance measures, a key indication is the achievement on standardized tests, which have determined the goals of proficiency *a priori*.

Further consideration might be given to some of the governing standards established by groups like the National Council of Teachers of Mathematics (NCTM) published during the 1990s and early 2000s. These standards are not necessarily specific examples of what to teach in the different content areas, but instead are guidelines trying to improve the quality of instruction in those different content areas. One point of particular interest for me comes from the establishment of assessment standards. Although NCTM's standards suggest that assessment should be more about student learning than about testing, the standard suggest that there are strands of content that can be audited by outside agencies.

I also consider the recent work of the Common Core State Standards Initiative, sponsored by the National Governors Association (NGA) and the Council of Chief State School Offices (CCSSO). Although not required for states to adopt these standards, these standards are directly influencing reconstruction of state testing and measuring of achievement. Although of different facades, these motions of educational reforms remain consistent to the concepts of goal steering

being discussed in this section because of the establishment of national goals, changing the accepted measures in those states and across the nation. These new standards do not change the goal-steering orientation of educational reform; to the contrary the Common Core State Standards support steering education from afar through predefined goals in efforts to standardize the learning across contexts and demographics. These standards have created a space for repetitious learning and quantified measures of that learning.

In this section I have claimed that outcomes-based education is a technology of goal steering particularly in the educational system of the United States. Read differently, I consider measuring performance of education on predetermined standardized tests to be a way that funding agencies, government organizations, and curriculum developers choose to direct the affairs of educating children and youth. The evidence for this claim lies partly in the legal court rulings and educational policy recommendations and mandates that have been a part of educational reform in the United States from the 1950s. In writing about outcomes-based education as a deployment of goal steering, I consider the works of Thomas Popkewitz. He suggests that educational reform is “a story of fluctuations and uneven movements, and unpredictable transformations as political rationalities are brought into the pedagogical discourses through multiple capillaries, capillaries that traverse distinctions between state and civil society.”⁸

Arguments about educational reform have come to rest on the assumption that quantification works to describe the conditions of education. For example, we are more interested in the percentages of children with “free and reduced lunch” than in the quality of their food or health of the children who eat those school meals. We are more interested in the percentage of students who are proficient on state testing as opposed to the means that teachers and children use to become proficient. Education is being steered by quantified results in meeting standards that have

⁸ Popkewitz, *Educational Knowledge*, 174.

been established prior to the test through institutions outside the classroom and the learning. Below I discuss how this quantification in education has come as an association of contemporary audit culture.

Measurement & Calibration

I have spent some time discussing the history of measuring human qualities, particularly as it relates to classifying humans through those measurements. I also addressed how those measurements are used to direct educational practices, through Popkewitz's ideas of goal steering. I have made these arguments under the assumptions that two key concepts are clear—the concepts of measurement and calibration. I do not consider these concepts to be fixed in the minds of my audience, and neither are they fixed in my mind. Here I provide some outline of how I understand these terms.

When I teach measurement to my pre-service elementary teachers, I suggest that measurement takes what is uncountable and makes it countable. I might suggest another way to think of this: taking that which is not quantified and making it quantified, thus allowing for comparison and application within probability models. Consider a heap of flour that will be used to make a bread loaf. The flour has no numerical value and thus the standardization necessary in the recipe does not exist. However, through the process of measurement, the heap of flour (which includes qualities of the flour and the qualities of the heap) are calibrated using standard weights and measures. Flour is quantified into countable, mathematical (and standardizable) units, such as the weight of the heap of flour or the volume it occupies.

Calibration is necessary in order to compare between measurements of different things or different times. Calibrations have been historically constructed to provide a common measurement against standard(s) determined *a priori*. Calibration allows for standardization across contexts and

situations. Calibration establishes a conventional standard by which comparisons might be made regardless of political, economic, or religious contexts. For example, the history of measuring lengths contains the tale of how each kingdom used a different metric for measuring distances, the king's foot. In this variously calibrated form, the measurements of distance and length were not comparable across political boundaries. Calibration makes comparison possible.

Moreover, quantification is necessary for calibration. Quantification is an act of counting or measuring, but in order to count, we must have units. Units are determined by calibration standards. Calibration standards then shape human observation and experimentation in order to facilitate comparisons across times and places. Put another way, quantification is a way to map experiences into numbers and sets that convert experiences into units of comparison. In this way, quantification has played a significant role in our ability to accept claims from argumentation. I provide next a brief summary of the social construct of quantification.

Human Qualities and Quantities

A key issue in this dissertation is the assumption inherent in measuring and calibrating human qualities, that is, the translation of qualities into quantities. I see a difference in suggesting a standardized method for measuring flour for baking and suggesting that various qualities of all humans can be measured in a calibrated way. It is in my readings of the history of quantification and statistics that I hesitate at the generalized acceptance of the potential for quantifying human qualities. Although much has been written about this subject, I briefly mention the work of Adolphe Quetelet, as summarized by Ian Hacking, and the construction of the average (or typical) man that is accomplished through measurement.

Hacking suggests that Quetelet adapted the ideas of astronomy to the working of quantifying human qualities. Hacking suggests that Quetelet applied the Normal (or Gaussian) curve that had

been used to measure phenomena like the position of stars (which are occupied in real space) or the infinite flips of a coin to the human world. But then Quetelet took another step that transformed quantification “[Quetelet] applied the same curve to biological and social phenomena where the mean is not a real quantity at all, or rather: *he transformed the mean into a real quantity.*”⁹ Hacking suggests that in Quetelet’s initial work, that the average was to describe the characteristics of subgroups of people (possibly by demographic distinction), which was an objective method for labeling and classifying these subgroups.

But, according to Hacking, it is the later works of Quetelet that changed the notions of measuring qualities. “He transformed the theory of measuring unknown physical quantities, with a *definite probable error*, into the theory of measuring ideal or abstract properties of a population.”¹⁰ This change from measuring with a definite probable error to measuring ideals will be discussed further in the next chapters. However, the conflation of average with the ideal is a crucial distinction in the history of measuring and calibrating human qualities. This second step transitioned mechanisms of chance into a way to establish what is ideal. I find this ethically disconcerting; by calibrated measurements, human qualities became quantified into classifying what is ideal. Thus through standardizing and measuring, the ideal can be determined and classified in such a way to suggest comparison and analysis.

Qualia

I believe that the philosophical concept of *qualia* might aid in considering the ethical difficulties that I have with quantifying human qualities. Quality shares its root with *qualia*. Although the debates about *qualia*’s existence are rich, I suggest that there is merit to consider *qualia* in the

⁹ Hacking, *The Taming of Chance*, 107.

¹⁰ Ibid., 108 emphasis mine.

context of measuring human qualities. In part, qualia depend on individual experiences of phenomena. The Latin *quale* (the singular form of *qualia*) refers to the kind or sort. *Quale* relies on the contexts and histories of the individual, allowing for the interpretation of events, objects, or situations through those personal, subjective lenses. Thus the qualities of humans are understood through the personal subjectivities of the observer. The difficulty comes when asking how does one calibrate these subjectivities in such a way as to make a measurement. Philosophically, the definition of qualia is that it cannot be measured. If measuring the “sort-ness” or “kind-ness” of an object is not possible, then what are the implications of quantification for the ways we think about what it means to be human?

In establishing these standardized measures, the influence of quantification can be felt in working with humans, and for the purposes of this dissertation working with humans in educational settings. In standardizing there must be some levels of accountability, determining if the standards are being met, and being met efficiently. This has given rise of the audit culture in education. I conclude this chapter with considerations of the audit culture and the influence of the audit culture on education and educational reforms, such as goal steering.

We Live in an Audit Culture

We live in an audit culture, a cultural and contextual manifestation that establishes within the culture the ability to establish standards and measure success in terms of meeting those standards by means of instruments of quantification and calibration. The word audit comes from the Latin for “to hear,” suggesting that there must be some evidence heard and judgment passed about what is owed or what is missing in the practice. But now auditing is not just hearing evidence; now auditing is conducted with reference to predetermined standards.

Consider the role of the audit in the manufacturing profession which randomly selects items being made and tests them for standardized quality. An independent group determines whether the standards of the manufactured item are meeting the requirements outlined by a governing body. We might also consider the role of the audit in the medical world, where patients' cases are systematically reviewed and the care provided is examined to determine if diagnosis or treatment is meeting pre-established standards. The audit culture has become one where explicit standards are stated (although I should stress that often the one being audited does not determine what those standards are, and sometimes the auditor has not determined what the standards are, either. Sometimes the institution or structure determines what counts.) with accountability for how one achieves those goals. The auditor then comes in and determines (sometimes independently of the institution) if these criteria are being met.

I find the concepts of auditing to be helpful in understanding how quantification and goal steering relate to current educational practices in the United States. I recognize that there are different ways that one might audit. In using quantification as a part of the audit, there is a (false) sense of efficiency and objectivity. In this mentality, quantification is seen as a quick source of information that can be used to determine what is being taught and how well. There are entire fields of educational study devoted to developing tests that "accurately" measure performance while allowing for those who determine what should be learned to quickly access data and determine how well certain demographic or regional groups of children and youth are doing. In this ability, the proficiency of the learning can efficiently be determined and audited. However, there is also a sense of trust that comes from quantified audits. Audits require some form of agreed standard(s), which can be verified through reliable consistency. Theodore Porter suggests

Objectivity...was a mechanism to exclude judgment. It could be "defined to mean simply the *consensus* among a given group of observers or measurers," and hence measured

(inversely) as a statistical variance. That is, if several accountants give nearly uniform figures for book value according to one measurement scheme, and rather diverse one according to another, the first is by definition more objective, whether or not it seems plausible. The importance of this kind of objectivity was not so overwhelming as to exclude consideration of “reliability,” meaning accuracy. But it could not be neglected, for without consensus there could be no reliability either.¹¹

Although speaking about accountants, I include Porter’s comment as it relates to the trust that comes through auditing in an auditing culture. Porter suggests that there is a consensus in what should be measured and how it should be measured. The community, such as a community concerned with directing educational outcomes, determines what measures are important and how to reliably and consistently determine what is of value, without the clouding of human judgment. Quantification allows for the appearance of clarity and objective judgment.

With this background, I turn to the ideas that auditing relates to goal steering. To recall, goal steering is determining the direction of education through predetermined goals, often by groups outside of the classroom or school. Auditing allows for those groups to consider how well the goals are being met and which groups are meeting those goals. The purposes of goal steering through standardized outcomes (determined *a priori*) are vetted through the abilities to gather reliable information about learning and achievement. Audits allow for careful consideration if education is moving in the “right” direction (where what is right has been defined by standards prior to the learning). Auditing seems to be a method to determine effectively if the goals are steering correctly. There must be some system of accountability for an audit to work. There must be some type of regulation.

¹¹ Porter, *Trust in Numbers*, 96.

Narrating the history of auditing in contemporary culture, Cris Shore and Susan Wright suggest

In the 1980s and 1990s, “audit” migrated from its original association with financial accounting and entered new domains of working life. We are witnessing an example of what we call “conceptual inflation”Audit has been released from its traditional moorings, blown up in importance and now, like a free-floating signifier, hovers over virtually every field of modern working life. Thus, we now have academic audits, company audits, computer audits, medical audits, teaching audits, management audits, data audits, forensic audits, environmental audits, even stress audits....In this case, as “audit” entered new areas of working life, what have become highlighted from its repertoire of meanings are “public inspection”, “submission to scrutiny”, “rendering visible” and “measures of performance”.¹²

In writing about this audit culture, I think that this term is aptly borrowed from the business world, as is suggested by Michael Power.¹³ In writing with this term, I recognize the desire to connect to the current trends in educational reform to become “grokked,” a term borrowed from Robert Heinlein’s *Stranger in a Strange Land* by Patti Lather in reference to the work done for education by the Bill and Melinda Gates Foundation, by the private and business influences.¹⁴ One of the callings of a business is to produce outcomes that hopefully generate profits, measured through the costs of creating quality products at prices that produce profit. In order to produce quality items, calibrations must exist to allow for measurements against predetermined standards, converting the quality of a product to a comparable quantity. In writing about the audit culture, I recognize that this business

¹² Cris Shore and Susan Wright, “Audit Culture and Anthropology: Neo-Liberalism in British Higher Education,” *The Journal of the Royal Anthropological Institute* 5, no. 4 (1999): 558.

¹³ See Michael Power, “Evaluating the Audit Explosion,” *Law & Policy* 25, no. 3 (2003): 185–202, doi:10.1111/j.1467-9930.2003.00147.x; Michael Power, *The Audit Society: Rituals of Verification*, 2 Sub (Oxford University Press, USA, 1999).

¹⁴ Lather, *Engaging Science Policy*.

model is deeply founded in the current contexts of US education. Ravitch, the Thernstroms, and Wagner all write about the increased recognition of business and privatization in public educational reforms. I recognize this deeply embedded notion of education, making education, as Michael Apple suggests, “more business friendly.”¹⁵

Peter Taubman’s book *Teaching by Numbers* has influenced my considerations of this audit culture and its relationship to education. His text suggests that educational progress has marched “under the twin banners of standards and accountability.”¹⁶ These two banners suggest important connections of the auditing culture, quantification, calibration, and goal steering. I will return to these connections in a moment. Taubman discusses three heuristics for understanding these two terms of educational progress. First, he considers Lacan’s concepts of quilting points.

The quilting points of ‘standards’ and ‘accountability’ stabilized what was still in the 1990s a conceptually open and fluid field in education, and recoded politically contentious issues. They made, for example, the ongoing racial problems in preK-12 public schools and the shocking resegregation of schools invisible, by recoding them into standards regulating diversity....What was once an open field has found closure in the unprecedented growth of local, state, and federal standards. Those standards implied a certainty of knowledge and required implementation of practices at all levels and in all aspects of education.¹⁷

Second, Taubman considers the concepts of standards and accountability through the lens of Foucaultian governmentality. Taubman suggests that considering accountability and standards

¹⁵ Michael W. Apple, “Education, Markets, and an Audit Culture,” *Critical Quarterly* 47, no. 1–2 (2005): 15, doi:10.1111/j.0011-1562.2005.00611.x.

¹⁶ Peter Maas Taubman, *Teaching By Numbers: Deconstructing the Discourse of Standards and Accountability in Education* (Routledge, 2009), 106.

¹⁷ Ibid., 106–107.

through governmentality constructs teachers as “dysfunctional, in need of intervention.”¹⁸ In this practice, “individuals are persuaded in the name of autonomy or empowerment to adopt surveillance and normalizing practices that gradually shape their own thinking and conduct to conform to a reality that is presupposed but that comes into effect as a result of these practices.”¹⁹ The final heuristic Taubman considers is that of the audit culture, as defined through British anthropologists. “Audit culture refers to the emergence of systems of regulation in which questions of quality are subordinate to logics of management and in which audit serves as a form of meta-regulation whereby the focus is on control of control. Institutions become auditable by abstracting performance objectives and focusing on the managing system for defining and monitoring performance.”²⁰

These heuristics about the audit culture allow me to consider the current rhetorical moment within educational reform texts in the United States. The first heuristic informs my thinking about how standards and accountability have become major forces in fabricating educational texts, particularly how the current rhetoric is closing potential for diverse discussion, valuing the same types of evidence and argumentation. The second heuristic positions the rhetoric of educational reform in light of power relationships that exist in this rhetorical context, shaping education as in need of some external salvation. In part, the saving grace comes from the educational reform being argued. Finally, the heuristic of an audit culture allows me to consider the ways that meta-regulating systems are influencing how education is steered by establishing predefined standards to meet. For me, these heuristics about the audit culture allow me to think in particular ways that have been established on the human relationships within education and not simply the numbers being audited.

¹⁸ Ibid., 107.

¹⁹ Ibid.

²⁰ Ibid., 108.

I spend the time writing about these three heuristics as they all help in illuminating the roles of audit culture in education and its particular relationship with quantification. I recognize, as does Taubman, the difficulties in understanding the ideas of accountability and standards through these three heuristics. However, we might consider all three as modes of regulating—regulating the curriculum, regulating the interactions within the classrooms, regulating the relationships between schools and government, regulating what counts as evidence of learning. In a way, these three heuristics offer what is acceptable evidence to support how education is directed from afar.

Education is being regulated, and Taubman's heuristics of the audit culture help us to understand the mechanisms of regulation that are in use. In the attempts to regulate the classrooms, there have been standards put in place, which are measured through systems of accountability. In the current educational discourse, that accountability is measured through predefined goals. Quantification has become a common tool of accounting for the success or failure to meet those standards. Test results are calculated through performance-based outcomes on standardized tests, which are analyzed and compared through quantifications. In the next chapter I will consider how these comparisons are made through descriptive statistics and how those comparisons are then converted to forecast future achievements. Both descriptive and inferential statistics have become common rhetorical moves in explaining/exploiting how standards are being met and who is accountable for what is lacking.

I contend in this chapter that auditing is now based on calibration, which requires measurement through quantification. Audits are designed to verify that certain standards and requirements are being met, whether those standards are industrial safety standards or standards of understanding mathematical operations; within the audit, however, is the assumption that there is a standard way to measure whether the standards are being met. Calibration plays a role in this audit

culture, particularly in education where the audits are quantifying qualities, particularly qualities of education and understanding.

It is tempting to say that two schools can be compared, especially when certain demographics are the similar between the two schools, such as the schools having the same proportion of students who are one race or students who have similar socio-economic status or are proportionate female-to-male or that they draw from the same neighborhoods. However, the calibrated measures used to compare these schools disregard the importance of context and circumstances. In describing “pockets of superb education” the Thernstroms provide a case from the Knowledge is Power Program (KIPP) school system.²¹ The text specifically states that the KIPP schools “draw from the same local population.”²² This drawing from the same local population supposedly allows for the comparison of the KIPP schools’ scores on tests to those that share the local population, such as the example of New York’s District 7 and how “only 9 percent in math and 16 percent in reading” are scoring above grade level in that particular district.²³

The Thernstroms describe two urban schools, which are supposedly similar in demographic characteristics, being compared. I suggest that there is a way to imagine in which these two schools are not comparable. In order to compare the two schools, it is necessary to take the perspective of an abstract calibrated standard that can be applied to all circumstances, a metric that strips the individuality of the schools and the contexts. In contrast, we could instead choose to focus on the differences in contexts and historical influences that would make it impossible to compare the two through an abstracted common metric. Focusing on uniqueness and differences, we would be unable to abstract the qualities of these two schools to force a comparison between them. Why

²¹ Thernstrom and Thernstrom, *No Excuses*, 45.

²² Ibid.

²³ Ibid., 50.

might I make such a claim? Yes, the schools may have similar proportions of demographics, like race, gender, and socio-economic status. The schools may also have similar textbooks. The schools may even have teachers who attended similar teacher education and professional development programs. However, the school is made up of unique people. People are not and cannot be the same unless we first construct a calibrated norm and convert an average into an ideal. By doing so, these measurements help to fortify demographic classifications; at the same time, the measurements refuse to acknowledge that people are irreducible to abstract categories. In this way, quantification is an ethical issue.

Not only does quantification allow for comparisons across schools but also within schools based on predetermined demographic characteristics. Educational literature is now concerned with achievement (which is synonymous with measured outcomes) of some demographic subgroup and determine, statistically, if there are differences between these groups, based on their performance on a standardized measurement tool. It seems that there are no bounds to classifying and comparing these subgroups, although the most common examples of comparison come from comparing achievement across gender or racial lines. One of the premises, and often declared benefits, of quantification is the ability to compare unlike objects. Quantification allows for a translation of qualities across contexts and situations to a unit of measurement, which has been calibrated in its measurement. For example, the history of measuring lengths is founded on this premise of calibration of measurements. The nations of Europe based the concept of a foot based on the ruling king's foot. However, in order to determine lengths, a standardized notion of length was calibrated to allow for comparisons across political, social, and environmental contexts.

Not only does quantification allow for comparisons across schools but also within schools based on predetermined demographic characteristics. Educational literature is now concerned with measured outcomes of some demographic subgroups, and they determine statistically if there are

differences among these groups based on their performances on a standardized examinations. It seems that there are no bounds to classifying and comparing these subgroups, although the most common examples of comparison come from comparing achievement across gender or racial lines. One of the premises, and often declared benefits, of quantification is the ability to compare unlike objects.

In education, this audit and accountability structure might be seen in the uses of student evaluations, random classroom visits by administrators (such as in Wagner's *The Global Achievement Gap*), the visits of student teachers and interns by an instructor who measures qualities of teaching, and the inclusion of national and state standards (such as generalized in Ravitch's *The Death and Life of the Great American School System*). This auditing culture in education relates to the practices in the corporate world that tends to regulate the production and trade of goods through the use of standardization—auditing allows for those standardized to be evaluated and exploit conditions where those standards are not met. In the auditing culture, there is a desire to produce numerical comparisons, such that even things like Likert scales are used to demonstrate when a teacher or student exceeds expectation, allowing for the audit to be conducted through quantitative reasoning and comparisons.

Educational reforms' focus has come to be expressed in terms of measurable outcomes and standardizing those measurements in ways that suggest growth or learning. In measuring these outcomes for comparisons, quantification has become a key player. Measuring and quantification share association. Quantification, as defined in the *Oxford English Dictionary*, is the process of measuring units.²⁴ Education, through mixing of auditing and learning sciences, is concerned with measuring the outcomes of students instead of evaluating learning in different forms, but this

²⁴ Oxford English Dictionary. quantify, v. Third edition, December 2007; online version June 2012. <<http://www.oed.com.proxy2.cl.msu.edu/view/Entry/155919>>; accessed 25 June 2012. An entry for this word was first included in *New English Dictionary*, 1902.

evaluation of outcomes occurs during a historical time when quantification has come to be recognized as a voice of comparison, generalizing, and prediction; I am not suggesting that there are not those who are concerned themselves with other forms of evaluating learning. However, the rhetoric of educational evaluation has become steeped in what can be measured (after carefully constructing a valid and generalizable measure), where that measurement concerns how well children and youth have learned goals that have previously been defined.

On the Epigraph

I conclude this chapter with some thoughts that draw from the epigraph at the beginning of this chapter. Although David Labaree is speaking about educational researchers, I believe that his concepts apply to educational discourse, particularly as this chapter considered the relationships between quantification, goal steering (through standards established *a priori*), and the current audit culture. Labaree suggests that the road of quantification is an easier road to travel (or the one having the least resistance), suggesting that educational rhetoric has become steeped in quantification because of the supposed clarity and appeal toward the general public and the policymakers. Quantification, taken in this light, suggests objectivity and trust.

This chapter explored some of the relationships of trust that are associated and assumed through the use of quantification in educational texts and arguments. I began by considering how the educational system of the United States is being steered by outcomes-based decisions. In the current educational climate, the purposes of education are predefined and standardized by bodies outside of the schools. Within this current evaluation structure, the failures and successes of these standards are measured through outcomes that are standardized, often in the form of a test. Students and teachers are evaluated not on individual merits but on abilities to replicate taught

material on these tests. Quantification plays a role in assessing levels of understanding and allowing for comparisons across demographics and geographic regions.

However, the meeting of these arbitrary standards are established through assumptions that human qualities can be assessed, or said differently, that these human qualities can be converted into mathematical quantities through calibrated measurements. The chapter considered, briefly, how the assumptions of human measurement resulted from efforts to take tools of measuring the physical world to measuring human qualities and characteristics. I highlighted the work of Ian Hacking and his consideration of Adolphe Quetelet's contributions of applying the Normal probability model to the interpretations of human characters and how these applications require the translation of human quality in mathematical quantity.

Considering these assumptions about quantification allowed me to consider how the audit culture is related to quantification and goal steering through the concepts of standards and accountability. I explored how three different heuristics of considering standards and accountability ultimately result in control. Education is being regulated and steered by the predefined goals and being held accountable by the outputs on measures of these goals. Why does this matter? The intent of these standards and accountability systems is to control the directions of education through comparisons and inferences. This chapter has served a point consider how quantification has become a standard feature of evaluating education and promoting a system of auditing established standards. I consider some of the conditions of educational quantification in an attempt to help show the differences between descriptions of the past and forecasts of the future.

Quantifications have become a hammer in the rhetoric of educational evaluation. However, this hammer has become a tool that seems to be used to homogenize all aspects of education, allowing for comparison of groups by establishing an ideal while promoting some sense of future forecasting based on the patterns in the past. I do not write this chapter to wag my finger at the use

of statistics and quantification in education. I believe statistics and quantification are necessary in order to understand many things about education. Because statistics and quantification are very useful, it is important to know what quantification can do and what it cannot do. No matter how many times you hit the bolt with a hammer, it only makes indentations.

CHAPTER THREE

RHETORICAL CONSIDERATIONS OF PREDICTION AND GENERALIZATIONS

But it is worth warning that both statistical and inductive inference could be naïvely expressed by the words, ‘we assume the sample is typical of the whole’. Perhaps part of the inclination to try to justify all inductive inference by statistical methods stems from failing to see that two different things may be referred to by those words.

- Ian Hacking¹

This chapter begins providing cases of the deployment of quantification as a trope of generalization and prediction particularly in offering implications and suggestions to education. How have the three texts generalized to the systems of education through their deployment of quantification? I writing the term generalized in this chapter, I recognize that there are different ways that statistics are used to generalize findings; first there is the generalization to the future and second, the generalization from the sample to the populace (whether that is generalized to the entire population or incorrectly ascribing general qualities to the individual within the population). I purposefully use the term generalize throughout this chapter as it related to both predicting future conditions based on the past descriptions and the use of generalizing to the population because the rhetoric of quantified generalization is deployed in both ways. This chapter considers the three texts and the role of implications in their arguments. Here educational consumers, such as policymakers, are provided with the golden ring after reading the book and swimming through the evidence. The texts provide answers to the questions of “so what?” or “what now?”

Throughout this dissertation I draw from the ideas of Ian Hacking who writes how Quetelet offered two important conceptual changes in the taming of chance. The first was suggesting that people were measurable and that within subgroups of people, there are and were characteristics or attributes (qualities) that could be measured (quantified through careful calibration). In this chapter I consider some of the educational implications for his second concept. This concept deals with the

¹ Ian Hacking, *Logic of Statistical Inference* (London: Cambridge University Press, 1976), 126.

shift from gathered measurements and applying to entire subgroups to the ideas of measuring ideals.

Hacking writes

It was Quetelet's less-noticed next step, of 1844, that counted far more than the average man. He transformed the theory of measuring unknown physical quantities, with a *definite probable error*, into the theory of measuring ideal or abstract properties of a population.

Because these could be subjected to the same formal techniques they became real quantities.

This is a crucial step in the taming of chance. It began to turn statistical laws that were merely descriptive of large-scale regularities into *laws of nature and society* that dealt in underlying truths and causes.²

There was a historical shift, according to Hacking, from inferring about natural quantities (which for Quetelet was astronomical measurements) that recognized and announced the chances for probable error into one where the probable error was lost in attempts to generalize to the entire population of interest. According to Hacking's interpretations of Quetelet's writings, there was an important shift of quantifying human qualities through calibrated measurements. In so doing, this shift to measuring qualities promoted changes in the application of these measures to all humans. In educational texts, the shift would be considered as apply to all students or all schools of a certain demographic or dynamic, etc. This chapter considers this generalization of qualities by considering how educational texts deploy this rhetoric of generalization supported through quantification.

The recognition of probable error distinguished these measurements from absolute certainty. In probable error, the qualities of these subgroups that were quantified were by no means certain; there was space for different findings, within certain probabilities. Generalization was not certain to happen but was probable. There was a shift in accepting that through precise, calibrated

² Hacking, *The Taming of Chance*, 108, emphasis mine.

measurements one could obtain the ideal. These ideals, which I contend are qualities, then become measurable quantities, allowing for comparisons to those ideals.

In quantifying qualitative characteristics of populations, quantifying to describe regularities, became “laws of nature and society.” In a way, these laws became generalized to the general populace (or put in a statistical way—population of interest). With this natural or societal law in place, there is potential to predict future events and make recommendations based on these predictions. Such recommendations are precisely that: predictions based on probability. Not predestined or predetermined outcomes. It is tempting to conflate these implications as certainty. Statistics were never meant to be certainties of things yet to come, instead probabilistic considerations. The discourse has made a monstrous leap to making predictions based on quantitative descriptions. The statistician in me wishes that all would recognize that predictions are viewed as stochastic rather than deterministic. However, this is not the view of many who research within education. I am not suggesting in this chapter that statistics is irrelevant in making predictions. It is in the improper use of these statistics within some research writings that diminish the random nature of generalization and prediction.

I should restate that I am not writing to demonize the use of quantification. The use of inference is not a problem, especially if one recognizes that the predictions are of themselves probabilities not certainties. In writing this chapter, I struggle with converting qualities that demonstrate some regularities into laws of nature, that is generalizing the conditions of the past onto the population of the present or the future. We might consider an example developed in *No Excuses*. The text argues that there is a past regularity that Black and Hispanic students graduate high school with an eighth grade-level education, whereas Whites and Asian students are at or above a twelfth-grade level. The regularity of Black and Hispanic students performing below their White and Asian

colleagues, has become in the text a law, suggesting that this is how conditions are presently in U.S. education.

A second problem associated with probabilistic predictions and generalizations is that it is tempting to take a generalized claim and apply it to the individual. One of the members of my committee suggested that an analogy of applying Newton's Laws of Motions improperly or misunderstanding the works of a philosopher and misusing the ideas. In these analogies I find a theme of misusing the tool. I think that a physicist would take cause at research that did not use correctly the notions of physical movement, without considerable theoretical and empirical work. The difficulty in this analogy within the rhetorical deployment of quantifications is that those are using quantifications in education are given introductory courses in quantification, without the necessary background in probability to understand that generalizations and predictions are not certainties but are findings available from the random nature. I am not suggesting that research inappropriately applies the general findings to the individual, but there is rhetorical use in educational writings that diminish the random in favor of making stronger claims about individuals or groups. The problem is not in the use of statistics; the problem is in the misuse.

I think that generalization forms a nice label (or way to sort) that can be applied to humans to create a way for comparisons, often in an attempt at gaining understanding about human behavior. For some reason labels continue to be deployed in discussions and thinking. I wonder if this is in part a result of our desires to be "better" than others, although the conditions of better require the abstraction of human qualities into some calibrated measure. The difficulty with labeling is that humans are unique, not able to be reduced simply to one classification or another. Humans enjoy individual qualities, individual agency, and individual thoughts that separate one from another, allowing for individual presence and contribution.

In generalizations of this sort we close the opportunity for individuals to come into a unique presence, a presence that only can be filled by that individual. In writing about coming to presence, Gert Biesta suggests that uniqueness can only come as we speak outside of communities that are bound together by commonalities.³ In existing outside of those commonalities the voice of the individual becomes a voice that only she can have, only offering insights from her experiences, learning and contexts. It is in these opportunities outside of common communities that the person's representative voice, the voice that could be heard by anyone that shares commonalities, that the person becomes unique. I mention Biesta's comments on uniqueness in light of generalization because generalizing assigns voices to those who have some form of commonality. If misunderstood, generalization, particularly statistical generalization, closes the potential for uniqueness, creating in its stead the representative voices of the common aggregate.

I can see how it would be tempting to consider a general trend as applicable to individuals who share some quality/qualities. I provide an extreme example that might be considered from reading *No Excuses*. The text describes how Black and Hispanic children do not score as well on standardized tests as their White and Asian counterparts. For someone who is not familiar with the probabilistic nature of statistics might apply the information presented in *No Excuses* to a personal level thinking that a neighborhood child who is Black or Hispanic would not be able to score as highly as a similarly aged White or Asian child also in the neighborhood.⁴ It might be tempting for a parent to read Wagner's work and consider the schools described in the last chapter as schools that

³ Gert J. J. Biesta, *Good Education in an Age of Measurement: Ethics, Politics, Democracy* (Boulder, CO: Paradigm Publishing, 2010), 80–90.

⁴ This statement could also be considered through different domains, such as those who are not familiar with the writings of Aristotle might misconstrue his rhetorical appeals in the arguments of *No Excuses*. Or we might suggest that those who are not familiar with racial battles of the nineteenth and twentieth centuries might misapply the works of Malcom X. This is part of the rhetoric associated with education—the appeals to what the masses want, which for the purposes of this dissertation, is the appeal to quantified data.

would benefit their child based on the success stories offered in the text, suggesting that they should move schools to something that is described as working. It might be tempting to consider the words of Ravitch and suggest that all charter schools are not fulfilling the purpose in educating children and youth.

My advisor termed this as a type of statistical *illiteracy*. This term represents a nice summarizing of this chapter, a look at how the problems associated with conflating probability and statistics for certainties in predictions and generalizing from sampled data. In my general statistics courses, a common theme I addressed is that an introductory course provides foundations for quantitative literacies. However, even after completing a course, there is a potential mindset of misunderstanding the random (and stochastic) nature of statistical inference. The general public, I believe, have a better literacy of descriptive quantification, which will be discussed in Chapter Four,—as they are exposed to descriptives earlier in their education and have a sense of what summaries are telling, but the general public exhibits illiteracy when it comes to excepting and considering the randomness involved in generalizing through quantification.

I contend in this chapter that the rhetorical nature of educational research has in many ways forgotten that generalization and prediction are stochastic (coming from the Greek for *to aim* or *to guess* meaning non-deterministic or sporadic). Educational reports offer suggestions based on the findings, generalizing to individual children and youth and the population as a whole; educational reform rhetoric offer predictions of what are to come if changes are not enacted. Further examples of this regarding the three texts will come throughout this dissertation; however, we might consider examples of policy researcher Eric A. Hanushek or the works of the Gay, Lesbian, and Straight Education Network's Safety in Schools reports for examples of this type of generalization.⁵ It is in

⁵ Eric A. Hanushek, "The Impact of Differential Expenditures on School Performance," *Educational Researcher* 18, no. 4 (May 1, 1989): 45–62, doi:10.3102/0013189X018004045; Eric A. Hanushek,

the forgetting of the stochastic natures of prediction and generalization that promotes within the discourse of quantified outcomes-based education a notion that applies quantified finding to all within the educational system.

This chapter considers how the rhetoric of prediction and generalization are deployed in *The Death and Life of the Great American School System*, *No Excuses*, and *The Global Achievement Gap*. This chapter is guided by the work of inferential statistics, which are the foundations for quantitatively arguing for prediction and generalization. As such, I comment briefly on statistical inference and the lack of appearance in the three texts before I move to consider more the notions of prediction and generalizability.

Inferences

In the traditional sense, inferential statistics come about through having a defined population of interest and selecting sample (theoretically a representative sample) from that population. Representative sampling suggests that the aggregate individual qualities are represented by those who participate. The sample is then quantified in an attempt to estimate the population as a whole. Thus measures of the sample are summarized in some form to a statistic, such as when a mean value is computed for the members of the sample (the sample mean serving as an estimate of the population mean). Probability models are then applied to the sample statistic to determine a probable estimate of how close the sample statistic is to the corresponding population summary, called a parameter. This estimation is built around sampling error which provides a measure of spread around the

“The Economics of Schooling: Production and Efficiency in Public Schools,” *Journal of Economic Literature* 24, no. 3 (September 1, 1986): 1141–1177, doi:10.2307/2725865; Gay, Lesbian and Straight Education Network, *2009 Nation School Climate Survey* (New York: GLSEN, 2010), www.glsen.org/research; “2011 National School Climate Survey: LGBT Youth Face Pervasive, But Decreasing Levels of Harassment | GLSEN: Gay, Lesbian and Straight Education Network,” accessed April 11, 2013, <http://www.glsen.org/cgi-bin/iowa/all/news/record/2897.html>.

statistic, suggesting a probable range of acceptable values for the populations parameter, often called by frequentist statisticians a confidence interval.

When originally considering the structure of this dissertation, I had hoped to write a chapter about the use of inferential statistics in the three texts. However, the three texts do not perform traditional inferential work, at least they do not report in terms of inferential statistics. I believe in part that this is a rhetorical move from the perspectives of the authors to the suggested generalist audience, who may misinterpret the findings of statistical tests or who may not choose to read the work in order to avoid wading through statistical jargon and application. This omission could also come as a result of the quoted data sources used in the texts' arguments. As I mentioned earlier in this dissertation, the authors did not conduct their own research studies for evidence; they drew from other publications, whether those publications were in research journals or newspapers or government reports. The lack of inferential work in the three books could also be an indication of rhetorical trust that the evidence provided comes from studies that have gone through the rigors of the statistical analysis. By trusting the research reports, the authors of the books were able to inform educational practice and policy by translating the research into more accessible terms. The Thernstroms for example did provide their own descriptive statistics work, whereas Ravitch and Wagner provide personal experiences while relying on the descriptive data of others.

The authors of these texts made rhetorical decisions to omit the details of inferential work from their texts. The decision to omit inferential statistical work does not mean that the opinions of the text are not influenced by inferences. The footnotes of the texts list some articles published in academic journals with a mixture of sources that are taken from popular mainstream media. I consider this an important rhetorical move from the authors of these texts, suggesting some hesitancy in supporting argumentation for general audiences with complex statistical inferences and tests. The authors assume that the generalist audience will not understand these statistical

computations and figures, choosing to err on the side of caution in creating an argument that does not require the prominent display of inferential statistical findings, with its inherent p-values, confidence intervals, and test statistics. The arguments may allude to or cite such findings, but that would require the reader to take upon her own initiative to find the source, read the cited authors text, and interpret the results. If the reader does not take this initiative, there is no readily available counter argument for the generalizations and predictions being offered in the texts. In these educational reform texts, I wonder if the authors assume that the general readership does not wish to be bogged down in the details of inference, instead offering a summary of inferential findings in support of the educational reforms offered in the texts.

However, as is almost convention in educational reform texts, the authors did provide generalized statements about what must or should be done as well as offering some predictive statements of what will occur if reforms are not enacted. Inferences allow some form of probability statement to be made about the population, whether that statement is a forecast of future events based on similar conditions or the application of findings to the population. However, a common (and statistically naïve) result of this inference is that the qualities that were summarized by the statistic for the population then get generalized to the individuals within the population. Thus generalization, through misapplied inferential statistics, applies measured qualities to the individuals within the aggregate.

This chapter has thus become a consideration of the generalizations that arose from the collection and analysis of quantitative data. To repeat a question I mentioned earlier: How do these three texts generalize to all education through the lens of quantification? Thus, the salient issue in this chapter is how are the descriptive quantifications of the past being used to argue in a more generalized way.

I write this chapter considering each of the three texts individually because the generalizations and predictions of the three texts are unique to the arguments within the texts. I should note that in reading the texts, the authors follow research writing conventions which establish a separate placement for the data and the interpretations of the data, suggesting that the data is presented collectively in the middle “analysis” chapters while the conclusions and implications are reserved for the latter part of the text.⁶ I recognize this to help the reader recognize that the generalizations and predictions of the texts are based on the support of quantification that had been deployed in the data sections of the text, although the generalizations and predictions may not directly refer to the quantifications. In part, the assumption is that before the reader will accept the generalizations to the population or individual or predicting the future educational condition, the reader will have accepted the quantified data as being valid and applicable.

I begin by writing about how predictions are used as a rhetorical tool, drawing from predictions offered by Diane Ravitch and how the Thernstroms engage in predicting the future conditions of Black and Hispanic youth and Children. I then consider the rhetorical use of quantification as a tool for generalizing findings. In writing about the generalizations, I have chosen to look at how *No Excuses* and *The Global Achievement Gap* chose to use generalizations. I note that the purposes of these two rhetorical features vary between the books. In Ravitch’s and Thernstrom and Thernstrom’s texts, the rhetorical use of prediction is one suggesting what the future may hold if changes pronounced in the texts are not enacted. Wagner and the Thernstroms offer generalizations of conditions across educational systems and populations. I believe that Ravitch is writing about what suggestions she would make to change education in the future. Wagner is writing about qualities that are seen that could apply to other groups. The Thernstroms do both.

⁶ I should highlight that this dissertation does not follow such a convention. I will refrain from having a separate analysis chapter in the hopes that the reader will see within each chapter appropriate places where analysis will be made.

Rhetorical Future Predictions: Predicting the Conditions of School Systems

The Death and Life of the Great American School System

Ravitch is writing *The Death and Life of the Great American School System* as a way to publish a major shift from her previously held positions favoring school of choice and school accountability as effective in educational reform. Thus, she concludes her work by offering a set of lessons that she has learned in the past forty years within educational research and policy analysis. Her lessons learned are generalized implications of the future of education if better reforms are not made.

Ravitch offers suggestions about the conditions of schools without really offering generalized predictions of the future. I am unsure if this is because of Ravitch's background in history or as an educational policy advisor, where her role was to offer suggestions without declaring certainties that would come, or if she's a smart and conscientious scholar who is aware of the pitfalls of statistical claims. Her suggestions are listed as ways that schools will or will not improve if changes are not made. For example, she suggests that schools would not improve if there were continual reorganization of schools' structure and purpose. She contends that organizational changes will not provide solutions to the real problems of lack of educational vision and curriculum. In this warning statement, the rhetoric is supported by her experiences listed throughout the text but not a forecasting of what will occur based on the numeric data.

Or we might consider Ravitch's recommendation that the future of education will not improve if the only focus is on mathematics and reading. "Schools that expect nothing more of their students than mastery of basic skills will not produce graduates who are ready for college or the modern workplace."⁷ She advises, based on the data accumulated since the passing of *No Child Left Behind*, that if schools do not encourage a liberal arts education, the past measurements indicate that

⁷ Ravitch, *The Death and Life of the Great American School System*, 226.

the future will not produce those who are able to create (both artistically and scientifically) and will not produce citizens who are capable of making thoughtful decisions.

A final example would be how Ravitch suggests that the schools “cannot improve if charter schools siphon away the most motivated students and their families in the poorest communities from regular schools.”⁸ This is coupled with the claim that schools cannot improve if they continue to be treated as a business. In these two claims Ravitch is forecasting a failure based on the data that was presented about the differences between charter and public schools and the influence of the billionaires in educational funding and achievement. The rhetoric here considers the past test score data to inform a future opinion about the failure if charter schools and businesses continue to infiltrate the schools.

Ravitch’s predictions are hedged in hypothetical language of “if” schools do not make the changes, they will not improve. The rhetoric involved in these forecasts is that there will be no change in how education is perceived if these changes are not met. I notice in Ravitch’s claims that the past data suggest the future will not be better if no changes occur. She states “we have known for many years that we need to improve our schools.”⁹ She is suggesting that the general public has known for years that the school system is currently not working, yet the difficulties arise from not agreeing on what should be done. In making this claim, she suggests that based on the past indicators, the past measures of learning and achievement, that if changes are not made, there will be no improvement.

⁸ Ibid., 227.

⁹ Ibid., 223.

One of the purposes of *No Excuses* is to decry the racial inequalities that exist in the United States, particularly concerning the differences that exist between the minorities, particularly the Black and Hispanic demographics. The generalizations found within the text extend inequalities to those who are Black or Hispanic, although the purpose of the text is concerning the Black-White achievement gap and not the racial achievement gap as a whole. The text generalizes that Black students will be four years behind their White counterparts after high school graduation. They suggest

An employer hiring the typical black high school graduate, or the college that admits the average black student, is choosing a youngster who has only an eighth-grade education. In most subjects, the majority of black students by twelfth grade do not have even a “partial master” of the skills and knowledge that the authoritative National Assessment of Education Progress says are “fundamental for proficient work” at their grade.”¹⁰

This quote opens the conclusion of *No Excuses*, suggesting a rhetorical finalization that there is a difference in performance on standardized tests between racial demographics. The Thernstroms use the term average, to suggest that although not all will be four years behind, there are some who will score higher and others who score lower. However, the generalized Black and Hispanic youth will be scoring at a level that is consistent with eighth-grade Whites and Asians. This generalization could not be made without quantification, particularly if we consider the term average to be a quantified summary. This quantification occurred in the first three chapters of the book, with the analysis of NAEP tests scores.¹¹

¹⁰ Thernstrom and Thernstrom, *No Excuses*, 270.

¹¹ Ibid., 12–14.

These test scores are quantified proxies of intelligence and learning to suggest a bleak future for those youth who are not White or Asian. The Thernstroms offer a future generalization, suggesting that there those Black and Hispanic youth are capable of becoming more, but are limited in by their access and achievement. In the quote that follows, notice how the text conflates probabilistic potential with a type of certainty. The conclusion of the text states:

African Americans today can serve as secretary of state, CEO of a major corporation, president of an Ivy League university, chief surgeon at a major hospital. But their access to positions of power and prestige—and to well-paying jobs in general—*will be* limited if they typically leave high school with an eighth-grade education. Americans with equal skills and knowledge have equal earnings today—whatever their race or ethnicity. But those equal skills and knowledge are the unfinished business of the civil rights revolution of the last forty years.¹²

Earlier the text considered NAEP data to infer that there is a difference between White and Black students when leaving high school. This rhetorical deployment of generalization depends on the trust in the numbers that were selected and deployed throughout the text, although particularly in the first three chapters, to suggest that the future *will be* limited for these Black and Hispanic students.

The Thernstroms argue that there are potentials for growth and achievement but the future prediction based on the quantification is that there will be no change for these youth unless the gap in schooling, which they have declared exists through reporting quantified results is not changed. Thus,

the alternative to a radical overhaul is an appallingly large number of black and Hispanic youngsters continuing to leave high school without the skills and knowledge to do well in

¹² Ibid., 274 emphasis mine.

life; doors closed to too many non-Asian minorities; the perpetuation of ancient inequalities.¹³

The conclusion of this text offers again a generalization of how doors *will be* closed without educational changes and reform. This rhetorical generalization becomes a condition of non-Asian minorities who will continue to be shaped through the assumptions that the future is a repeated past.

In selecting this example from *No Excuses* there is a rhetorical issue of basing the future qualities of educational performance and achievement on the quantified counts of the past. As if offering a suggestion that in the future there will be limited career options for Black and Hispanic students if changes are not made because of the test scores of the past or that the future of urban youth are bleak if these schools are not freed from public school constraints because the past scores indicate that the certain charter schools score better on standardized tests. This type of prediction of the bleakness for Black and Hispanic students is serving as a rhetorical foundation for a Pygmalion Effect pre-labeling and predicting Black and Hispanics as behind their White and Asian counterparts, creating a self-fulfilling prophecy of Blacks and Hispanic youth and children continually being behind.

Critical Considerations of Predicting

Forecasting is one of the difficulties in inferential statistics. Data do not speak for themselves; they are given meaning from the interpretations, biases, and subjectivities. It is in this given meaning that certain arguments are made through the data, potentially telling the story that the author or audience wants to hear. I believe that this is part of the deployment of data in

¹³ Ibid.

argumentation, purposefully pieces, as Maggie MacLure suggests, to fabricate a purposeful story.¹⁴

However, in this piecing and fabrication there has become a trust in the accuracy of forecasting particularly through numerical data.

One of the hallmarks of inferential statistics is found in the concepts of prediction and forecasting. In reading a recent work by Nate Silver, I have a new perspective in my consideration of the differences between these two terms, particularly through deploying them as distinct and not interchangeable terms. Silver suggests that prediction was something prophetic or celestial (such as the statements offered by a soothsayer) whereas forecast comes from the connection that man is the master of his own fate, alluding to Shakespeare's character Cassius from *Julius Caesar*. Silver suggests

The term *forecast* came from English's Germanic roots, unlike *predict* which is from Latin.

Forecasting reflected the new Protestant worldliness rather than the otherworldliness of the Holy Roman Empire. Making a forecast typically implied planning under conditions of uncertainty. It suggested having prudence, wisdom, and industriousness, more like the way we now use the word *foresight*.”¹⁵

Here Silver suggests that the difference between forecast and foresight comes in part from informed conditions amidst uncertain conditions. Thus, when meteorologists make suggestions about the weather conditions of the next day, they are making a forecast based on certain conditions of wisdom and modeling. However, the problem with forecasting is that they are not certainties, instead they are educated and informed thoughts based on certain conditions, particular assumptions about behavior and history, and data constraints.

¹⁴ Maggie MacLure, *Discourse in Educational and Social Research*, 1st ed. (Maidenhead: Open University Press, 2003).

¹⁵ Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail--but Some Don't* (New York: The Penguin Press, 2012), 5.

I am suggesting here that it seems in this data-dense world that those who are able to use data to support predictions are believed. Consider for a moment the work of political pundits and pollsters who through dense data sources are able to forecast (whether non-biased or not remains a large and potentially fruitful question) the outcomes of political races. For example, we might consider the polling during the 2000 presidential election where Al Gore ran against George W. Bush. In forecasting, many pundits suggested that Al Gore would win, however the descriptive counts provided enough electoral votes for George W. Bush. Here, forecasts suggested one result while the historical counts provided for another. The forecast was uncertain.

This work of political pundits is very similar to the work of creating gambling odds and handicapping. Each person establishing the odds has their own models and prediction tables on which they establish their odds and margins. In doing so, the models do not always succeed, such as when a basketball player is injured in the middle of the game. There are those who forecast accurately, through the careful use of data, such as the work done by the blog fivethirtyeight.blogs.nytimes.com (run by *New York Times* analysis Nate Silver). These forecasts correctly understand the limitations of forecasting through data while making informed (and for the most part accurate) declarations.

The forecast for the 2000 presidential election could have come through quantified data in two forms, each serving a different rhetorical purpose. First, the data could have come from public opinion polling, such as the use of Gallop polls to determine the likely winner. This case is making predictions based on the ability to quantify the qualities of support and beliefs. The second type of forecast would come from taking counts of previous elections and predicting based on the past. Both forecast the future political condition. The first uses descriptive quantification of qualities to consider the population's tendencies and trends, while the second looks at the historical influences. The first is a demonstration of the problem of having quantities stand in proxy for qualities and the

second is a problem of assuming the future will be like the past. I will return to these topics more fully in Chapter Five.

Human conditions and human responses are not random and cannot be modeled completely accurately through probability models. Randomness is a difficult condition to meet; the statistician in me recognizes that even “random number generators” are not random, instead being a string of numbers that closely resemble randomness. However, humans are not merely predetermined strings of commands or outcomes. Humans are not mechanical. Humans are capable of creating and capable of acting under their own agency. They are not simply the result of a random occurrence. They are not coins that when flipped land on a side based on the force of the flip, the number of rotations in the air, the weight of one side versus the other. Humans are not mechanized objects programs to respond to conditions without thinking and choosing. Humans are capable of influencing the future in which we live.

I do not throw out this concept of humans not being random carelessly. I recognize that this is a deeply philosophical issue, one that has been brought up through religious discussions and writings of diverse human philosophers. I am content for the purposes of this dissertation to contend that humans are not random and exist in complex power relationships. Foucault suggests “the human subject is placed in relations of production and signification, he is equally placed in power relations which are very complex.”¹⁶ It is in these complex relations that I contend humans respond without randomness, responding based on the contexts, conditions, and historical influences.

¹⁶ Michel Foucault, “The Subject and Power,” in *Michel Foucault: Beyond Structuralism and Hermeneutics*, ed. Hubert Dreyfus and Paul Rabinow, 2nd ed. (Chicago: The University of Chicago Press, 1983), 209, <http://foucault.info/documents/foucault.power.en.html>.

Accuracy, Precision, and Construct Validity

One of the difficulties of forecasting comes through the concepts of accuracy and precision. When I teach this concept to my introductory statistics courses, I have often used the example of an archer shooting at a bull's-eye target, a target with concentric circles of alternating color. Ideally the archer would like to hit the center of the target, successfully and continuously. Accuracy describes the ability of the archer to surround the center of the target while precision refers to the ability of the archer to hit near the same spot consistently with consecutive shots. Thus, an archer who surrounds the center of the target while not “clustering” the shots in the same space would be accurate but not precise. On the other hand, an archer who is able to cluster the shots but not around the center would be precise but not accurate. Ultimately inferential statistics desires to be both accurate and precise, that is estimating occurrences regularly, which may be possible when dealing with machines and mechanical functions of the natural world, but which cannot account for human ingenuity, creativity, innovation, and unpredictability, all human qualities. I am not suggesting here that inferential statistics cannot be used to inform human practices. I take issue with the issues of taking human qualities, which cannot be measured, and calibrating them to a quantified measure.

Some might argue that is the purpose of construct validity, a term which has come to be regarded as measuring what you intended to measure, consistently. This concept comes from Cronbach and Meehl in the 1950s. They consider such validity to be “established by showing that the test items are a sample of a universe in which the investigator is interested. Content validity is ordinarily to be established deductively, by defining a universe of items and sampling systematically within this universe to establish the test.”¹⁷ I read Cronbach and Meehl’s statement to mean that construct validity validates how close a measure is to what is desired to be measured. Put another

¹⁷ Lee J. Cronbach and Paul E. Meehl, “Construct Validity in Psychological Tests,” *Psychological Bulletin* 52 (1955): 282.

way, is the measure measuring what you intend or measuring something different. Yes, it is possible to measure accurately some human characteristics that are desired. The difficulty comes when measuring some quality, like human intelligence or ingenuity.

Cleo Cherryholmes writes about construct validity and different responses from different epistemological camps. I find his treatment of construct validity to engage important issues of how we understand educational research. In speaking about how construct validity might be seen through the postmodern, Cherryholmes suggests that in considering the construct, there must be recognition of the socio and political influences. Drawing from his reading of Foucault, Cherryholmes suggests the following

Mainstream approaches to construct validity, one might argue, are technical choices based on expertise, rationality, and authoritative knowledge behind which ethico-political choices lurk. Implications of Foucault's interpretive analytics for construct validity locate constructs historically and politically. Attempts to validate ethico-political choices are pursued by asking questions such as Which ethico-political choices are hidden? How are they hidden? How can they be illuminated? Decisions about construct validity cannot be disentangled from ethico-political decisions. This is not an argument against using devices such as Campbell and Fiske's multitrait-multimethod matrix, but it suggests that techniques such as these may often obfuscate, confuse, and mislead more than clarify and validate.¹⁸

Cherryholmes suggests that through a postmodern lens decisions of construct validity cannot be disentangled from the ethical and the political. The difficulty, then, with using quantification as a rhetorical tool is not that it is impossible to measure some human characteristics. The difficulty for me and the working of this dissertation is that human qualities cannot be separated from the ethical and political dimensions and thus become convoluted as rhetorical evidence.

¹⁸ Cherryholmes, "Construct Validity and the Discourses of Research," 440.

Returning to the ideals of accuracy and precision, inferential statistics require conditions to be the same. The statistical parlance would be to not infer beyond the population represented by the data, as outlined by specific contexts and conditions. The conditions and assumptions of the data play a part in understanding when and where forecasts might be made. Suppose for the past thirty years a company has made baby cribs at the same location using the same materials. The company has produced say 200,000 cribs during that time and had 2 that failed after purchase. It seems that this company could claim that their cribs are safe; the odds of owning a safe crib seem favorable. A new owner of this company decides to move the company to a different location and to use different materials than had originally been used. It seems that this company would want to market the thirty years of successful building. It would seem almost natural to claim such a safe crib. However, the contexts and conditions have changed. Is it enough to suggest that the new materials would not work just as well or better? No. But the point of this example is that it is unknown. The changes in contexts, the changes of conditions might alter the forecasting abilities of this company. I am not suggesting that this company could not market that the name had been around for thirty years or that they have a history of crib making. The danger of this inference is forecasting beyond the data the new cribs would be as safe or safer than the old cribs, if based on the historical (and descriptive) claims of high safety rates. Yet it seems the current deployment of quantification would allow for such claims to reach and possibly influence public opinion and perception. A counterargument to this statement might be that claims that inform public opinion that are based on say historical research does the same work as the use of quantification as a type of evidence. I contend that the results may be the same, but the paths through which the results are obtained are rhetorically different.

Consider the example of *The Death and Life of the Great American School System*. Ravitch, a trained historian, could have solely argued through historical means to challenge the conditions of

school reform in the United States, suggesting that the reforms for a national curriculum that is fostered by schools of choice, should occur through the historical claims of school conditions. However, in this text, Ravitch does not solely turn to the appeals of educational change through the use of historical arguments; her arguments drew support from the presentation and interpretation of data. The changes in schooling have not worked, or so Ravitch argues, which is demonstrated not by the historical documents and historical appeals, but instead because the numbers become a rhetorical tool of lack of change in calibrated achievement. The rhetorical use of quantification in educational reform texts is a huge leap to move from quality control in factories to predicting the behaviors of humans. I will return to this leap in Chapter Five.

Arguing through Generalizations: Generalizing School Systems

No Excuses

The purpose of the argument in *No Excuses* is that there are ways that racial achievement gaps can be overcome through careful reconstructions of educational systems and principles. The Thernstroms argue through the text that there are changes that would be beneficial in all cases. For example, they contend that urban schools must attract “more smart, articulate, hardworking people, eager to teach the kids who most need academic nurturing.”¹⁹ They suggest that teachers who work with urban youth spend too much time in learning how to teach as opposed to becoming masters of a content area. They suggest, based on unreported data of parochial and private schools, that teachers are capable of doing an excellent job even without taking teacher education courses. They suggest that all schools should consider some type of merit-based pay to attract smarter and stronger teachers who do not just understand how things work for children.

¹⁹ Thernstrom and Thernstrom, *No Excuses*, 251.

A second generalization offered by the text is that the states need stronger academic standards and standards of accountability than are currently being used, drawing this generalization from the scores and abilities of students of Massachusetts. This generalization is that all states must follow the stricter reforms of such states in attempts to tear down the roadblocks to educational success within those states. The Thernstroms argue from the success of one state suggesting that all other states should follow this particular case.

A final example offered by Thernstrom and Thernstrom is the generalization, based on numeric data, of what reforms should occur within schools to aid in closing this racial gap. They suggest, for example, that “*every* urban school *should* be a charter.”²⁰ They base this claim on data that was explored in the earlier part of the book that included glimpses of how well certain charter schools did on standardized tests, again a deployment in this book as a proxy for intelligence and learning. Again, I have emphasized the use of generalization in this text with words “every” and “should.” What would happen in the future if these schools were converted? I do not claim to know. Thernstrom and Thernstrom suggest, however, that the freeing from traditional constraints associated with charter schools would open doors to diminishing the racial achievement gap and creating better spaces for learning for those youth and children who are marginalized in the current system.

The Global Achievement Gap

Wagner writes differently from the other two texts analyzed for this dissertation. He does not rely on quantification as much as the other two texts, drawing from his personal experiences in the classrooms as qualitative evidence instead. I purposefully chose to include Wagner’s text in this dissertation, in part, because of this difference. He does use some form of rhetorical quantification,

²⁰ Ibid., 265 emphasis mine.

as will be described in Chapter Four. He does include comparisons of counts and test scores. He does include information and facts that set a tone of what he is about to discuss and recommend. However, the difficulty in analyzing Wagner's text through rhetorical quantification is that his generalizations are not based on statistical inferences; they are based on small cases, valuing descriptive statistics of successes found within these cases, instead.

For example, he concludes his book with a discussion of three schools (two of which are charters), which in his opinion are meeting the demands of preparing youth for a more global economic future. He does generalize certain traits that are found in these schools, such as how these schools are "learning and assessment focused" instead of memorization and test-preparation focused. These schools were driven by student-motivated projects instead of following a curriculum in that served to prepare for the test. And finally, these schools base accountability on how students perform on real world problems rather than on standardized tests.²¹

He does make some generalizations based on the data throughout the text. In his conclusion, he considers how "all students need new skills to thrive in a global economy," "using new information to solve new problems matters more than recalling old information," and "today's youth are differently motivated when we compare them to previous generations."²² In these statements Wagner generalizes these qualities to youth, suggesting that in this new generation need new skills, new tasks, and new motivations.

As I mentioned in Chapter One, this text seems to be written from the perspective of a sales pitch, offering a new product for the future success of children in the global market. Thus, the generalizations of the text are suggestions of how to incorporate Wagner's seven success strategies into school reform and performance. The skills that he is generalizing support his notions of success

²¹ Wagner, *The Global Achievement Gap*, 258–259.

²² Ibid., 256–257.

found in any possible job in the market. He suggests that “the most successful businesses want to hire as many employees as they can with these skills.”²³

Beyond the Sample?

Rhetorically, all three texts use descriptive quantifications to generalize beyond the sample, offering statements that apply qualities to all, whether that be a generalization to schools, youth, people who have similar racial characteristics or are in the same age demographic, or to curricula. They apply the measures of the past to a general population. However, two of the texts took a leap producing forecasts based on the measures of the past, offering what would be if changes were not made, based on past counts. In this rhetorical deployment of quantification to predict the future based on the past in educational argumentation conflates probabilistic inferences with statements of certainty.

Consider how the rhetoric of *No Excuses* would change if the argument had concluded with something that assumes probabilities instead of certainties, something like: It is most likely that Black youth will graduate from high school at an eighth-grade level. What is different from using the term typically in the quoted passage above? I see that the Thernstroms’ argument invokes the ideas of what Black youth are certainly, as opposed to the concepts of writing about the Black youth as a probability. What is being assumed in this statement? I recognize that the statement is a condition of probability, based on conditions. In this statement there is a chance for a Black youth to graduate at or above the level of her peers. It is also begs to ask when conditions are available for a student to graduate above an eighth-grade level. It assumes that there are individual stories that could be told. However, the rhetoric is not as impactful hedged in the statement of probability. It is easier to gather supporters for change if the problems are generalized to the entire population of interest.

²³ Ibid., 266.

Critical Issues of Generalizing

Generalizability and the natural sciences are connected, and from the natural sciences the social sciences have developed a desire to generalize findings to human populations. In the field of US educational research, scientific research has taken hold particularly through established norms and documents that describe methods of funding and accepted practices. Earlier I mentioned the work of the National Research Council to establish methods of scientific research in education, particularly in building “models or theories that can be tested.”²⁴ One of the results of such scientific studies, according to the report, is “how individual findings generalize to broader populations and settings.”²⁵

Lynn Fendler begins her argument through consideration of federal standards for educational research, found in the US Department of Education’s *What Works Clearinghouse* (WWC), a collection of research studies that are almost solely based on experimental and quasi-experimental design principles. Specific standards are established by the WWC to support what counts as scientific research in education. Fendler suggests that “when research designs meet these standards, they are called ‘scientific’ by the WWC. This is a particular, and historically specific, definition of science.”²⁶

It is in this setting that Fendler argues about the nature of generalizability in educational research and writing. She contends that, “in educational research, generalisation is an example of inductive thinking because it is a process that seeks to find an overall pattern across an array of

²⁴ Committee on Scientific Principles for Education Research, *Scientific Research in Education*, 2.

²⁵ Ibid., 4.

²⁶ Lynn Fendler, “Why Generalisability Is Not Generalisable,” *Journal of Philosophy of Education* 40, no. 4 (November 2006): 437.

specific examples.”²⁷ I interpret this statement in educational rhetoric as a desire to find the general from the particular, seeing particular educational data and outcomes to generalize to the entirety of that population. Fendler considers some of the works of analytic philosophers, such as Hume and Russell, in discussing some of the critiques of induction. Below I consider this idea in more detail as the concepts of induction will structure the analysis of the three books I am studying.

Although there are different interpretations about what is meant by generalizability, the purposes of the What Works Clearinghouse, according to Fendler, is “to provide direction for policy.”²⁸ Educational policy-making looks at quantifications as authoritative in determining what steps should be undertaken and what changes should be made. Education is being steered by the research, which is deeply shaped by rhetorical deployment of quantification. However, in policy considerations, quantification takes on new forms. Fendler offers this insight

Educational policy discourse converts probability to certainty in the process of decision-making.... Generalisation is unquestionably a stochastic process. Therefore, within statistical modelling, there is no basis for trust or certainty in the generalisability of findings;

probability is precisely not certainty.²⁹

I read this statement as the changing of the probable bound within probable error to that which is without randomness or change for error, to that which denies the random nature of statistical comparison or computation. I appreciate Fendler’s comment that generalization is a stochastic process, a process that is non-deterministic or sporadic. I have noticed in my teaching introductory statistics that the concepts of generalization as a stochastic process have been not inherent. The concept of generalization seems that it should be deterministic, suggesting that when tests are run,

²⁷ Ibid., 438.

²⁸ Ibid., 442.

²⁹ Ibid.

claims should be made about the entire group represented by the sample. The tendency to prefer certainty goes along with Dewey's observation about our "quest for certainty."³⁰

In Chapter Two, I provided an example of watching a roulette wheel for many rounds providing a setting for generalizing in statistics. In the example, I explored how a gambler watches the roulette wheel and determines that either the ball should follow pattern and land on the color most observed or suggest that the other color is "due" to be spun, a misunderstanding often called the Law of Averages. In this example, I suggest that simply because the data suggest that one color is more prominent in a certain set of observed data does not eliminate the probability and potential outcome of a different color being called. In this example the result is stochastic—not pre-determined.

In the educational rhetoric of quantification, this stochastic nature of outcomes often is replaced with the potential to inform social conditions and suggest social changes, such as is the case in the educational reform texts analyzed in this dissertation. I am not writing this dissertation to suggest that educational policy or educational research cannot be informed by the use of statistics and quantification. In writing this dissertation, I suggest, however, that the use of quantification has become dominant in educational writing and research to portray the qualities of education as measured quantities, creating within the educational discourse a change from individual stories to stories that of generalized conditions. In generalizing to all, educational reform authors assume two key theoretical implications in their use of this rhetoric.

First, they assume that human qualities can be measured and that unique human qualities can be implied onto others. In prior chapters I have considered some of the difficulties that arise from quantifying qualities. However, the assumption carries additional connotations of being able to apply

³⁰ John Dewey, *The Quest for Certainty: A Study of the Relation of Knowledge And Action* (Lightning Source Incorporated, 2005).

the general to the specific. A particularly salient example of this application of quality to individuals comes from *No Excuses*. Thernstrom and Thernstrom suggest in the conclusion of the book that “Indeed, *every* urban school should become a charter.”³¹ What is being generalized in this case? The text draws from examples, particularly found in Chapters Three and Four of *No Excuses*, of successful charter schools in the urban setting, using standardized test scores as proxy for learning to compare traditional public schools to charter, with the resounding conclusion that every urban school would benefit from becoming a charter. Here, Thernstrom and Thernstrom took the quantified findings of test results for different schools and then rhetorically applying the generalization to all present and future urban schools.

Hacking suggests that this application of generalization to humans occurred in four steps: (1) Suppose that repeated measurements were taken on a single individual, creating a distribution of measurements clustering around the average height. (2) Quetelet compared this to taking “repeated observations of a single astronomical quantity. (3) Suppose there were many height measurements, although it is unknown if the heights came from the same individual or many individuals, assuming that the many individuals came from a homogenous population (a population sharing similar qualities and characteristics). The observed heights will still cluster around an average in either case. (4) “Here we pass from a real physical unknown, the height of one person, to a postulated reality, an objective property of a population at a time... This postulated truth unknown value of the mean was thought of not as an arithmetical abstract of real heights, but as itself a number that objectively describes the population.”³²

The second theoretical underpinning of generalization is that humans behave like mechanical machines to produce random results, such as the results of flipping a coin, allowing for findings

³¹ Thernstrom and Thernstrom, *No Excuses*, 265 emphasis mine.

³² Hacking, *The Taming of Chance*, 108–109.

from one context or condition to be applied to others. In probability theory, the flipping of a coin is a Bernoulli event, an event having only two possible outcomes (in this case landing on heads or tails). Repeating a Bernoulli trial n times creates a binomial random variable, where probabilities can be computed based on the knowing the probability of success (such as landing on the head of the coin), knowing the number of flips being made, and knowing that the flips are independent of each other. As the number of trials approaches infinity, the natural result is that the coin flips can be modeled through the Normal probability model, thanks to the Central Limit Theorem.

However, one of the keys to this application of the Normal model is the assumption that the trials must be independent. I recognize that as a statistician there are many central limit theorems that do not require independence, such as the set of weak-convergence theories of probability. In my work as an educationalist, however, I have noticed that these weak-convergence theorems are not important considerations in educational research and implications. The commonly accepted theorem in educational discussions requires independence; I proceed from that common educational assumption. This is apparent in the tossing of a coin. The previous tossings do not have influence on the tossing of the coin in the future. Each individual flip is independent of the flips in the past and the flips in the future. Often in statistical analysis, assumptions, such as independence, will be simplified under the banner of progressing knowledge or the further the discussions in the field. I do not dispute this simplification in general practice because not even in nature do we find data that fulfill all of the assumptions of using the normal distribution. However, One of the difficulties that I have with generalization is that the assumptions are made as if every action of a human being were independent of past events. This cannot be simplified in my mind. The past influences the agency and decisions of the present and creates non-independence. I personally cannot see how the educational reform rhetoric can assume that students sitting in the same classroom or sitting in the same school or come from the same community are independent, that they have no history, and that

they have not been affected by previous experiences. Even in the students were randomly assigned to classrooms, such as the case with the Tennessee Project STAR, there are still interactions and relationships that cannot be accounted for in the assumptions of independence. Additionally, humans do not act or respond stochastically. Humans have the ability to think and choose and respond. This response limits the random nature of humans. I am not suggesting that all of the decisions made by humans are rational, but I cannot suggest that human responses are random or mechanical either. In the case of flipping a coin, the responses of each independent trial are random, where a strict definition of random suggests that the possible outcomes are known beforehand without knowing which possible outcome will occur. In working with humans, I contend that it is neither possible to know all of the possible outcomes before they occur nor is it possible to suggest that what does occur is predictable, suggesting that human behaviors and qualities are random limits interactions and relationships.

Fendler suggests that in policy discourse, probability gets convoluted with notions of certainty.³³ Instead of recognizing that quantified evidence is probabilistic in nature, the conclusions become a type-generalized practice, applicable to all systems of education that meet certain descriptors or categorizations. If the audiences are to believe the suggestions that are being made to the point that changes are enacted, then there must be enough reasoning to suggest the change occur. The deployment of quantification takes on the role of predicting future benefits that would come through the desired changes.

Conclusion

This chapter concluded the second section of this dissertation, considering the rhetorical deployment of quantification within three educational texts. This section considered how

³³ Fendler, "Why Generalisability Is Not Generalisable," 442.

quantifications are deployed in two different ways: (1) Counting to describe the past, and (2) Using the past to forecast future conditions and recommend changes. This chapter then considered how the counts of the past are used in a conflated way to predict the future educational conditions. This chapter considered how the rhetoric of educational reform uses the data from analysis sections of the text to derive support for the future forecasts. In this chapter, I explored how there were hidden assumptions of statistical inference that allowed for generalized statements to be made, which included a generalization of qualities toward the general populace and how those generalizations permit the authors to consider how the future if recommended changes are not made.

However, in the context of modern social sciences, generalizing human qualities forms the basis for an assumed ideal, an ideal that becomes interpreted as normal. In creating this generalized ideal, people who do not share these ideal qualities become marginalized and educational policies are designed to normalize—to bring people closer to the average—as is highlighted by the current educational trend to establish a uniform standard throughout the United States with a common core. I concluded this dissertation by considering the ethical nature of generalization. This consideration was based on the difficulties that I see in assuming that human qualities can be quantified through calibrated measures.

CHAPTER FOUR DESCRIPTIVE RHETORICS OF COMPARABILITIES, REVEALINGS, AND JEREMIADS

They proceed in accordance with models or concepts borrowed from biology, economics, and the sciences of language; and they address themselves to that mode of being of man which philosophy is attempting to conceive at the level of radical finitude, whereas their aim is to traverse all its empirical manifestations. It is perhaps this cloudy distribution within a three-dimensional space that renders the human sciences so difficult to situate, that gives their localization in the epistemological domain its irreducible precariousness, that makes them appear at once perilous and in peril.

- Michel Foucault¹

Quantification in educational texts is an indication of what is happening in classrooms and in learning. Quantification has a crucial role in representing US education to the patrons of US education: the general public and the rest of the world. However, the cultural norms of educational writing for general consumption almost takes for granted that quantification is being deployed in the arguments, having come to accept its presence in the arguments. Quantification is useful depending on the questions and claims, but in this current historical context, quantification has become almost ubiquitous in educational arguments allowing for quantification to not be a useful rhetorical tool for claims about qualities. It seems that quantification has been accepted in educational rhetoric, and in some cases the acceptance is unquestioned. With acceptance can come un-recognition and assumed performance.

This chapter considers how quantification is used to describe conditions of the past. In writing this chapter, I consider different rhetorics that are deployed through these quantified descriptions. Ultimately, my intentions of this chapter are to consider more fully the issues of trying to quantify human qualities. This chapter then provides a look at three types of descriptive rhetorics that are involved in educational argumentation. I recognize that there is a trust in using quantification in argumentation, exploiting rhetorical clout in educational writing with the use of

¹ Foucault, *The Order of Things*, 347–348.

quantification. This chapter considers how descriptive numeric summaries are part of this rhetorical clout and the ethics involved in using such clout in describing and inferring general qualities.

In Chapter Two I considered the impact of Quetelet on applying quantification's summaries to human qualities, particularly noting the difficulties in establishing a metric for quality. This chapter considers the three books that deploy quantitative summaries as evidence in their arguments. The descriptive statistics provided in these texts are used as tropes in making the larger arguments about educational practices and suggested changes. I examine the rhetorical deployment of quantification to argue about the conditions of education in three current texts that address conditions of educational achievement. The three texts that I have selected each argue differently about the conditions of education, although important similarities exist. In this chapter I will explore how *The Death and Life of the Great American School System*, *No Excuses*, and *The Global Achievement Gap* use quantification to in the arguments about the characterizations of educational groups.

This chapter is not an argument about the intentions of the authors of these works or the claims of the books; instead I focus predominately on the printed texts. It is not my intent to challenge or enshrine the works of these authors, instead to consider the arguments being made and how quantification plays a role in those arguments. In this chapter I will consider how the three texts use descriptive quantifications to examine measured quantities. The issues addressed in these books are issues that get encapsulated in, and thus labeled as, educational reform. I am not writing this chapter to suggest that the arguments of Ravitch, Thernstrom and Thernstrom, and Wagner are accurate or inaccurate, appropriate or inappropriate, recommended or discouraged. Instead this chapter is about how these books use the trope of descriptive quantification to argue their stances.

I see the purpose here as about how the arguments of the texts deploy quantification as a rhetorical trope. As such, I provided a general overview of the arguments of the three texts in Chapter One. I recognize that the three books offer different arguments about the current

conditions of education. Although the arguments are different, there are common deployments of descriptive quantification throughout all three texts. I will then consider some of these approaches to argument, drawing examples from the three books.

Different Deployments of Descriptive Quantification in the Three Books

I restate that I do not see a problem with the use of quantification in educational writing and research; quantification can consider different types of questions and provide different perspectives on how education is deliberated and enacted. I see several problems in the use of quantification within rhetorical arguments. I will discuss some of the theoretical implications of those problems in the next chapter. This chapter deals with one of those problems, specifically how educational argument uses quantifications to measure human qualities, under the guise that measurement can be calibrated as descriptions of these qualities. In the current educational construction and consideration, outcomes are positioned as proxies for qualities like learning and growth. In Chapter Two I considered how the audit culture has taken hold in education, focusing on standards and accountability of those standards. Within an audit culture, there must be accountability for meeting standards effectively, which becomes the duty of the teachers and the students.

Both teachers and students are responsible for demonstrating learning—via the medium of a standardized test (in an age outcomes-based educational age), where the outcomes have been predetermined—in order to account for meeting the standards. Thus, the teachers and the students are audited through quantified measures, which may or may not represent educational qualities, such as learning. Assumptions about the purposes of education shape standardized tests while student scores on those tests determine school success and failure, where teachers (who in some cases are being paid on the merits of these scores) and students (who are put under pressures to show growth

in the school's march toward full proficiency by 2014 as outlined by *No Child Left Behind*) are limited in what appeals can be made in the process.

In this chapter, I consider the rhetoric of describing educational conditions. I have developed three different themes to explore how descriptive quantification is deployed within educational arguments. The analysis is organized not by book, but by themes that cut across all three books. The three themes are “Rhetoric of Descriptive Comparabilities,” “Rhetoric of Descriptive Transparency,” and “Rhetoric of Jeremiad.”

A Rhetoric of Descriptive Comparabilities

Thernstrom & Thernstrom

No Excuses emphasizes the existence of a racial achievement gap through the demonstrations on the NAEP, which is “the best evidence” to determine how much learning is occurring in schools.² The testing data used to bolster their argument comes from averages of subgroups, which do not represent “fixed, innate traits that are independent of the environment and cannot be changed.”³ The argument is not merely about the existence of such an achievement gap, instead directs for changes in educational systems that should be fixed. A key to understanding the purposes of this text is to recognize that comparability has been achieved through quantification and calibration, and differences have been classified through test scores.

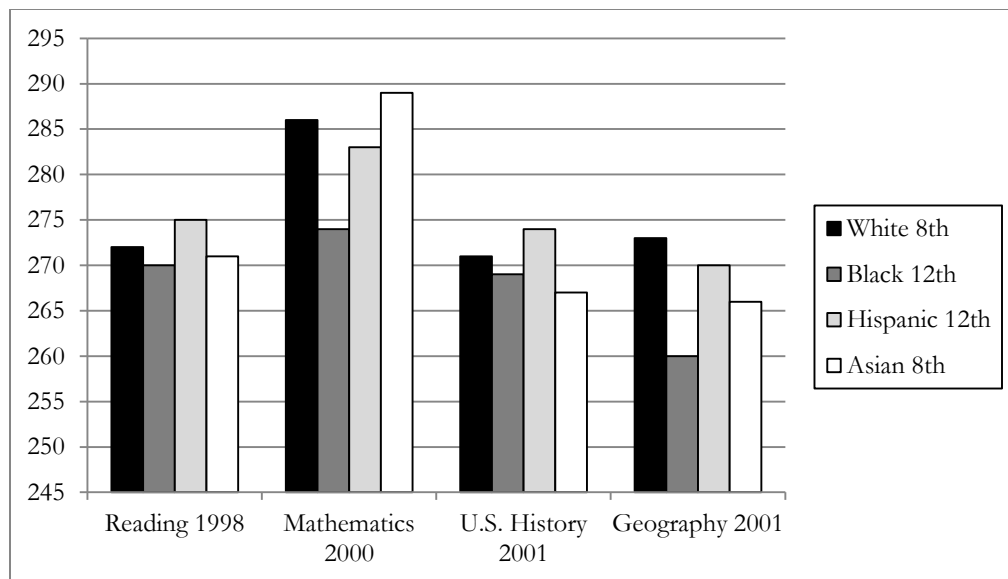
For example, the text argues that Black and Hispanic youth are four grades behind their White and Asian counterparts, or restated, that on average, 12th grade Black or Hispanic students demonstrates similar or worse scores than 8th grade Whites and Asians in reading, mathematics, US

² Thernstrom and Thernstrom, *No Excuses*, 12.

³ Ibid.

history, and geography. This claim is reflected simply by a bar graph (recreated in Figure 1) that shows the mean scores in the subjects for 8th grade White students, 12th grade Black students, 12th grade Hispanic students, and 8th grade Asian students. No numerical descriptions are made in prose, leaving the graphic image to supply the numerical data. Thernstrom and Thernstrom suggest that the figure “reveals” the findings concerning racial learning.⁴ Within the text, general arguments about how Blacks perform slightly worse in subjects like reading and history but much worse in mathematics and geography are provided; similarly, the text mentions that Hispanics “do only a little better than African Americans.”⁵

Figure 2. Mean NAEP Scores Demonstrating Four Year Racial Gap



When the argument considers those students who are classified as “below basic” on the NAEP scoring scale, the rhetorical quantification changes. The claim is that the percentage of youth

⁴ Ibid., 13.

⁵ Ibid.

demonstrate below basic skills on the NAEP exam is greater for Blacks and Hispanics than for Whites and Asians. In order to support this claim, the text again includes a bar chart separated by race and subject, this time including science, writing, and civics, but the text also writes specifically about some of the disparities that exist by race.

The figures for whites and Asians are worrisome. But the rather disappointing scores of many whites and Asians look good when compared with those of Blacks and Hispanics.

Only in writing is the proportion of African Americans lacking the most basic skills less than 40 percent. In five of the seven subjects tested, a majority of Black students perform Below Basic. In math, the figure is almost seven out of ten, in science more than three out of four.

These are shocking numbers. A majority of Black students do not have even a “partial” master of the “fundamental” knowledge and skills expected of students in the twelfth grade.

In most subjects, but particularly in math and science, Hispanic students at the end of high school do somewhat better than their Black classmates, but they, too, are far behind their white and Asian peers.⁶

In this statement, quantification is in a different format, one being a description of only data and the other being accompanied with a graphic while highlighting what the Thernstroms consider important in the display, but both are being rhetorically deployed as a comparison of 8th grade Whites and Asians to 12th grade Blacks and Hispanics. Here there is explicit use of numbers to argue the point that proportion of Blacks and Hispanics scoring below basic are far greater than Whites and Asians. There is specific mention of the “almost seven out of ten” or the “more than three out of four.” Although these numbers are not the exact numbers displayed on the graphs, the

⁶ Ibid., 15.

assumption is that the numbers are egregious enough to warrant specific display in the text as well as in the data display.

Both sections deploy univariate graphics to provide or tell or reveal the answers to questions concerning the achievement scores of these racial subgroups. I purposefully use the terms provide, tell, and reveal in this sentence because they are the terms that are used in *No Excuses*. The text posits questions to the reader about the topic at hand, whether that is Black and Hispanic students being four years behind or the impacts of NCLB, and then informs the reader that the graphic will tell or reveal the answer. The rhetoric becomes that the graphic speaks. It is assumed that the reader will interpret the graphic in the ways that support the argument in that section of the text.

Ravitch

Ravitch, however, argues about differences not among racial groups, but through the differences that exist in arbitrarily selected periods of time. This might be a result of differences in educational specialization—Ravitch is a historian whereas Abigail Thernstrom is a political scientist (Stephan Thernstrom is an historian). In arguing about the ineffectiveness of *No Child Left Behind*, Ravitch considers comparability on the national yardstick, the NAEP, prior to the passing of the mandate to the time following its passing. This concept of a national yardstick demonstrates calibration in measurement that was discussed in Chapter Two. In order to “measure” the comparability between states, districts, schools, classrooms, and individual students, calibration allows for the abstraction of qualities—such as qualities of knowledge or aptitude or performance—to quantities. Let me return to the deployment of this national yardstick in Ravitch’s argument that the national mandate was not producing the desired results, as defined by improvement on the NAEP. She states

Test score gains on the National Assessment of Educational Progress—the only national yardstick for this period—were modest or nonexistent in the four years after the adoption of the law. In fourth-grade reading, NAEP scores went up by 3 points from 2003 to 2007, less than the 5-point gain from 2000 to 2003, before NCLB took effect. In eighth grade reading, there were no gains at all from 1998 to 2007. In mathematics, the 5-point gain by fourth-grade students from 2003 to 2007 did not match their 9-point gain from 2000 to 2003. In eighth-grade mathematics, the story was the same: The gain from 2000 to 2003 (5 points) was larger than the gain from 2003 to 2007 (3 points).⁷

Here the rhetoric deployed is a quantification of comparison of individual state selected exams to the nation, the comparison becomes the increase of points on the NAEP. Ravitch is arguing at this point that NCLB is currently not working. It is not working, according to her, because the test gains on the NAEP are smaller than before the legislation was passed. During the period of 2000 to 2003, there were almost double-digit gains in fourth-grade mathematics, where after the mandate, those gains were almost cut in half. Ravitch describes the conditions of the United States, suggesting that there is ineffectiveness throughout the US educational system, not only within individual states. Ravitch is describing conditions of the past as part of her rhetoric of comparison. She is comparing test results in the countable past, although the determination of the year bins seems rather arbitrary to me. Ravitch is comparing responsibly through quantified measures.

Wagner

Wagner deploys this comparative rhetoric differently than Ravitch or Thernstrom and Thernstrom. Wagner uses citations rather than quotations in comparing different schools and offering suggestions of how to make changes. For example, in talking about the “alarmist studies”

⁷ Ravitch, *The Death and Life of the Great American School System*, 109.

that have appeared over the recent years, the text cites PISA and NAEP while not quoting the actual numbers and scores. This is not appealing to quantification instead appealing to the authority of other texts. I mention this as Wagner's descriptive rhetoric of comparison is embedded in endnotes or broad claim, such as "Other studies show that, overall, students' achievement has not significantly improved as a result of the implementation of NCLB. In fact, 12th graders' reading scores were lower in 2005 (the last reading test year for which we have data) than they were in 1992, and their writing test scores remained unchanged between 1998 and 2002 (the year of the last national writing assessment)." ⁸

The text concludes that statement with an endnote, which states that this claim is supported by "data provided by the National Assessments of Educational Progress (NAEP), a series of assessments and resulting 'Report Cards' on education, sponsored by the National Center for Education Statistics, a division of the US Department of Education." ⁹ *The Global Achievement Gap* uses percentages and rates to support some claims, but often those claims do not relate to performance on standardized assessments. Instead, the argument draws from these assessments to support claims about the current state of education, placing data in an endnote, which compels readers to look up the data themselves.

This suggests, in part, that in this argument, Wagner deploys quantitative comparisons as not primary voices in his narrative. Wagner shifts from rhetorically valuing quantification to valuing quantified comparisons as notes for reference but not as the prominent pieces of evidence. Although all three texts are writing for a lay audience, Wagner takes a different rhetorical view of quantitative comparisons, suggesting that the stories of the individual are more valued than the summaries of the aggregate. I wonder if Wagner is making assumptions about the quantitative

⁸ Wagner, *The Global Achievement Gap*, 12–13.

⁹ Ibid., 292.

literacy levels of this lay audience or if he is displacing quantified comparisons for qualitative ones out of ethical commitments.

Comments and Considerations

I am drawn to the graphics in *No Excuses* as they are major pieces in the arguments; there are graphics and figures spread throughout the introductory chapters. These charts become an important part of the argument, although not all of the charts are explained in detail. The text, if an explanation is given, draws the reader to key points of the graphic by restating the numbers demonstrated in the graphic. The deployment of this quantification is such that there is a chart or table that can “speak” more than the interpretations of the author. Edward R. Tufte suggests “Explanations that give access to the richness of the data make graphics more attractive to the viewer. Words and pictures are sometimes jurisdictional enemies, as artists feud with writers for scarce space....Words and pictures belong together.”¹⁰

As a statistician, Tufte’s words have been ingrained into my statistical presentations—explaining to the reader precisely what I intend for them to glean from the graphic. The author of the report explains predetermined purposes so that those who do not know might. This becomes a type of teaching. However, the inclusion of graphical displays changes the focus from only reading the information the author intends to be read to including opportunities for the readers to explore and compare based on the availability of more data as opposed to just what is being shown.

An important consideration in the role of rhetorical comparison is the use of graphics in the arguments. As noted, *No Excuses* uses graphics in the first section of the book to support the comparisons between racial demographics, using them to reveal the conditions of the racial gap.

¹⁰ Edward R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. (Cheshire, CT: Graphics Press, 2001), 180; see also Edward R. Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, CT: Graphics Press, 1997).

However, it is important to note that the other two texts do not use graphical representations in the texts. Although all three texts compare educational outcomes and conditions through numbers, only the Thernstroms place a visual representation within the text, allowing for additional comparisons beyond what is included in writing. Here the rhetoric shifts from explaining what the author intends by only including statements of numbers that help fabricate their arguments opening a space for additional interpretations, interpretations of interest to the reader. Allowing for the reader to explore data for themselves through presenting it in some form to the reader positions the reader not as a learner, coming to the argument in a deficiency but as a potential contributor to the discussion, able to consider more the argument being made as they have the potential to engage the data in thoughtful ways instead of just being told what the comparisons mean.

The inclusion of graphical displays shifts the rhetoric of comparison from the rhetoric of explanation to the potential rhetoric of exploration, creating an equality between the reader and the author. This equality is not an equality of physical things, instead an equality of intellectual ability and contributions. In making these rhetorical moves to explain explicitly, I suggest a counter-thought based on my readings of some of the works of Jacques Rancière.¹¹ Rancière provides important considerations of equal intelligence and how anyone can inform their own practices and understanding. I mention Rancière as I consider the rhetorical work of description to be a space where there could be intellectual equality and not forced explication or explanation; as such, he is an important counter balance to the explicative nature of educational rhetoric, particularly in considering the use of explaining descriptive quantification.

¹¹ Jacques Rancière, *The Ignorant Schoolmaster: Five Lessons in Intellectual Emancipation* (Stanford University Press, 1991); Jacques Rancière, *The Politics of Aesthetics: The Distribution of the Sensible*, trans. Gabriel Rockhill (Continuum International Publishing Group, 2006); Jacques Rancière and Steven Corcoran, *Dissensus: On Politics and Aesthetics* (Continuum International Publishing Group, 2010).

In his essay *The Ignorant Schoolmaster*, Rancière suggests that the key to an ignorant schoolmaster is in the interpretation of reasoning. Gert Biesta and Charles Bingham suggest an interpretation of Rancière's work. "The usual aim of pedagogical logic is to teach the student that which he or she does not know, to close the gap between the ignorant one and knowledge. Its usual means is explanation."¹² For Rancière, the act of explanation supposes "limited capacities" of the receivers.¹³ Explanation creates a mentality that separates those who know from those who don't. In offering explanation about quantitative descriptions, the argument becomes embedded in foreclosing possible interpretations and questions about the data presented. There becomes a specific desired outcome, which is presented through the writing—declaring a moral to the story. In my reading of Rancière, I believe that there are opportunities for multiple interpretations.

From this reading, the rhetoric of comparison in educational reform texts could be seen as rhetoric that assumes a pedagogical logic, where the purpose of the rhetoric is to explain to the reader what the quantitative descriptions explicitly mean in the contexts of education and the desired reform. However, from the work of Rancière, I suggest that there are ways to deploy rhetoric of comparisons without limiting the capacities of the readers by beginning with the assumption that the reader is capable of understanding and exploring and offering a way for readers ways to suggest they are capable of being in the know, beyond having what should be done in education explained to them. This rhetorical shift I think includes the presentation of the data in ways that allow for comparative exploration from the reader, such as with the inclusion of graphics.

¹² Charles Bingham and Gert J. J. Biesta, *Jacques Rancière: Education, Truth, Emancipation* (London: Continuum International Publishing Group, 2010), 3.

¹³ Ibid.

A Rhetoric of Descriptive Transparency

A second kind of move I call a rhetoric of descriptive transparency. I have struggled with this label, not because the term is anything magical, but because the connotations and associations with this label shape considerations of this rhetorical deployment. In using the term transparency, I consider how quantification might be deployed as a clarifying rhetoric, being used within the argument as a tool explaining what current conditions actually *are*. I believe that this rhetoric might be understood through the objectivity and generalizability desired in scientific research, which have become desired in educational research and educational reform texts.

Here I suggest that quantification can be deployed as a mechanism for transparency, clarifying, from the rhetor's perspective, the truths about the current conditions of education, particularly in this historically specific notion of outcomes-based education, are exhibited through the achievements and demonstrations on calibrated, standardized exams. This section, then, explores some of ways that Ravitch, Thernstrom and Thernstrom, and Wagner use quantification to make transparent the current conditions of educational achievement. I should caution that this section is neither an endorsement or decrying of this rhetorical use; I contend that this is a choice in argumentation, which I describe.

Arguing for School Proficiency Transparency

In *The Death and Life of the Great American School System*, Ravitch suggests that current school reforms are not working, particularly in demonstrating proficiency on standardized tests. In one chapter, Ravitch writes about the impacts of *No Child Left Behind* on US schools. She argues that the reforms of this mandate were not working. She suggests that she realized this after attending a conference held at a conservative think tank. During the conference, scholars “presented persuasive

evidence” on the ineffectiveness of these reforms.¹⁴ *The Death and Life* continues by presenting to the reader pages of quantified evidence suggesting the NCLB was not helping and possibly hindering education. Although the text contains many quantified references, I focus on the invocation of NAEP scores for the moment.

Ravitch suggests that *No Child Left Behind* could never work unless the states adjusted what it means to be proficient in content areas. Ravitch suggests “most states devised ways to pretend to meet the impossible goal.”¹⁵ This was allowed in the mandate as each state was allowed to pick its own standards, determine which tests would be used in assessing those standards, and defining state-by-state what proficiency meant for that state. Ravitch gives the example of Mississippi “Mississippi claimed that 89 percent of its fourth graders were at or above proficiency in reading, but according to NAEP, only 18 percent were.”¹⁶ Since the mandate allows states to determine what examinations are administered and what the levels for proficiency are, NAEP again appears as a source for testing the claims of the validity of NCLB. Ravitch suggests that the NAEP is the only national yardstick (alluding to the concepts of calibration, which I discussed in Chapter Two) for such a comparison as states diminish requirements for proficiency so that yearly progress is made.

Given the necessity to report gains, many states reported steady—and sometimes amazing—progress toward the mandated goal of 100 percent proficiency. Texas, for example, reported in 2007 that 85.1 percent of its students in grades four and eight were proficient readers, but on NAEP tests, only 28.6 percent were. Tennessee claimed that 90 percent of its students were proficient readers, but NAEP reported that 26.2 percent were. Similarly, Nebraska told

¹⁴ Ravitch, *The Death and Life of the Great American School System*, 99.

¹⁵ Ibid., 106.

¹⁶ Ibid.

the public that 90.5 percent of students in these grades were proficient, but NAEP said the number was 34.8 percent.¹⁷

Here the use of quantification develops the differences between the standards established by individual states and those established by NAEP. Given an impossible goal, the states were required to demonstrate, through students' representations on a test, that there was growth and improvement, that more students were becoming proficient.

However, in this the term proficiency came to mean completely different things, measured at different levels. The assumption of the argument is a basic concept of percentages, suggesting that the reader has an understanding of what it means for only 35 percent of students to be proficient. Thus, the argument proceeds logically that there is inflation occurring within states. This inflation was not a result of manipulations by the states, instead a lack of consistency of what it meant to be proficient between the national assessment and the individual state assessments. It might be a cliché to invoke a commonly mentioned phrase about lies and statistics here, but in Ravitch's argument it seems that the states are claiming that children and youth are more proficient than would be demonstrated in other testing situations. Here quantification is being deployed to make transparent the lack of consensus in establishing what is meant by proficiency in the US educational system. The numbers, for example, of Tennessee make clearer one way the actual learning and achievement can be perceived as true.

In order to understand this rhetoric, I provide a little background about the discrepancies in comparing state outcomes to national outcomes. In the passing of *No Child Left Behind*, each state was authorized to choose what standardized test it would use in measuring how well the students are making progress toward becoming proficient in the core subjects of reading and mathematics. Each state was then authorized to establish its own level of what it meant to be proficient based on the

¹⁷ Ibid., 161–162.

test chosen. Thus, a student in Massachusetts would be measured on a different test at a different level for proficiency than a similarly aged student in say Michigan. The states began with a different definition of proficiency than what is reported as proficient on the NAEP. There was no national standard of what it meant to be proficient or to pass the state test, each state having the ability to determine that on its own. Thus, the states created a system to meet an unobtainable goal of getting all of the students to be proficient by the year 2014.

I believe that having the term proficiency mean completely different things depending on the states' interpretations and testing versus the standard established by NAEP is an example of how this descriptive rhetoric relies on the attempts to measure qualities (student learning of mathematics and literacy) into measured (and auditable) quantities. Here the states and NAEP are both trying to determine if children and youth are proficient in mathematics and literacy, relying on the meeting of certain numeric quantities that represent learning qualities. I ask how is it possible to measure, count, and analyze a person's abilities without such discrepancies and biases. It is not difficult to see how one state would measure the quality of aptitude differently than another which could be completely different from a national measurement.

Arguing about the Transparency of Choice

This type of rhetoric is also deployed in Ravitch's comparisons of school choice. A key portion of her argument about school reform comes through consideration of school choice and marketization. Quantification is deployed in her argument about the conditions of school choice through the description of "the data wars" which compared the effectiveness of charters to traditional public schools.¹⁸ The text is filled with examples and evidence from these wars. The

¹⁸ Ibid., 138–144.

rhetoric is different, often citing claims made by other researchers but not the actual quantifications that were used to determine these claims.

As part of these data wars, Ravitch presents differing descriptions of the educational conditions between traditional public schools and charter schools. The descriptions are deployed by different groups to suggest that public schools do just as well as charter schools or that charter schools perform better than public schools. In considering this rhetoric of descriptive transparency, I summarize quickly the diversities in reports that are used in Ravitch's argument. I do so in an attempt to highlight how this rhetoric of descriptive transparency suggests different findings that is it describes the qualities of learning differently.

- The American Federation of Teachers “learned that NAEP showed no measurable differences on tests of reading and mathematics between fourth-grade students from similar racial/ethnic backgrounds in charter schools and in regular public schools....Overall, charter and public students performed similarly in reading, but public school students performed better in mathematics.”¹⁹
- “Caroline M. Hoxby published a comprehensive study comparing charter schools and their nearby public schools. Hoxby...found that they (students at charter schools) were more likely to be proficient in both reading and math than public school students.”²⁰
- In July 2006, the US Department of Education released findings that compared public to private. “Public school students performed as well as or better than comparable children in private schools. Private school students scored higher on average, but their advantage

¹⁹ Ibid., 138.

²⁰ Ibid., 139.

disappeared when they were compared to public school students with similar characteristics.”²¹

- In reporting a study performed by Thomas Kane in 2009 which studied Boston charter and public schools, suggesting that charters had significant impact in the middle and high school years. “The gains were especially large in middle school mathematics, where students moved from the 50th to the 69th percentile in performance in one year—about half the size of the black-white achievement gap.”²²
- Citing a national study from 2009, the text concludes that most students in charter schools perform no better than those in traditional public schools. The study “found that 37 percent had learning gains that were significantly below those of local public schools; 46 percent had gains that were no different’ and only 17 percent showed growth that was significantly better. More than 80 percent of charter school in the study performed either the same or worse as the local public schools.”²³

In the rhetoric of descriptive transparency, quantification gets deployed through fabricating a received truth as to which school system is better. One source of evidence, the US Department of Education, suggests that public schools do just as well or better than private schools. Another source suggests charters promote percentile change in the middle and high schools. Which description is to be believed? How is it that both schools claim benefit?

It is possible that all of these claims could be correct, depending on what perspective and focus one refers. This becomes a question of what is does good education look like and what types of reforms are necessary for what types of education (discussed in further detail in Chapter Six).

²¹ Ibid., 140.

²² Ibid., 141.

²³ Ibid., 142.

Within this question, there is an additional layer of how does good education get evaluated in this age of measurement. In this data war, as Ravitch describes it, there are different perspectives and points of interest and depending on how the data is portrayed (such as what ranges are chosen to represent end points for comparison), depending on what aspects and qualities of education are valued, and depending on the how the argument is constructed deploying these numbers, all of which could be correct. I find this interesting as part of the point of trusting numbers, according to Porter, is that there is an objectivity in numbers, yet as this bulleted list points out, the objectivity is lost depending on what is valued in evaluating education and how those data points are portrayed in the text. In a call to be objective, the trust in numbers has become manipulable.

Arguing for Transparency of Global Knowledge

In contrast to Ravitch's considerations of charter and private schools, Wagner's argument suggests that there are schools that prepare US students to compete in the global knowledge economy, at least schools that are measured through the implementation of Wagner's Seven Survival Skills. His argument concludes that there is a way to prepare future workers to take part in this global climate, where most of the successful examples offered in text are charter schools. Wagner concludes that these schools are shifting focus toward learning instead of on memorizations for a test, these schools are motivating students, and these schools are thinking about how schools are being accountable for the preparation of their students for college.

Wagner's implications come from the conclusion and afterword (published for the paperback edition). However, the evidence that supports these conclusions is found in the final numbered chapter which offers a glimpse at three successful, small schools. Drawing from these schools, Wagner's argument portrays effective education as educational systems that evaluate the work of teachers through the competencies of those youth within the classrooms. He suggests,

through quoting a Virginia Beach superintendent, “the only real evidence of critical thinking happening in the classroom was in *what the students were doing*.”²⁴ Thus, the argument of *The Global Achievement Gap* portrays that schools should spend more time assessing through graded but not standardized portfolios or “performances” instead of measuring through standardized examinations. As part of this publication, Wagner suggests that students be required to have digital portfolios that are accumulated through the high school years, particularly through the collaborative efforts of the teachers within the schools. This would allow for “the concept of performance standards and performance-based assessment” in the schools.²⁵ “I would like to see much more attention paid to the idea of performance standards.”²⁶

Wagner writes against evaluating student learning on test scores, suggesting, “test-score improvements don’t tell us very much about what students know and are able to do.”²⁷ However, little is mentioned of how evaluations are offered beyond the fact that there is some form of portfolio with collections of student work and performances that are graded by some form of rubric. I wonder if these portfolios are graded qualitatively, based on suggestions and improvements or quantitatively, through assigned scores and metrics. I mention this as the later suggests a different use transmutation of the qualities of learning into a different type of standardized quantity.

The argument tells the stories of three schools that are successful in implementing these ideas into their construction. Wagner pays particular attention to their graduation rates and mentions their achievements on standardized tests, although Wagner does not provide quantified data, just mentioning the comparable rates of these schools with other students in traditional public schools.

²⁴ Wagner, *The Global Achievement Gap*, 280.

²⁵ Ibid., 287.

²⁶ Ibid.

²⁷ Ibid., 280.

Although Wagner argues about the success of these schools through interviews with students and school administrators, I provide only his quantified summaries. I will summarize two different schools, High Tech High and The Met, before I consider how this is a rhetorical deployment of transparency.

The first object for his conclusions comes from High Tech High, a high school that has been viewed as successful in San Diego. “Since graduating its first class in 2003, 100 percent of High Tech High students have been accepted to college—80 percent to four-year colleges....The national average of college graduates who get a technical degree from college is 15 percent. High Tech High’s is 27 percent, the result of the very different approach the school takes to teaching math, science, and engineering.”²⁸ This school provides an example of a technologically focused high school, which has projects that require not only student collaboration but also teacher collaboration to complete. Or consider the success of a school system in Providence, Rhode Island—The Met. This system of schools is a conglomeration of several small schools which have been influenced by charitable organizations. “The Met,” Wagner suggests, “has gained a national reputation for graduating nearly 100 percent of its students, with 95 percent of its graduates accepted into a two- or four-year college.”²⁹ The final school suggested as a model for change is the Francis Parker Charter School in Boston. Wagner summarizes this school as having an impressive track record.

It has consistently ranked among the top-ten schools in the state’s MCAS tests—while absolutely refusing to teach to the tests. Since graduating its first class in 2000, 100 percent of its students have been accepted to college, and 95 percent of Parker graduates have gone

²⁸ Ibid., 207–208.

²⁹ Ibid., 232.

on to college—with 96 percent of those students attending four-year colleges. The college graduation rate is 85 percent.³⁰

In this example, Wagner uses quantification in making transparent the successes that are available to certain schools that follow non-traditional models. The role of quantification in this rhetoric comes from specific consideration of graduation and acceptance rates into a college or university. In this case, the number of graduates and the numbers attending college represent some proxy for the qualities of effective education, being deployed rhetorically as a tool for making the successes of the schools more transparent to the general public, offering a transmutation of what good education looks like in a world desiring audible measures. The issue becomes one of how can graduation rates represent the qualities of good education that Wagner spends chapters highlighting. Wagner suggests that success is measured by having rates higher than the national average. Wagner's argument has been about implementing his survival skills into public schools. However, the inclusion of quantification does not address that argument, instead providing graduation rates.

Wagner's point is different from Ravitch's described above. He is suggesting that charter schools make a difference, basing his contentions on the graduation rates, college acceptance rates, and test scores of several selected schools. These rates and scores are standing in proxy for desired educational qualities. Whatever the quantity being deployed, and whether for or against charter schools, these rhetorical deployments are consistent in the transmutation of human quality into quantity. Ravitch argues against charter schools; Wagner argues for them. Both authors make their claims on the basis of comparison and ranking, comparing students' successes (whether through test scores or graduation rates) and ranking which school type is deemed more beneficial for youth and children. Regardless of ideological position, it seems that quantification is being valued highly by

³⁰ Ibid., 242.

authors of educational reform texts. The claims of needing educational change and making that change in the desired direction of the author are supported by the inclusion of numeric data.

Arguing with the Transparency of Test Scores

I conclude with an example of descriptive transparency from *No Excuses*. I recognize this deployment is different from that outlined above; I do so to demonstrate a broader range of deployments of this rhetorical trope. Thernstrom and Thernstrom use this rhetoric of descriptive transparency to suggest clarification supporting their thesis of standardized testing as an effective and efficient way to measure qualities of learning. Although they use test score descriptions throughout the text, an appealing use of this rhetoric comes as they portray that the US public accepts the thoughts of using standardized tests. I mention this as a different example than the examples of test scores and measures.

In Thernstrom & Thernstrom's Chapter Three, the focus shifts to current discourse of standardized testing and why such testing is necessary to "tell us what students, educators, parents, and the general public need to hear."³¹ It seems that test scores take on the role of a voice telling the tales that should be told. *No Excuses* portrays that there are other methods of evaluating students, such as through portfolios or grades, but the argument of this text suggests that such methods are incomplete in informing the public's understanding of what children are learning and doing. Thernstrom and Thernstrom describe how the qualities of public support have been quantified, suggesting a transparency in supporting the inclusion of standards and testing. I share two quotes from *No Excuses* that demonstrates this rhetorical principle.

³¹ Thernstrom and Thernstrom, *No Excuses*, 39–40.

Table 1. Support or Opposition to Standardized Assessments

Indeed, the polls do indicate high levels of support, in principle, for both standards and tests. Public Agenda is a nonpartisan, nonprofit, widely respected public opinion research organization. In September 2000, it found that only 11 percent of parents thought “schools today place far too much emphasis on standardized test scores.” Seventy-one percent supported “testing students at a young age...because struggling students can be identified and helped.” Fifty-five percent said there was “nothing wrong” with teaching to the test, since it measured “important skills and knowledge.” In a national survey commissioned by the Business Roundtable a month earlier, 65 percent of parents and 70 percent of the general public said students should “pass statewide tests before they can graduate from high school.” Those percentages went up when people were told the students “could take the tests several times.” The surveys did not break down their results by race, but a Public Agenda poll in the winter of 1997-1998 found that 78 percent of black parents agreed that testing “calls attention to a problem that needs to be solved.”³²	A 2001 survey by the Business Roundtable found that a large majority of Americans were opposed to relying solely on tests to determine high school graduation, and 80 percent believed that some students don’t show what they know on standardized assessments.³³
---	--

The rhetoric of descriptive transparency is deployed in the first quote to clarify the position of the author. This quote, for me, demonstrates how this rhetoric can be seen in educational texts that do not derive from quantifying learning. In this case, the use of quantification describes public opinion in attempts to clarify the authors’ stance that standards and standardized testing is necessary and beneficial. In doing so, the authors of *No Excuses* clarify their argument through quantifications that the adult US public agree with them in this venture, appealing to a sense of numeric consensus.

However, I note two hindrances of this rhetoric.

³² Ibid., 26.

³³ Ibid., 30.

First, *No Excuses* was written after the passing of *No Child Left Behind*, which is the messenger “of a glaring racial gap” that exists in the United States.³⁴ However, the use of descriptive transparency is based on opinions gathered from reports published in 1997-98 and 2000. The passage of *No Child Left Behind* was not until after this support for standards and standardized testing was claimed. In the use of these quantified opinions, the text is suggesting that the public supports the use of standards and standardized testing; however this poll was taken prior to a mandated implementation in the nation’s schools. It seems as though this deployment for supporting standardized measures is hiding behind historically different contexts and circumstances. Thus, the Thernstroms are arguing that with *No Child Left Behind* there is greater emphasis on national standardized testing, which is part of their suggested recommendations for education. However, how do these polls and surveys calibrate for the changes of time?

Second, this deployment comes with caveats, which the text recognizes several pages later and shared as the second quotation in Table 1. The text considers that there is large support for testing but not at the cost of limiting high school graduation or as accurate measures of what is known by the students.

It appears that there is large support for standardized testing, which shows where problems are, yet there is a large percent of people who believe that standardized testing does not allow all students to show what they know. Quantified evidence is thus suggesting differences. There is an appearance of public support in the use of standardized tests to measure learning. The crucial question in this situation is if it is really possible to convert the qualities of support into measurable quantities and the rhetorical value of the quantified proxies. Are adults truly satisfied with the directions of schooling and the implementations of new measures and standards? Many would suggest finding this answer through polling, another method for converting the quality of support to

³⁴ Ibid., 25.

descriptive counts. In dealing with humans, polling has become a standard method for converting qualities, like support, into comparable quantities.

Of interest in this argument becomes that in the U. S. educational system, the value has been placed on these quantified proxies. The Thernstroms ultimately are writing about the issue of civil rights. They could have written from an ethical perspective arguing about the nature of inequality and effects on humans. Instead the text deploys comparisons of test scores and public opinion polls in efforts to support their claims that there are people who are being treated horribly and unethically. Here two respected authors, academics, and fighters for civil rights are arguing that inequalities exist and must be changed through trusting in quantified arguments. I might ask what qualities are lost in the creation of these quantities. Again, the Thernstroms suggest that “indeed, the polls do indicate.”³⁵

Comments and Considerations II

I believe that all rhetorical texts are put together with selections about what evidence to deploy where and when. There is careful construction of the text to best portray and persuade. This is not to suggest that the arguments are written to deceive but there are always more stories available than the one presented in the argument. The descriptive quantifications were selected to aid in the telling of a certain story, such as the narrative that states are choosing to define proficient in different ways than national tests or when quantification aids in demonstrating public support.

To me, it seems apparent that arguments are structured purposefully with their evidence carefully selected. That is the point of rhetorical study. However, I argue that although quantification in arguments is no different from other forms of evidence, the deployment of this rhetorical trope suggests that there are other portrayals that might be considered through the deployment of these

³⁵ Ibid., 26.

numbers. In both of the examples I provided, different interpretations are provided within the argument. I originally titled this section using the words sleight-of-hand. I recognize the nefarious nature of this term, often connected with trickery. I believe that this rhetoric is misdirecting the readership of the argument through appeals to numbers, as arbitrary as they may be, rather than to the qualitative and humanistic ethics of treating people equitably.

A Rhetoric of a Jeremiad

I conclude this chapter by considering a third type of rhetoric based on the use of descriptive quantifications. I recognize that there are other tropes that could be considered. I, however, offer this as a look into rhetoric that I consider a result of educational argumentation in the United States and US educational reforms. Educational reforms in the United States are written for the purpose of changing current educational conditions and results. It is not difficult to see how educational argumentation has been influenced by fields like psychology and the learning sciences. However, I also believe that educational argumentation has been influenced by the rhetorical stylings of the United States. As such, I conclude this section by considering the arguments of *The Death and Life of the Great American School System*, *No Excuses*, and *The Global Achievement Gap* through the concepts of Sacvan Bercovitch's text *The American Jeremiad*.³⁶

Bercovitch analyzes carefully a rhetoric pattern common in the United States, particularly the ability for US authors, public speakers, religious leaders, and politicians to invoke within the listener or reader a desire to change based on "scriptural" basis. I write the word scriptural in quotes as it relates not necessarily to the Christian canon, but some basis of moral or ethical or societal law that is currently broken. The form of the American jeremiad, Bercovitch suggests, is that the rhetoric establishes some law or norm that is being broken, suggesting punishment or retribution or casualty

³⁶ Bercovitch, *The American Jeremiad*.

that will come if the norm is not restored, followed by the distinctly US rhetoric of a promise of salvation and benefit from returning to the law or norm. In this rhetorical deployment there is a sense of punishment and fear for what might come if changes are not enacted.

This section is not about how to fix these problems or the generalizations of these problems, instead it is about how quantification is used to instill in the public a desire for educational change based on edict and fear of what might come if change is not enacted.

How Did the Home Team Do?

Wagner argues about this ineffectiveness of the US educational system also through test scores, but not the NAEP. He takes from an international comparison test offered by The Organisation for Economic Co-operation and Development (OECD), the Programme for International Student Assessment (PISA). PISA is a standardized assessment of reading, mathematics, and science given to 15-year-olds. Wagner quotes PISA's findings to suggest that about one in five 15-year-olds demonstrate, according to the test, skills that would classify them as reflective and communicative. One in five suggests that in the international frame, there are youth who are capable of highly marketable skills.

“How did the ‘home team’ do?” asks Wagner, drawing on a sports metaphor. “Badly. Very badly, indeed.”³⁷ It becomes difficult to cheer for your team when winning is less than likely. It becomes difficult to spend the money on a poor performing group. It becomes difficult to see hope when performances are diminished. In Wagner’s case, he is not considering a local professional team, instead the conditions of education.

Nearly one-quarter of US students scored below a level 1—a level far lower than that achieved by students in the OECD countries. A lower percentage of US students than

³⁷ Wagner, *The Global Achievement Gap*, 74.

OECD students scored at levels 2 and 3. And in four countries (Finland, Hong Kong—China, Japan, and Korea), 30 percent or more of students performed at level 3 in problem solving, compared to only 12 percent of US students. This analysis reveals that even the kids we consider to be our most academically talented are not even close to the competition “On average, US high achievers for problem solving (those scoring in the top 10 percent in the United States) were outperformed by their OECD counterparts. To be in the top 10 percent of students in the United States, students needed at least a score of 604...but 675 or better in Japan.”³⁸

How badly did the home team do? Wagner suggests that one-quarter of US 15-year-olds scored below level 1, with the statement that this level is far lower than other participating countries. Wagner does not support the claim that the United States is far below other countries by providing a comparison point from those countries. However, the percent of high scoring 15-year-olds in the United States was reported at 12 percent while in four other countries the rate was 30 percent. This rhetoric returns to the use of percentages and being able to compare across subgroups by declaring what percentages of 15-year-olds fit within certain qualities.

However, in this paragraph, Wagner also turns to the rhetoric of percentile, suggesting that in order to score in the 90th percentile (or top 10 percent according to Wagner) that the students must demonstrate a score of 604, 71 points lower than in Japan. This rhetoric suggests that the best players for the home team, to carry Wagner’s metaphor slightly further, are playing far below the demonstrated skills of other nations, particularly exemplified through Japan’s scoring. What does this mean? In Wagner’s argument, this way of reporting statistics suggests that the top scoring US 15-year-olds are not close to the top scoring Japanese 15-year-olds, suggesting that “If I’m an

³⁸ Ibid.

employer of a multinational corporation,...,all other things being equal I'm likely to locate my new facility in a number of other countries before I'd consider coming to the United States.”³⁹ Wagner is making a leap from the higher test scores to the better an employee a person would make. It seems that only by abstracting qualities into quantities can this leap be made effectively. How can the score on the NAEP determine the qualities of an employee, such as the quality of punctuality, determination, proficiency, leadership, and work ethic? In abstracting qualities of learning to quantities of measured on certain tests, Wagner is suggesting that future desired qualities of a worker could be inferred.

This type of argument positions the United States ultimately in comparison to other countries. As mentioned in Chapter Two of this dissertation, valid comparisons require calibration and calibration requires quantifications. However, as I have mentioned throughout this dissertation, quantification of qualities can be problematic for all types of information. The difficulty in comparison is that objects are qualitatively different (they differ in qualia). But within this framework of the jeremiad, there must be comparison to support the author's call for change in order to return to a “scriptural” standard. Thus, comparison requires calibration which requires quantification. I contend that part of the reason the United States holds to descriptive comparison is due to the rhetorical traditions of the jeremiad.

Wagner is appealing to the emotional connection that the general readership in the United States would have to their home country. In appealing to the emotional connections to the general US readership, Wagner pays particular attention to the potential loss of US workers, particularly to other industrial countries. If all else were equal, he would hire workers from other countries. I question the type of argument that is being made by evoking the equality of everything else. All else

³⁹ Ibid., 74–75.

being equal is not possible in any condition. This type of rhetorical deployment cannot serve as a logical appeal, but it may be an effective appeal in the way a jeremiad is effective.

A Shamefully Ignored Issue

This rhetoric is not only found in international comparisons and futures. Thernstrom and Thernstrom argue through this type of rhetorical quantification as well. They describe how racial inequality in educational achievement has been treated as a dirty secret in the past. However, “this shamefully ignored issue has moved to the front and center of the educational state. In part, the new attention is simply a response to an altered economic reality.”⁴⁰ Moreover, they suggest the individual and nation’s future is jeopardy because of the demonstrations on these standardized tests.⁴¹

They argue that the future is in jeopardy through racial comparisons of scores on an adult literacy test (testing both prose and quantitative literacies) and through a comparison of earnings by race. Again these comparisons come through graphics. They suggest

the average black college graduate, Figure 2-2 indicates, is no more adept at reading prose than the typical white who attended college only briefly, leaving before receiving a two-year degree. Even worse, Figure 2-3 shows that the quantitative skills of the average black with a bachelor’s degree are no stronger than those of whites who only graduated from high school and did not attend college—a four-year gap.⁴²

There were few differences between Blacks and Hispanics, still below White levels. How might an “all else being equal” argument look in this case? Of course they do not mean to imply that Whites

⁴⁰ Thernstrom and Thernstrom, *No Excuses*, 3.

⁴¹ Ibid., 40.

⁴² Ibid., 36–37.

only be hired in demanding jobs because they have the reading comprehension and quantitative literacy. However, given the framework of their argument, that preposterous conclusion could be drawn.

Before those questions are answered too hastily, the Thernstroms provide argument that the future of racial minorities is also impacted through the amount of education they receive. *No Excuses* suggests that there is a “striking relationship between income and schooling,” which is made clear by the figure in the text.⁴³ In the argument of *No Excuses* it is not enough to suggest that as the average education level increases, the average income also increases. The argument specifies how this relates to racial minorities (not providing any information for Asians). They explain:

Among those with fewer than nine years of education, Latino incomes were 9 percent below those of whites and black incomes 24 percent below. White college graduates earned 15 percent more than African Americans with a college diploma and 23 percent more than comparably well-educated Hispanics.⁴⁴

The explanation in writing suggests to the reader key pieces of information that should be taken away from the reading. The writing maintains that not only do Whites earn more money, but also there is a large disparity in income between Whites and those classified in this evidence as minorities.

Comments and Commentaries III

In talking about the rhetoric of monstrous, Edward J. Ingebreetsen suggests

The theater of fear, then, is pedagogical, teaching by preemptive example. It is also participatory and interactive, intended to be habit-forming.... Rather, ceremonies of fear,

⁴³ Ibid., 37.

⁴⁴ Ibid., 38.

like other social theatrics, adopt their own ends conventions, motifs, and images pilfered from many sources.⁴⁵

Fear-filled rhetoric still informs arguments, although through its own accepted conventions and norms. These norms have filled the deployment of quantification through fears of what is to come. This rhetorical deployment has become almost standard in educational texts, particularly in texts that desire to address policy concerns in education. This rhetoric is attached to the future and the conventions of this rhetoric depend on appealing to the nature that education is a public good that is necessary for maintaining economic presence and development.

Jeremiad rhetoric tries to persuade us of peril. In arguing about educational reforms, descriptive quantification reminds the audience not only of these perils, but also provides a way to summarize the magnitude of that peril. In this section I have written about how the educational reform texts analyzed in this dissertation have deployed numeric descriptions in an attempt to provide a warning of what is wrong and an invitation to ward off impending peril. This rhetoric invites comparisons, which require calibration and quantification, in an attempt to warn of what educational qualities are to be changed.

Rhetorical Descriptive Quantifications

This chapter was not an inclusive consideration of the possible uses of descriptive quantification in educational texts. However, I believe that this chapter has aided in understanding that quantification has taken an active role in educational rhetoric. I do not consider this inappropriate. However, the difficulty arises when human qualities are converted into quantities for purposes of comparison. In this chapter, I have considered three ways descriptive quantification is

⁴⁵ Edward J. Ingebreetsen, *At Stake: Monsters and the Rhetoric of Fear in Public Culture* (Chicago: University of Chicago Press, 2001), 21.

deployed rhetorically. I recognize that there are other possibilities. It was not my intention to construct a full and comprehensive list of quantified deployments. In considering the rhetoric of descriptive comparability, the rhetoric of descriptive transparency, and the rhetoric of jeremiads, I have presented different ways descriptive quantitative summaries are deployed in educational text. Ultimately, these three rhetorics demonstrate the assumptions that are possibly overlooked in measuring human qualities through quantified means.

These texts all deploy different rhetorics of comparison, transparency, and jeremiad. Yes, all three texts do use quantification to make these arguments, but this is not the only way that these arguments could have been made. In writing these books, the authors could have used rhetoric of comparison, rhetoric of transparency, and rhetoric of the jeremiad without relying on quantification to consider the conditions of racial inequality, traditional public schools versus charter schools and vouchers, and the conditions of U. S. students in their readiness compared to international markets. The important point is that the authors of these texts chose to rely on quantifications to support their arguments, placing rhetorical value on the presentation of descriptive quantifications.

What is the difference in using quantification within these arguments and not using quantification? This question becomes an issue of ethics and values in argumentation. The major difference in the deployment of quantification is that it is valued in the educational reform discourse above arguing through humanist appeals. As was discussed in Chapter Three, the purposes of this deployment is the ability to infer (whether that is to infer about a generalized characteristic or to infer future conditions) based on the descriptions of the past. Instead of valuing arguments made by appeals to the human condition, appeals to the individual, these texts have argued about changing human condition through appeals of finding general trends in efforts to delocalize the philanthropic arguments being made.

Chapters Three and Four have been considerations about the rhetorical deployments of quantification in educational reform texts. These two chapters consider three texts and how the authors of these texts have rhetorically valued the objectivity of quantification in making their arguments. Again, I am not suggesting in this text that quantification is appropriate or inappropriate, as quantification does offer contributions in understanding and arguing educational reform. However, I note that there are some ethical considerations when quantification becomes so valued in educational reform writing. The remainder of this dissertation considers five of those problems of valuing quantification within educational reform texts.

CHAPTER FIVE

YUCK: A CHAPTER ON THE ASSUMPTIONS OF QUANTIFICATION

We need individual stories. Without individuals we see only numbers: a thousand dead, a hundred thousand dead, “casualties may rise to a million.” With individual stories, the statistics become people—but even that is a lie, for the people continue to suffer in numbers that themselves are numbing and meaningless.

- Neil Gaiman¹

In Chapter Two I discussed how education in the United States has become dominated by audits and how those audits draw from quantification in accounting for how well predetermined standards are being met. That chapter was about relationships. This chapter considers more the mechanisms of quantification, particularly two types of quantification: the use of quantification to describe and the use of quantification to infer and forecast. In a way, this chapter is about differences between these two uses of quantification, but it is also a chapter about the different ways these two uses are viewed rhetorically and the assumptions that govern them.

It might be tempting to consider descriptive and inferential statistics as one and the same; after all, both are often taught in the same introductory quantitative methods courses or statistics courses. I recognize that they contain similarities and will recognize those similarities as appropriate. However, the importance of this chapter comes because in the rhetoric of educational writing, the acts of description are conflated with the acts of inference. I seek, then, in this chapter to explore that conflation and the assumptions as it relates to educational writing and argumentation.

I recognize in writing this chapter that I cannot jettison my past experiences with quantification. In writing this chapter, I have considered my studies in statistics and in education; I have considered my course instruction in introductory statistics as well as in educational research. This chapter is written from a perspective that includes my role as an observer and instructor. This chapter will consider both quantification and statistics. I will then consider two ways of using

¹ Neil Gaiman, *American Gods: The Tenth Anniversary Edition: A Novel* (HarperCollins, 2011), 285.

statistics: explaining the past through description and stochastically considering the future. I will conclude the chapter with thoughts about how these types of statistics are governed in educational writing, particularly writing that positions evaluations through outcomes. I write this chapter to untangle, but in untangling, I recognize the potential to create new webs for future consideration.

An Experience

For the past four summers, I have had the opportunity to teach introductory undergraduate courses in statistics and probability. I enjoy teaching these courses as an opportunity to continue my study in statistics and its education. I usually expect a range of responses to the question: “Why are you taking this course?” Among the answers I usually hear that the course is a requirement for a major or because it would be useful for graduate school applications; I also hear the statements about liking mathematics and enjoying the chance to see what statistics is about. At the same time, there are always the responses of “Yuck.”

I have often thought about the disgust that is shoveled out when someone thinks about statistics. I have theories about why they might consider statistics to be unbearable or an unpleasant course to take. Whatever the reasons, there is a sense of not wanting to be in the class and only being there because one must. (I have thoughts that some who read this dissertation might be thinking the same thing.) I have summarized this general feeling of disgust as Yuck!

I recognize that Yuck! might not be the word choice of the current population of undergraduates, but the sentiment and emotional evocation suffice. This response is similar to the stepping in something unwanted or the viewing of something considered personally profane. Yet, on the first day of lectures there is a sense that the course is possibly worse than anything found on the street. In writing this chapter I consider a bit more than undergraduate angst toward statistics, but also the feelings of Yuck! gathered from the convoluted nature of arguing with/through statistics.

Quantifications and Statistics

Michigan State University's College of Education has a required course for all Ph.D. students: CEP 932—Quantitative Methods in Educational Research. Although I would love to see someone write about this required course, the purpose in mentioning this course is that among the graduate students it is referred to simply as “statistics.” The American Educational Research Association asks if a study is quantitative, qualitative, mixed-methods, or theoretical. The audience knows that those studies that are labeled quantitative will be immersed in charts, figures, and summary statistics. In the introduction to their book about educational research, Paul Smeyers and Marc Depaepe use quantification interchangeably with statistics when they claim “one has to admit that the kind of research that uses quantitative, i.e. statistical techniques, has gained most prestige in the 20th century.”² It seems that there is no difference between the terms quantification and statistics.

I think there are subtle differences that should be mentioned in this chapter before I consider the differences between descriptive and inferential statistics. I have been cautious in the use of my terms quantification and statistics in this dissertation. This chapter will focus on the use of statistics instead of solely quantification. Consider for a moment some of the differences that exist, taken from two common English reference sources

² Smeyers and Depaepe, “Representation or Hard Evidence? The Use of Statistics in Education and Educational Research,” 1.

Table 2. Comparison of Quantification and Statistics

Source	Quantification	Statistics
Oxford English Dictionary	“The action of quantifying something,” where quantifying is “to measure or determine the quantity of” ³	“In early use, that branch of political science dealing with the collection, classification, and discussion of facts (especially of a numerical kind) bearing on the condition of a state or community. In recent use, the department of study that has for its object the collection and arrangement of numerical facts or data, whether relating to human affairs or to natural phenomena.” ⁴
Wikipedia.com	“In mathematics and empirical sciences, it is the act of counting and measuring that maps human sense observations and experiences into members of some set of numbers.” ⁵	“The study of the collection, organization, analysis, interpretation, and presentation of data. It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments.” ⁶

I consider quantification to be the act of converting entities into some numeric units, which then allows for calibration, comparison and inference to be performed through statistical analysis.

Quantification, for me, is determining the amounts that are then used to describe conditions or to consider through probabilistic models. One of the purposes of statistical analysis is to estimate a parameter that describes a population of interest through the estimation of those parameters through sampled statistics, often called inferential statistics. However, a key to making inferences

³ "quantification, n.". OED Online. September 2012. Oxford University Press.
<http://www.oed.com.proxy2.cl.msu.edu/view/Entry/155915?redirectedFrom=quantification> (accessed November 12, 2012).

⁴ "statistics, n.". OED Online. September 2012. Oxford University Press.
<http://www.oed.com.proxy2.cl.msu.edu/view/Entry/189322?redirectedFrom=statistics> (accessed November 12, 2012).

⁵ “quantification”. Wikipedia.com. <http://en.wikipedia.org/wiki/Quantification> (accessed Nov. 1, 2012).

⁶ “statistics”. Wikipedia.com. <http://en.wikipedia.org/wiki/Statistics> (accessed Nov. 1, 2012).

comes through the use of things like descriptive statistical analysis, such as finding a mean or a proportion. When I teach this concept to my introductory statistics students, I mention that quantification is the process of taking that which may not be numeric and creating counts from that information, while inferential statistics is the process of estimating the population's characteristics through sampling.

Why do I mention this? In educational rhetoric, as I mentioned above, the terms quantification and statistics are often used interchangeably. I do not see this as overly problematic, but I do wish to mention that during this chapter I will be addressing two types of statistical use: descriptive statistics and inferential statistics. In outcomes-based education, the studies require scores measuring performances on testing instruments, and those quantified scores are treated as proxies for education or learning. This is an attempt to quantify, or take something non-numeric and make a countable and mathematical quantity, learning in an attempt to compare and evaluate through the processes of descriptive and inferential statistics.

The remainder of this dissertation considers the rhetorical differences between using descriptive statistics and inferential statistics in educational arguments. In suggesting this framework, it will be necessary to consider the differences between descriptive and inferential statistics in construction and governance. The remainder of this chapter considers these differences. I do this by considering five different ethical problems that arise in the use of quantification:

- (1) Converting human qualities into quantities without altering the qualities,
- (2) Inferring qualities beyond the sample,
- (3) Assuming the future will be like the past, allowing for forecasting future events based on the past conditions and summaries,
- (4) The tendency to conflate probability and certainty, and

(5) The positivistic assumption that humans can be studied with the same technologies and measurements that apply to the natural world.

I recognize these problems as part of the vernacular that uses the term statistics to describe three different things: first, there is the term descriptive statistics, which counts the past and summarizes the past conditions as what is. This type of statistics might allow for simple correlations determining the strength of linear relationships among numerical measures. The second type of statistic is used to generalize from a sample to the desired population. Finally there is the type of statistics that is used to infer from the past to forecast a probabilistic future. All three are seen as statistics, being taught often in the same quantification or statistics courses. Yet, these three types of statistics serve different rhetorical purposes.

Descriptive Quantifications

Converting Qualities into Quantities

Descriptive statistics are not unfamiliar; if I were to look at the exposure to descriptive statistics I would find even in elementary school basic descriptive statistics are taught from an early age, including concepts like having children count how many of their classmates' favorite color is blue or making a pictograph of the numbers of dog owners are in a class. Elementary-aged children are also exposed to such descriptions like finding the mean (average), the median (middle), or the mode (most). However, I believe that descriptive statistics continue well beyond elementary school, particularly in academic research and the writing and rhetoric involved in that research.

Descriptive statistics aim at portraying things as they were, that is ascribing through some numeric analysis the conditions in the past. For example, one type of descriptive statistic might be counting the number of deaths as a result of pneumonia in 1918, the year of the influenza pneumonia pandemic. Or we might consider describing annual incomes in the United States by

describing the average income, which is found through adding incomes and dividing by the number in the sample, or through the use of the median income, the 50th percentile of household incomes.

Education is also concerned with the use of these descriptive statistics. For example, we find that it is common to hear reports that counts (often to be reported in percentages) the racial demographics that compose the school district or schools counting the number of children that qualify for free or reduced lunch. I see in these reports potential problems. Yes, it is possible for individual children to be counted; the difficulty with this is the assumption that the qualities of these children can be converted into measures that accurately describe these qualities. For example, in creating a demographic variable called race, the assumption has already been made to establish an abstraction and classification of individual qualities. Or in considering the counts of children who qualify for free and reduced lunch (which is often seen as a proxy in statistical analysis for Socio-Economic Status) the quantification assumes that qualities and conditions of poverty are able to be abstracted based on an arbitrary formula which establishes a classification for poverty.

I recognize that this is not only an issue relating to quantification. In all fields there is always some form of how to define certain terms and what is to be counted as evidence in the arguments. In the field of anthropology there is of course specific (and purposeful) choices that are made in what counts as evidence and what stories are pieced together to fabricate the findings of the research. Anthropologists might look at abstract notions like gender roles in the society and highlight the shortcomings or strengths of the group of interest. Of course these abstract notions are available for criticism. Some would argue that these criticisms are problematic. I do not offer criticisms of quantification in educational rhetoric to demote these evidentiary sources; instead, I consider this an opportunity to explore and question the role of descriptive quantification in educational arguments.

In an educational system which is founded on auditing standards, rhetorical reliance on descriptive statistics and quantitative reporting (such as reporting the average scores for a school district on a state's standardized achievement tests or the percentages of schools who meet adequate yearly progress as defined by *The No Child Left behind Act*) becomes standard. The appeals of these quantified values is not difficult to see, as these test scores allow for comparisons among schools and make it possible to determine if standards are being met. However, test scores serve as a proxy for qualities related to learning, such as the abilities to compute mathematical formula or comprehend a piece of literature. Thus, whether the description is that of a racial comparison or that of comparing achievement, there are inherent problems of accurately mapping qualities through calibrated measures.

However, another problem arises in the use descriptive statistics: the comparisons of demographic subgroups based on arbitrary descriptors. Consider the often used research input of socio-economic status (SES) when comparing and auditing student performance on standardized tests. In analyzing achievement scores, one of the common comparisons is that of rich versus poor, in attempts to recognize how to come to the rescue of the marginalized poor. Within SES there is an arbitrary definition of what it means to be upper-class or middle-class or in poverty. Some researcher decides what level of household income coupled with education level and number of people in the household determines where to label the individual so that comparisons might be made. Even worse to me is the practice is educational research to suggest some arbitrary measure like students who qualify for free or reduced lunch is an appropriate representative of SES. So, there are arbitrary classifications for who qualifies as a free lunch student and those who are labeled as reduced lunch students representing the arbitrary labels and classifications of which students are labeled poor.

The numbers assume that individuals share common characteristics that can be used to label and classify. The difficulty in this is that these children and youth are individuals, which individual

characteristics and qualities. These people have lived lives and have experiences that require more than simply summarizing them under the guise of a common demographic. It is possible to count (and report) the individuals who are given free lunch, but the ethical constraint ends there. It is upsetting in educational rhetoric to diminish the qualities of humans in an attempt to summarize through assumed commonalities.

In his *Trust in Numbers*, Theodore Porter provides a different perspective of the problems associated with converting qualities into quantities through descriptive statistics. He suggests that the although the motives of early statisticians might have been for improvement, the problem became that the rhetorical use of these descriptions were used to highlight qualities that were in need of intervention and change, often classifying groups of people that the researcher had little desire to associate with personally.

Much, probably most, statistical study of human populations has aimed to improve the condition of working people, children, beggars, criminals, women, or racial and ethnic minorities. The writings, especially private ones, of early social statisticians and pioneers of the social survey exuded benevolence and goodwill. In print, though, they generally adopted the hardheaded rhetoric actuality, which permitted women as well as men to assume the role of the scientific social investigator, and not merely of an agent of charity.⁷

Here Porter is describing the use, particularly the rhetorical use, of quantification particularly descriptive statistics which have been used in changing conditions of those less fortunate, non-normal but times have been in the margin. Converting qualities into quantities became a rhetorical server of offering evidence for change. In describing these qualities, descriptive statistics also provides a sense of actuality and objectivity which is perceived as authoritative. Descriptive statistics also allows for a dismissal of moral closeness in favor for impartial distances. Porter suggests that

⁷ Porter, *Trust in Numbers*, 77.

descriptive statistics were used to describe and investigate members of society “whom they did not know, and often did not care to know, as persons.”⁸ The use of descriptive statistics such as averages or percentages seems to gain favor as a vehicle for describing populations that lacked “strong and interesting personalities.”⁹

However, the issue of this rhetorical deployment is that it is impossible to measure the qualities of these people. However in the social sciences this has become the rhetorical norm to consider human qualities as countable through calibrated measures. This is not to say that there are things in education that are countable, such as the number of desks in the classroom or the number of text books provided or the amount of money used within the schools. These objects are countable without the ethical concerns of diminishing human qualities into quantities that can never accurately portray the qualities being measured.

Education’s rhetorical use of descriptive statistics continues to be used in attempts to describe conditions within schools and classrooms. Later these descriptions are often interpreted as bases for inferences that portray conditions that need to be changed or altered. The deployment of descriptive statistics in education is used as a tool for summarizing the past conditions within schools or summarizing the past assessments of learning and achievement.

Inferential Leaps

Inference refers to the potential to infer about conditions. Infer means to conclude from data. I contend in this chapter that these conclusions can take the form of concluding qualities toward the entire population based on data or conclude about conditions in the future, based on the descriptions of the past. In educational reform writing and research, the purpose of making

⁸ Ibid.

⁹ Ibid.

inferential leaps is to take the described qualities and conclude about others within the population, such as other students in a certain grade, or to conclude that the potentials in the future, such as suggesting that unless schools change structures, based on past test results, the school will not continue to make adequate progress toward having all students be labeled as proficient based on test scores.

The problem of using descriptive statistics in educational research is that it assumes that human qualities can be quantified, making ethical assumptions that the transmutation of qualities into quantities does not diminish the qualities. However, when statistics are used to infer information there are different assumptions and problems that surface beyond the issue of being able to count qualities. In this next section I consider the remaining four issues that were outlined above, suggesting that they relate to the commonly used phrase inferential statistics. In general, the issues of inferential statistics in educational argumentation are connected to the notions of probability, which are lost in the hopes of making statements that are applicable and generalizable beyond the context of the study.

Assuming Qualities beyond the Sample

One of the dangers of writing using quantification is the attempt to generalize qualities to the entire population. Recall that one of the hallmarks of scientific educational research, as outlined by Shavelson, Towne, et al. is that the findings are generalizable to other contexts and situations.¹⁰ This assumption comes from assuming that contexts do not differ and change. This is not to suggest that changes in contexts cannot be informed by research and work. In medicine, for example, it might be beneficial to test smokers for lung cancer although the contexts have changed. The difficulty in this is when rhetorical statements are made that concern the majority or the entirety. Recall the example

¹⁰ Committee on Scientific Principles for Education Research, *Scientific Research in Education*, 71–73.

I provided in Chapter Four from *No Excuses* where the Thernstroms suggest that “every urban school should become a charter.”¹¹ This is a problem of the use of generalization in inferential statistics, as was mentioned in Chapter Four. This might be the result of authors’ misunderstanding quantification, but this problem is prevalent in educational writing and rhetoric. If something works in one condition it would work for all. In this statement, the qualities of charter schools are generalized through data to suggest that charter schools produce better qualities of learning. Thus, the generalization becomes one of shifting attention from the individual conditions and contexts to one of generalizing what should be for all urban settings.

In Thernstroms’ study, there was a sampling of urban schools, both charter and traditional public schools, and the descriptive data suggest that those students who are in charter schools score higher on a standardized test, that is that the outcomes-based education is more effective than the traditional public schools. This descriptive data is used to make a generalization about the conditions of all urban schools. The Thernstroms deploy the descriptions in ways to suggest beyond the sample to all conditions. Let me restate my position in this dissertation. The problem is not found in descriptive or inferential statistics or their use. The problems come from those who use statistics rhetorically without considering the assumptions that produce the statistics and the ethical complexities that are associated with quantifying qualities.

This problem is not unidirectional, meaning that this problem does not only exist by taking the qualities of a sample and applying broadly to the entire population. This problem also exists in taking the general findings and imposing the qualities summarized to the individual. Hacking calls this the “looping effect.” An example from my personal teaching might help in understanding the bidirectionality of the problem of generalization. In my teaching of introductory statistics, I often mentioned to the students that there was a potential in statistics to infer based on the data.

¹¹ Thernstrom and Thernstrom, *No Excuses*, 265.

However, it seems that in the work of learning introductions to statistics there is the potential to conflate forecasting based on a probability with predicting with certainty.

In asking my statistics students to write about findings based on data, the students struggle with the lack of concreteness in the statistical findings; struggling with making statements that are certain to happen instead of statements of being probable. I would hear students make comments about how after finding a linear regression comparing a vehicle's weight to the gas mileage per gallon suggest that a car weighing a certain weight would run an exact number of miles per gallon of gasoline. Or I would see students read a statistical output suggesting a certain medication has a 90 percent chance of improving a medical condition if taken. The struggle in this learning is that these statements are statements of probability based conditionally on the contexts. There is no guarantee that a car weighing so much will get a forecasted miles per gallon or that a given medication will work. Here the responses are seeking to generalize beyond what conditions are known and especially beyond those that are not to a specific individual or case.

This is always a sticky situation for me as I consider statistics through probability. It is easy for the students in these introductory classes to learn the lingo, such as "I am 95 percent confident that the true population parameter (say, the mean) is somewhere within the given interval," but I find that the students in these courses struggle with what is meant by these phrases and under what assumptions these statements are made. In part, I wonder if this is a holdover from prior mathematics classes, such as arithmetic and algebra, where solutions to equations are usually exactly found through repeatable algorithms and processes. Yet, there are applications of algorithms and processes in the work of statistical/quantitative analysis but with different implications in the findings. I feel that these students have a better feel for what I will call descriptive statistics, which simply put are found to describe the past, because of the certainty based on collected data.

Descriptive statistics, erroneously, becomes an educational rhetorical tool for applying qualities to the entire population and to the particular individual. The ethical problem with this comes through limiting the individual nature of humans by suggesting that there is something called a representative sample, a sample that is able to represent the qualities of the population as well as the qualities of the individual. The statistician recognizes that there is really no way to test and know positively if the sample represents the desired population. The statistician recognizes that random sampling is a technology that enables such representation of the population. This careful consideration of the statistician is not found in educational discussions, particularly when claims are made beyond the same contexts.

I contend that this is not possible when considering human issues. For example, what makes the sample representative? In statistical jargon there would be thoughts about representing the populations' demographics, such as the proportion of a certain race or the proportion of males and females or representing socio-economic status in the sampling. In design theory, this concept might be worded as suggesting that if the population of interest contains males and females, then the sample should also contain males and females (with more credibility to those studies who stratify to represent the proportions of males and females). If the population of interest were only male, then the sample should only select from that desired population.

The concept of representative sampling then suggests that the desirable qualities (as desired by the researcher) have been effectively accounted for in the sampling, and that the opinions and reactions of the general population would be similar because humans that share these desirable qualities are going to behave similarly. However, the problems come in assuming that just because a human has a certain chromosome combination or a certain standard of living that the general human will behave in that way. Just because I am male does not mean that the findings of a study

can be generalized to me as an individual or to the entire subpopulation of males. I am an individual with abilities to behave and respond and react. I am capable of thinking and determining directions.

Assuming the Future will be Like the Past

Next, I turn to the problem of the assumption that the future will be like the past. This problem is based in the ideas that history repeats itself, suggesting that because humans behaved one way in the past that they will behave that way in the future. This assumes that there is a core of what humans are and that they will behave the same in all conditions, such as described through the philosophical concepts of progress. Determinism would suggest that we as humans are condemned to repeat history, no matter the choices or current historical interactions. Essentialism would suggest that within the group being studied there are necessary attributes required to function, assuming that the core attributes demonstrated in the past would be present in the future. Statistical inference about the future might be based on any of these foundations. I believe the continuity assumption is usually just out of habit.

The ideas of inferential statistics suggest that in likelihood characteristics can be applied beyond what has occurred in the past to what might occur in different situations and times. Inferences begin with descriptive statistics, such as in counting the number of people with influenza. However, inferential statistics take a leap of generalization away from describing the counted past to a speculate about the number of people that might contract influenza in the future. This speculation—or inference—is based on the assumption that the future will be like the past.

Why might political scientists read the works of Thomas Jefferson or Abraham Lincoln if the present or future is not the like the past? This is not to say that ideas and constructions cannot be informed by the thoughts of others. It might be possible to not look for insights from others. But in considering the text or experiences or ideas or contributions of others shapes our understanding

of the world and the desires of the human. Although the past is not like the future or the future like the past, political theorists read Jefferson or Lincoln to begin, as Gayatri Spivak notes, the “rearrangement of desires.”¹²

Inferences build off other descriptive statistics and their probability models to create these generalizations. Included in this process are assumptions: assumptions about the nature of time, such as the suggestion of historical continuity—the assumptions that the future will be like the past—to generalize a predictable future; assumptions about continuity among contexts—suggesting that conditions and contexts are transferable. In the influenza example the inference comes through predicting some number of deaths based on the data collected and applying a probability model to create an upper and lower bound of potential influenza deaths in the upcoming year—as if we could be sure that next year will be the same as last year. Inferential statistics are not certainties. Instead they are probabilistic potentials of what is likely to occur assuming the future is like the past.

As an educationalist, I notice that educational arguments deploy both descriptive statistics and inferential statistics, although often not distinguishing between the two, except possibly with a different section title or methods subheading. For example educational leaders in Michigan who are concerned with the number of schools making Adequate Yearly Progress (AYP) toward having all students being measured as proficient on the state's assessment, the MEAP, might first present descriptive statistics about comparative test scores. Then they may infer from those descriptions to try to predict how students might improve within a grade band and longitudinally. That is, for example, looking at predicting how fourth-grade students might improve year-to-year and how students would improve as they progress through their educational careers would be to assume that next year is the same as last year, and students will behave next year the same as they behaved last year. Through selecting representative samples these educational leaders might attempt to predict

¹² Gayatri Chakravorty Spivak, *Other Asias* (Malden, MA: Blackwell Publishing, 2008), 3.

the proficiency rates based on applications of the normal probability model. Educational leaders may also use inferential statistics to forecast such things like the average test scores based on random sampling. In both cases, inferences are based on the assumption of continuity: that next year will be the same as last year.

Ian Hacking suggests that descriptive statistics of the nineteenth century were used for philanthropic work, describing worthy interests of improving the laboring classes. Hacking mentioned that Quetelet and others “thought that they could do so by exercising a new kind of control. Discover what are the statistical laws that govern crime, disease, vice, unrest. Then find ways to alter the conditions.”¹³ Descriptive statistics historically then could be seen as a tool for changing conditions that were viewed with social disdain. In this example offered by Hacking I see how descriptive statistics might be used to make inferential leaps in suggesting changes in social conditions. I suggest that this example also demonstrates the assumptions of these philanthropic statisticians that the future will be like the past, that the conditions of the crimes, vices, and diseases will repeat based on the descriptions of the past and the conditions of the future being based on these past descriptions.

This use of quantification assumes that the conditions from the past will be the same as the conditions of the future, allowing for a prediction of what the future holds. Based on the experiences and descriptions of the past, the future repeats what has happened, unless something calamitous happens as an intervention. This assumption denies the complexities of a “set of relations that delineates sites which are irreducible to one another and absolutely not superimposable on one another.”¹⁴

¹³ Hacking, *The Taming of Chance*, 118.

¹⁴ Michel Foucault, “Of Other Spaces: Heterotopias,” trans. Jay Miskowiec, *Architecture / Mouvement / Continuité* (1984), <http://foucault.info/documents/heteroTopia/foucault.heteroTopia.en.html>.

Conflating Probability and Certainty

When sampling is done and inferential statistics are analyzed, the results create a type of statement based on the probabilities associated with a certain model, such as the Normal distribution. These statements are bound to the probability model from which they are generated. When writing about quantified finding the sample must meet the required conditions for the probability model to hold. Only when the conditions of a probability model are met can the inference be made. However, within meeting these conditions, the inference becomes one of likelihood not one of certainty.

I find that illustrating this concept with a probability example helps to demonstrate how these forecasts are bound by the conditions of the probability models on which they are based. This example takes descriptive data, data that could be used to describe the conditions of the past, to predict behaviors in the future. I recognize that this example is not about humans, instead I provide an example of a truly random object—a game of roulette. In choosing a non-human example, I provide a case where the forecast would be recognized as being a probability in an attempt to foreshadow the difficulties of trying to measure humans through the natural world, a discussion held in the next section.

I recognize that there is a way to consider the gambling scenario through exact probabilities—one that would not require the use of statistical inference. Those who have a greater understanding of probability and randomness would recognize in this example that the probability of getting red is constant. However, I provide the example as a layperson might consider going into a gaming/gambling scenario. It always amazes me to talk to my statistics students about gaming and the naïve thoughts that are had about how a gambler is “due” for a win. I provide this example to

demonstrate ways that observational inferences might be tools in confusing what is probable with what is certain.

Suppose a gambler goes to into a casino to play roulette. She takes an opportunity to watch several rounds of the game, say ten, and observes that most of the spins landed on a red-colored number, again say eight or nine. Yes, this gambler has collected evidence to report conditions of what has happened in the past, what I have been calling descriptive statistics. It might be tempting to suggest that the wheel might be biased toward red, based on those descriptive statistics. However, should the gambler assume that the next spin will also result in a red, thus suggesting that the gambler should place her bet on red?

The novice might read the descriptive data as suggesting that the bet should be placed on red. However, the difficulty in this is that the spinning of a roulette wheel is a random occurrence where there is a potential for a different color to be spun, each time. Based on probability, the chance of the ball landing red would be 16 out of 38 (if using the United States' standard of having a 0 and 00). Because the gambler collected data, it is tempting for her to suggest there is a certainty that the ball will land on red. However, the prior landings of the ball do not influence the future landing of the ball. There is still a 16 out of 38 chance that the ball will land red, but it would be regarded solely as a chance not a certainty. Even if the wheel were biased toward red, there is still probability or potential for the ball to land in a different colored space. I am not suggesting that the gambler in this example will lose because of the bet on red; at the same time I am not suggesting the gambler will win because of the bet on red. I do not know what will occur and thus can offer only a probabilistic guess.

Now some may suggest that this example is faulty because there were only ten trials, obviously not enough evidence to show the whole picture of chance occurrences. Those familiar with the concept of The Central Limit Theorem might suggest that I do not have enough trials in

this example to invoke the normal probability distribution. I would agree. If for example, the gambler had sat and watched the game of roulette for 1000 turns or one million turns and still found 90 percent of the results to be red, the inference still is only a probability and not a certainty. Yes, we would be more confident that the wheel was biased red. I recognize this; however, this does make the conclusion of the spin a certainty. As a student of probability the suggestion of certainty based on this data or any data seems inappropriate. However, I note that it is far too common for interpretation of statistical data to be interpretation of certainties instead of interpretation of probabilities.

I have mentioned throughout this dissertation that the application of the measuring natural phenomena for human qualities was influenced by some of the work of Belgian astronomer Adolphe Quetelet. I mention again that Quetelet suggested that there was some form of natural law associated with the use of quantification and its applications toward measuring human qualities. This shift suggests to me that there is a tendency in deploying quantification rhetorically to make the findings some form of law that must be obeyed. This is a problem in educational reform writing and research. There is a tendency in writing about educational reform to suggest that if changes are not made, based on the data that provided, the conditions of education will not improve. The difference becomes a rhetorical deployment of certainties based on quantified data as opposed to statements of possibility.

This is a dangerous rhetorical conflation. The problem assumes not only that the future is like the past and that human agency has no relation in those forecasts, but it also assumes that the probability models that are foundational for such inference are not random, instead overlooking assumed aspects (such as randomness and independence) in a search for solutions to social problems. I am reminded of the rhetoric of the jeremiad described in the last chapter. The American jeremiad assumes that there must be some change in order to receive the promised blessings. The

rhetoric loose influence if the promised changes are hedged as statements of probability. The suggestion that it is likely that schools that do not change will not improve in educating youth is far less impactful in the general eyes than making it a statement of certainty. It is far too common in the use of quantified discourse to see interpretations consist of words like “must” or “will.” This is not the intent of inferential statistics. Statisticians recognize the assumptions of probability that abound in making statements about the future conditions and qualities.

Positivist Views that Humans can be Studied Like the Natural World

Inferential statistics then become a way for us to predict into the future with some sense of confidence, a probability based on assumptions of historical continuity and humanistic essentialism. When I taught this concept in my statistics courses a common misconception was that just because the data suggest something *might* occur does not mean that it *must* occur. It is possible although not necessarily likely, that something unpredicted might occur, such as when a weather report suggests an 80 percent chance of rain. The forecasting of rain is based on certain conditions and interactions that produce a likelihood of producing rain. In this, a weather forecast(er) considers the complex interactions of temperatures, humidity levels, air pressure, wind speeds, etc. to predict the chances of rain. To borrow from the frequentist notions of statistical inference, this concept of 80 percent chance of rain suggests that given the same conditions (as listed above) an infinite number of days, there would be rain on 80 percent of the days. A Bayesian would suggest that with informed *a priori* opinions, the probability given the conditions would be 80 percent. I am not writing about the difference between these two statistical inferential systems, but suggest that in both there is some sense of consistency in condition, whether that is through an infinite number of such conditions or that based on the prior information, the conditions suggest such that likelihood of rain. In either case, the coming of rain is not guaranteed, but instead is a leap into generalizing based on the

conditions at hand and the continuity of historical and contextual influences. But meteorology is not a social science, and inferential statistics mean something altogether different when they try to predict human behavior instead of weather patterns.

It would be nice for us to be able to predict the future with a certainty that is not contingent on the meeting of prescribed conditions and contexts, particularly for those who invest in the stock market; however inference through statistical data is not and never can be certain. Inferential statistics then become a way for us to predict into the future with some sense of confidence, a probabilistic confidence based on assumptions of historical continuity. When I taught this concept in my statistics courses a common misconception is that just because the data suggest something might occur does not mean that it must occur. It is possible although not necessarily likely, that something unpredicted might occur (such as when a weather report suggests an 80 percent chance of rain). It would be nice for us to be able to predict the future with certainty, particularly for those who invest in the stock market, however inference through statistical data is not certain.

If the meteorologist suggests an 80 percent chance of rain, does this guarantee that there will be no rain? Does this also mean that there will be great chance of rain for a neighboring county or state or nation? Does this mean that if there is rain that the meteorologist was incorrect in reading and presenting the data? I purposefully use the example of the meteorologist because of the uncertainty of the weather and the acceptance of rain even when rain is deemed unlikely through the models. I do not believe it is uncommon for the weather report to suggest a low chance of rain and to in reality rain. However, this example also provides an important consideration about the importance of inferring from similar contexts and situations. The chances of a neighboring county or state experiencing rain would not be determined from the data reported by the meteorologist. This seems understandable as the conditions and contexts of the neighboring states are different; the barometric conditions will change, the elevations may be different, the influences of different

metropolitan areas will be felt. It almost seems absurd for us to suggest that simply because it there is an 80 percent chance of rain in Michigan that Ohio will have the same forecast.

Weather forecasts are built on models that are built around factors, such as the temperature, the barometric pressure, the humidity, the time of year, wind speed and direction, etc. These factors interact to generate a model. Forecasting means that, given these precise environmental factors, it has rained eight out of ten times. However, when one of the factors changes, the model is no longer as reliable in forecasting the future conditions. Either a different model is needed or a new model must be created. These models rely on weather patterns, behaviors of the weather that are predictable and quantifiable. However, there arises a problem in using this theory in describing human behaviors. The weather does have times when it behaves unpredictably, but the unpredictable behavior of weather is not a result of agency and thoughtful decision, as is the case with humans. Human behaviors are not random instead the human behaviors are results of agency, relationships, and qualities. There is an assumption that human behaviors are consistent. When we try to predict human behavior using inferential statistics, we forget that people learn, change, create, grow, and innovate such that my actions in the past cannot predict my actions in the future.

I remember reading about an experience of R. A. Fisher who went to a tea party and was introduced to a woman who claimed she could tell whether milk had been poured into tea or whether tea had been poured into milk. Of course this causes some clamor at the tea party as many suggested that it was impossible to distinguish the order. Fisher carefully constructed a type of experiment where the woman was offered ten cups of tea prepared through different infusions. Fisher drew from this experience while writing his *Designs of Experiments*; I however relate this experience as Fisher developed this experiment based on probability models which would allow him to predict how the woman tasting tea might respond to future tea infusion samples. This is not to suggest that Fisher's model would guarantee correct predictions every time, but the probability

suggests what is likely, alluding to the problem mentioned earlier of claiming probabilities as certainties. However, this is an example of using inferences for predicting human behaviors, suggesting that the woman's qualities of tasting tea could be measured like the natural world, like measuring the wheat yield of a field planted in Fertilizer A compared to the yield of wheat in the field using Fertilizer B.

The Problems of Predicting and Forecasting

Inherent in this work of forecasting and inferring from data is *risk* and *uncertainty*. Risk, according to Nate Silver, “is something that you can put a price on. Say that you’ll win a poker hand unless your opponent draws to an inside straight: the chances of that happening are *exactly* 1 chance in 11 (in a Texas Hold ’em game with one card to come).”¹⁵ Whereas uncertainty “is risk that is hard to measure. You might have some vague awareness of the demons lurking out there. You might even be acutely concerned about them. But you have no real idea how many of them are or when they might strike.”¹⁶ Put another way, uncertainty confronts complexities and external influences, whereas risk can be probabilistically measured. Inferring through risk, those inferences and forecasts which are based on knowledge of potential outcomes, can be beneficial for growth (particularly monetary growth) however there is calculated chance of failure.

It is in the uncertain that forecasting can be detrimental, such as forecasting about the housing market during the last 2000s. The social sciences are not dealing with risk (unknowns that can be calculated, such as calculating the cards needed to beat an opponent in a game of chance), instead dealing with uncertainty (with its complex relations of power, signification, and production). Humans are not mechanisms, which includes the prior knowledge of potential outcomes. As I have

¹⁵ Silver, *The Signal and the Noise*, 29 emphasis in original.

¹⁶ Ibid.

mentioned, humans have abilities to choose, suggesting that they are not bound to a set of potential outcomes, such as the potential outcomes of flipping a coin.

John Dewey classified modernity as “the quest for certainty.”¹⁷ It might be tempting to say that there is trust in forecasts from inferential data because, as Silver suggests, “we abhor uncertainty, even when it is an irreducible part of the problem we are trying to solve.”¹⁸ If we believe Silver’s comments, then the assumptions of not liking uncertainty drive an acceptance of what is being forecast as a way out of the uncertain. The governance of inferential statistics comes through a willingness to believe forecasts as certainties as technologies that challenge the nebulae of uncertainty, suggesting that chance has been tamed, as Hacking contends, so that we believe that through quantification we are measuring risk instead of the hard to measure uncertainty. This is problematic to me. Forecasts are probabilistic statements of events that might occur given certain conditions and constraints. They are not certainties, although there are those who wish they were. I recognize that this fear of uncertainty is a modern preference for seeing and understanding the world. I mention this here as a connection to the episteme of some of the educational research world, particularly those who suggest that educational issues might be understood through empirical lenses.¹⁹

I contend that with human beings—the demons that lurk—uncertainty cannot be ignored as we consider the mis/use of forecasting and inferring. Human beings do not behave according to Newtonian laws of motion. However, as inferential statistics continue to be used in the quest for certainty, there is potential for the probability to become assumed certainty, suggesting that

¹⁷ Dewey, *The Quest for Certainty*.

¹⁸ Silver, *The Signal and the Noise*, 20.

¹⁹ Consider Committee on Scientific Principles for Education Research, *Scientific Research in Education*.

environments of social interaction, such as with education, are becoming more a world of risk, where the outcomes can be seen to be computed with exactness.²⁰

Thus inferring from data can be useful (and accurate). However, there are some cautions that would be warranted

- When quantification is used in generalizations, the individual qualities of humans are conflated with the abilities to apply general qualities to populations beyond the sample.
- When quantification is used to forecast with certainties instead of recognizing the probabilistic nature, human qualities are diminished by treating them as mechanic objects.
- When quantification is used to forecast and predict, there is an assumption that the future will be like the past.
- Forecasts may not be correct.
- When quantification is used the qualities of the contexts and conditions are not transferable beyond the current contexts.
- In the current outcomes-based evaluation of education, a final thought may be beneficial “We need to stop, and admit it: we have a prediction problem. We love to predict things- and we aren’t very good at it.”²¹

Governance in Education

The sociologist Nikolas Rose suggests that numbers are linked to governance and democracy. From his *Powers of Freedom*, he suggests “numbers have achieved an unmistakable

²⁰ See Porter, *Trust in Numbers*, 209–213.

²¹ Silver, *The Signal and the Noise*, 13.

political power within technologies of government.”²² Here he outlines four types of political numbers. Although the four types of political numbers are of general interest to me, this chapter draws specifically from his third classification and its application to educational argumentation.

Numbers make modern modes of government both possible and judgeable. Possible, because they help make up the object domains upon which government is required to operate. They map the boundaries and the internal characteristics of the spaces of population, economy and society. And other locales—the organization, the hospital, the university, the factory and so on—are made intelligible, calculable and practicable through representations that are, at least in part, numerical. Judgeable, because rates, tables, graphs, trends, numerical comparisons have become essential to the *critical scrutiny of authority* in contemporary society. Liberal political thought has long been characterized by skeptical vigilance over government. This vigilance is increasingly conducted in the language of numbers.²³

I understand Rose’s third point of the role(s) of numbers within the technologies of government as a delicate balance between the being able to operate and being able to judge those operations; I would suggest an alternative to judgment, that of being able to audit those operations. Within the judgeable qualities of numbers is the ability to have scrutiny and accountability toward set (often *a priori*) standards and characteristics. Numbers make it possible for modern government through the establishing of what might be governed, and the characteristics through which governments are deployed. Ultimately, these deployments within the technologies of governance tie to the concepts of the audit culture that were discussed in Chapter Two of this dissertation.

²² Nikolas S. Rose, *Powers of Freedom: Reframing Political Thought* (Cambridge University Press, 1999), 197.

²³ Ibid. emphasis mine.

How does this apply to writing about educational reform? Rose is speaking about how quantifications are used in operations and evaluating those operations. Educational reform rhetoric is concerned with the operations of education, particularly the operations within educational policy and educational practices. Texts about educational reform, such as the three texts analyzed for this dissertation, are concerned with changing the operation of education, trying to make education good. Educational reform rhetoric has developed a trust in quantification as quantification makes it possible to evaluate and judge the effectiveness of the current educational systems and challenge the authority and credibility of the current policies and practices. The rhetorical use of quantification becomes valued as quantities are a representation of what good education should look like, although in Chapter Six I argue against this proxy relationship between quantification and good education. It is in this quantifying culture in education that we find a form of “skeptical vigilance.”²⁴

It seems that descriptive statistics and inferential statistics are sometimes conflated as being the same. I partially wonder if that is because they share a word (statistics) in their name. They are neither the same nor share the same purpose. Descriptive statistics describe phenomena of the past, describing through countable data conditions as they were; inferential statistics apply probability models to abstract uncountable qualities as quantities in an attempt to generalize what is likely to occur based on given conditions and constraints. I do recognize that both descriptive statistics and inferential statistics use similar technologies, such as the use of visual displays or the uses of summary statistics, such as the mean. However, it is how these technologies are deployed and the ultimate end purpose that makes the distinction important.

To me the rhetorical deployment of these two statistics hinges on a type of temporality: describing the events or conditions of the past through descriptive statistics and forecasting the events or conditions of the future based on the assumed continuity of future events based on the

²⁴ Ibid.

events of the past and the assumptions that humans are truly random, such as with the flipping of coin. Rhetorically, then, the descriptive statistics serve as evidence to describe conditions “as they were.” This deployment focuses on questions that wish to explore through certain conditions constrained by certain definitions, such as defining through demographics. Thus, as Rose suggests, those areas like the hospital or the psychiatric ward become manageable and descriptions become simpler through the summaries of numbers. This type of work does have merit for understanding our circumstances in systematic and careful ways; however, descriptive statistics are one thing, while inferential statistics are something else altogether.

One of the struggles of working with descriptive statistics is the implying beyond the conditions of the summary, which is to suggest that the descriptions and summaries generalize for cases outside of the given conditions, contexts, and relationships. Not only is this true for forecasting future conditions or events but also when these descriptions are generalized across historical conditions, allowing them to be taken as generalized truth of the conditions at the time or in the demographic. Such an example would be when texts consider test scores for minority children and only use data from a specific city, say New York City or Detroit and imply based on these findings that conditions for all minority children in the United States are reflected in this summary.

This chapter has explored common rhetorical uses of rhetorical quantification in educational reform. It has become a common argumentative construction to suggest that human qualities can be converted to quantities through calibrated measurement without losing the nature of qualities. It has become common to suggest that there are characteristics that can be used to describe subgroups within a population, often called demographics. These characteristics are often seen as generalizable to those within the same demographic or characteristic.

In my experiences with educational reform rhetoric, I find suggestions of quantitative summaries by demographic applying beyond the sample to those who share similar traits, such as

suggesting that all students of color are academically lower achieving than White or Asian youth, based solely on the measures of the National Assessment of Educational Progress (NAEP) and the determination of a person's racial classification. It is common in educational reform rhetoric to suggest that the future will be like the past, suggesting that the conditions and the contexts will repeat. It is common in reformation rhetoric to suggest that the quantified data are inferring a certainty in the future if reforms are not made instead of offering statements of likelihood. Finally, there is the common assumption in educational reform that humans can be measured like the natural world, assuming that human behaviors and qualities are predictable and measureable instead of recognizing that humans exhibit agency and choice that cannot be completely known beforehand. Humans are not machines. They interact and react to complex relationships beyond what can be foreknown.

Yuck! This chapter has considered some of the confluences that exist in the use of statistics in writing and research. I have considered how concepts, such as quantification and statistics or descriptive statistics and inferential statistics, can be convoluted and governed as the same. This results in situations where assumptions about distinctions are lost. This chapter has been written to explore some of the assumptions that exist in these convolutions.

In previous chapters, this convolution of descriptions and inferences was explored. In particular, the dissertation shifted from exploring the concepts of auditing and quantification to exploring more the rhetorics of description and inference involved in three educational texts. This chapter considered some of the assumptions involved in the convolution of descriptive and inferential statistics. In the chapter that follows, I will continue to explore how quantifications have played a role in the governance of children and youth in determining the *a priori* educational goal

steering and how they have, as Rose suggests, “map[ped] the boundaries and the internal characteristics of the [educational] spaces.”²⁵

²⁵ Ibid.

CHAPTER SIX

WHAT KINDS OF EVIDENCE FOR WHAT KINDS OF REFORMS?

Speech leaves no mark in space; like gesture, it exists in its immediate context and can reappear only in another's voice, another's body, even if that other is the same speaker transformed by history. But writing contaminates; writing leaves its trace, a trace beyond the life of the body. Thus, while speech gains authenticity, writing promises immortality, or at least the immortality of the material world in contrast to the mortality of the body.

- Susan Stewart¹

This dissertation has been about rhetorical uses of quantification, questioning educational reform texts' uses of quantification because of the dependence to quantify qualities. In writing this dissertation, I have considered quantification as a type of evidence that is used for argumentation in educational texts. I recognize that there are huge discussions in educational research circles about the issues of evidence and methods and structures. This dissertation faced those debates but not to engage them head-on. As I have stated throughout this work, I believe that there are questions that are appropriately considered through quantitative work, meeting the assumptions that allow for probability models to be applied to the analysis and deriving results that inform educational practice, knowledge, or theory. This is not the problem of this dissertation.

The problem for this dissertation has been that in the educational reform discourse, a common rhetorical tool is in play, the use of quantifying to describe conditions, allowing for rhetorics of comparison, transparency, and jeremiad, and the use of quantification for generalization, based on the assumptions of inferential statistics which are conditioned on probabilities and not accepted as certainties. The discourse of educational reform has taken up their charges of and for change through collecting numerical data and considering the need and benefit of given reforms through the lens of the quantified.

¹ Susan Stewart, *On Longing: Narratives of the Miniature, the Gigantic, the Souvenir, the Collection* (Durham, NC: Duke University Press, 2005), 31.

I have suggested in Chapter Two of this dissertation that this is a method of goal steering education. Goal steering suggests that with predetermined standards and measures of those standards, which include auditing to ensure that those standards are being met, the purposes of education are being led in directions that have been defined before teachers and students enter the classrooms. The rhetorical trope of quantification is a tool used within this discourse to take unknown qualities of educational contexts, educational learning, and educational practice into currently recognized numerical summaries and inferences. Thus, as Chapter Three suggested, the rhetorical use of quantification is used in taking the countable past and inferring qualities into a probable future.

Chapters three and four considered the deployment of quantification in three educational reform texts, drawing from texts that were written for a more general audience to consider how this rhetorical use of quantification is deployed for those who may or may not have more than basis quantitative literacy. The assumptions of educational researchers are that there has been exposure to quantification and thus having a better understanding and appreciation of quantification in argumentation. I was not concerned with this audience, and as such chose works that were written for audiences who did not have an assumed quantitative literacy level.

I have struggled with how to write a conclusion for this dissertation, never having written something longer than a journal article and never having tried to combine several into one larger work. It would be appropriate to restate the claims that have been made up to this point, reinforcing the need for quantification and how it is deployed. I have considered that to be a possible path for this conclusion. However, I chose to conclude by raising further questions for inquiry, questions asking what types of evidence are appropriate for what type of reforms.

An Age of Educational Reform

I have contended throughout this dissertation that we live in an age of educational reform, an age where education has been challenged and changed in attempts to “better” educational achievement and attainment. In order to change to “a better,” there is a comparison in qualities, and in the case of educational reform and achievement, this comparison requires calibrated measurements that then translate qualities into quantities, even to the point of creating the assumption that qualities can be measured. In classical philosophy, qualities are by definition immeasurable. However, in living in a world of educational reform, there is space for educational argumentation, attempting to change perceptions about how education should be considered and changed.

In Chapter One of this dissertation, I began considering Michael Leff’s comment that rhetoric exists where judgments are to be made about common concerns. I stated then that educational reform texts are such a space, a place for rhetoric to invite judgments about the conditions of education in hopes to influence opinion. Although this is not the purpose of all educational texts, and not the purpose of all educational reform texts, I suggest that the arguments involved in educational reform texts for general consumption are precisely designed to change the public’s judgments about education.

I purposefully selected three texts to analyze which offer changes in the educational system, whether the change concerns public and private schools, racial inequities, or competing in the global market, the three texts argued not only for change but suggested key changes that should be made. In making this argument, the roles of quantification were explored as a rhetorical trope in transmuting the complexities of education into understandable summaries and generalizations based on some quantified inferences. These texts had an agenda and sought to implement the changes of that agenda through different rhetorics, which for the purposes of this dissertation were broken

down into descriptive quantitative rhetoric and generalized quantitative rhetoric. These are not the only rhetorics that were deployed in the texts to suggest change is necessary. However, the use of numbers and quantified data were important parts of understanding how arguments are being made to persuade people how to conduct educational reforms.

Educational reform is about changing the conditions of education. The texts that were selected for analysis for this dissertation wish to change the future based on the measurements of the past. I appreciate Thomas Popkewitz's consideration of how educational reform is about cultural changes. He suggests

The age of reform is a historical trajectory of things of difference in the cultural theses about modes of living. The inclusive dream in planning lifelong learning, the learning society, or the information society in the contemporary landscape of school reform is not produced through the same assemblies, connections, and disconnections that ordered American progressive education and its sciences of the child. Nor are those past and present reform programs merely to find effective paths to a utopian future. What is taken as natural and sacred in the commonsense of reform are particular internments and enclosures. Theories of learning, development, community, and problem solving articulate double gestures as comparative principles about the honored feats of the noble with the fears of the threats and dangers to the civilized future.²

Educational reform suggests a difference in the ways that education is performed, or as Popkewitz suggests, the modes of living. These reforms are not to find educational paradise. Instead, what is assumed as part of normalcies of educational reform, which I have contended in this dissertation

² Thomas S. Popkewitz, *Cosmopolitanism and the Age of School Reform: Science, Education, and Making Society by Making the Child* (Routledge, 2008), 172.

have been the rhetorical use of quantification, are in actuality constraining the potential for future growth and understanding.

The title for this chapter alludes to a chapter in Patti Lather's *Engaging Science Policy: From the Side of the Messy*. In her chapter "What kind of science for what kind of policy?" Lather considers a specific type of policy work that "interrupts a state of instrumentalism where policy researchers are situated as the handmaidens to the state and/or entrepreneurs."³ In this world of instrumentalism, the research is put in a position to be a servant to those holding funding, such as the funding available from national research organizations or foundations. Her work suggests a way to interrupt this sense of being a servant to the research questions these groups want to have investigated using the methods they desire.

Thus, I ask: "What type of evidence produces what kind of policy?" and "What kind of policy produces what kind of evidence?" This dissertation, for the most part, has considered a certain type of evidence used in argumentation of education reform. It seems to me that there is a sense and desire for objectivity in education to produce supposedly unbiased reforms that would further the cause of educational equality and equity. Objectivity is a word that connotes freedom from the constraints of personal beliefs and subjectivity; objectivity is supposed to be fair for everyone, not privileging any particular group. In this age of educational reform, the desire for a rhetoric composed of quantification takes root in the desires for knowledge to be objective. I believe that this rhetorical deployment of quantification is political in nature.⁴

In this desire to deploy quantification, there is a sense of speaking in a rhetoric that is freed from the constraints of social and moral subjectivity. In attempting to speak through these

³ Lather, *Engaging Science Policy*, 73.

⁴ See Rose, *Powers of Freedom*; Porter, *Trust in Numbers*; Theodore Porter, *The Rise of Statistical Thinking: 1820 – 1900* (Princeton, NJ: Princeton University Press, 1988).

quantified rhetorics, there is a sense of protection in becoming autonomous, or being able to speak for oneself. Porter suggests

When academics take up a branch of practical quantification, they commonly complain that their predecessors were moralists and lacked objectivity. This, for example, is the way sociologists have customarily interpreted the early history of the social survey, arguing that disciplinary autonomy is needed to attain a proper state of objectivity. The converse may be more nearly true: The Weberian language of objectivity was adopted in part as a defense of the incipient discipline against political interventions. Moving away from a descriptive, empirical style and using ever more recondite quantitative techniques brings similar advantages.⁵

Porter is suggesting that many rhetorical and epistemological styles have come to quantification because of the assumed advantages of becoming autonomous, which is becoming free from the constraints of human subjectivity. I contend that this is true for the writings and rhetoric of educational reform texts, especially as educational reform texts have created in this current historical and rhetorical condition a value in quantified arguments, shifting rhetorical attention away from ethical arguments.

Measuring What?

In my preparations for this dissertation, I read a work by Dutch educational theorist Gert Biesta that asked the question: Do we value what we measure or do we measure what we value?⁶ In this age of educational reform that is informed by quantified comparisons of human qualities, I take Biesta's question as important in the conclusion of this dissertation. If educational reform continues

⁵ Porter, *Trust in Numbers*, 199.

⁶ Biesta, *Good Education*, 12.

to use quantified rhetoric to inform arguments about conditions of education along some demographic line, such as race or charter vs. public schools or international comparisons, then the question of what we are measuring and what we are valuing is deeply important to consider. I might rephrase this paragraph to suggest that if educational reform rhetoric insists on using quantification for evidence, is the quantification really measuring what is desired or have the arguments of educational reform adapted to value what is measured.

Biesta highlights two problems of basing argument and shaping educational practices and perceptions solely on the quantified data. First, he suggests that “when we engage in decisions about the direction of education we always and necessarily have to make *value* judgments—judgments about what is educationally *desirable*.”⁷ In making value judgments, educational rhetoric considers what is beneficial in education and beneficial to the particular points of the authors, again suggesting that educational reform is not autonomous of political influences. However, the second point is a validity of measurements. Biesta continues by suggesting

More than just the question of the *technical validity* of our measurements—i.e., the question whether we are measuring what we intended to measure—the problem lies in what I suggest is referred to as the *normative validity* of our measurements. This has to do with the question whether we are indeed measuring what we value, or whether we are just measuring what we can easily measure and thus end up valuing what we (can) measure.⁸

This second problem, then, is not only measuring accurately, but also considers whether the measurement is what we desire in the first place. In the sense of rhetorical quantification, do we place value on the rhetorical evidence being deployed or is that evidence being deployed because we can easily measure it.

⁷ Ibid.

⁸ Ibid., 13.

Although I consider these two problems appropriate in the consideration of quantification as a rhetorical trope, I further suggest a problem of possibility. As I have mentioned throughout this dissertation, there is a problem of quantifying qualities, often creating a numeric proxy for those qualities. I argue here that the difficulty in educational reform rhetoric that relies on quantification is that quantification relies on measuring qualities to evaluate how well educational reforms are performing in what we value in education. However, measuring what we value is difficult, if not impossible, suggesting that educational reform rhetoric places value on what can be measured instead of what is ethically and morally of value.

Let me return to the three objects of analysis for this dissertation relating them to these three problems with using quantification within educational argument. If I were to ask the question “What is education for?” of this text, what would I get?⁹ This question is a common theme throughout Biesta’s *Good Education in an Age of Measurement* and serves nicely as a consideration for the conclusion of this dissertation.

First, consider *The Global Achievement Gap*. In this text the comparisons are between schools that prepare students for work within a global economy. What is education for concerning *The Global Achievement Gap*? Wagner’s view, based on a list of quantified statements, “is that the numbers cited..., taken together, point to a new and little-understood challenge for American education:...all students need new skills for college, careers, and citizenship.”¹⁰ What is education for? In Wagner’s argument, education is for an economic and civic preparation, although I view the second point as only a buzzword in Wagner’s overall argument. The point of education is to prepare future workers, which is traditionally known as “efficiency” in curriculum studies. In Wagner’s text, the values of readiness for international competitiveness and future employment are quantitatively measured and

⁹ Ibid., 19.

¹⁰ Wagner, *The Global Achievement Gap*, xxi emphasis in the original.

argued through the abilities that students demonstrate on standardized tests, like PISA. These measures become proxies of what counts in employment, such as counting the performance on a mathematics exam to determine how well people will perform in the future.

Wagner's text is interesting in the fact that the quantification is not the focus of the schools that are deemed effective in preparing for economic future. He does provide qualitative cases of what he terms successful schools. Always, however, within these qualitative cases, he provides evidence for why these schools are beneficial and considered succeeding through the deployment of quantification stating how well the schools do on standardized tests, their high graduation rates, and the rates of acceptance of students into colleges. The argument uses these numbers as proxies for what the text considers educational value: the preparation for the future economic impact of US students in the work force. Wagner can measure graduation rates and acceptance rates found within his example successful schools, although I doubt that he measured the rates himself instead relying on the data published by the schools. Is the graduation rate really important or is it given rhetorical importance because it could be measured?

This question should not be read as a statement that does not value an individual's graduation from high school. I do believe that graduation is an important personal achievement. But it is a personal achievement. As a human I rejoice in the successful completion of high school, with the potential for future progress, regardless of the direction that progress takes. The issue here is that the rhetoric is using a graduation *rate*, a quantified proxy of a personal achievement. Wagner's rhetoric shifts attention from the values of achieving graduation to that of a measurable summary, allowing for the individual to be consumed within the summary.

Second, consider the purposes of education in *No Excuses*. Thernstrom and Thernstrom argue that the achievement gap in between Black and White students is "the most important civil

rights issue of our time.”¹¹ If I were to consider what is education for according to the argument of *No Excuses*, I might suggest that education exists for a sense of equality, a way to prepare students to be considered equal in achievement and ability. In short, the value of education is to prepare all students equally so that the average student of any race is at equal levels with other students in efforts to promote equal opportunities in future jobs and attainment. The text uses test scores as proxies for the qualities of learning and comprehension to suggest that there is a racial difference of four years by the time students graduate from high school (suggesting that Black and Hispanic students are at an eighth-grade level when they graduate from high school).

These measures of test scores serve as proxies for what is really valued in the Thernstroms’ argument, that of racial inequality. The text assumes that the quantification that is found is an objective measure of racial abilities and allows for clustering and comparison. Thus, in order to measure racial inequality, the Thernstroms value test scores as a way to measure racial gaps and inequality. They suggest in the text that standardized testing is the best current option, suggesting that “blaming the messenger, or at least denying the validity of the message, is far easier than figuring out how to deal with the problem that the test scores have identified.”¹² In the rhetoric of *No Excuses*, the quantification of the test scores becomes a messenger of the problems of racial inequality, a messenger that could be ignored or vilified. Has the test score become of value in the rhetorical argument because it is easy to measure racial demographics based on a score?

Finally consider the purposes of education in Ravitch’s *The Death and Life of the Great American School System*. Ravitch concludes the text with a similar call from the other two texts that

¹¹ Thernstrom and Thernstrom, *No Excuses*, 274.

¹² Ibid., 25.

education serves an economic purpose, holding the “key to developing human capital.”¹³ For Ravitch, though the purpose of education is to

create a renaissance in education, one that goes well beyond the basic skills that have recently been the singular focus of federal activity, a renaissance that seeks to teach the best that has been thought and known and done in every field of endeavor.¹⁴

What is the goal of education? For Ravitch’s argument the value of education comes in developing human capital through the teaching of the best throughout the different fields and domains. From the text, the policies that are currently in place have become focused on a few core subjects instead of creating individuals who have their learning based on diverse experiences. Quantification is used as support for why the systems that are in place are not working. Ravitch is writing this work to announce her position change. Fair enough. Quantification becomes a tool for demonstrating why her positions have need to change, serving as proxies for the qualities found within her educational observation. I see in Ravitch’s work appeals to the quantification as support for why Ravitch is suggesting her changes in positions, suggesting that Ravitch, a trained historian, is assuming that quantification of qualities is more persuasive than the use of other arguments, such as historical or philosophical warrants and structures. The data becomes more valued evidence suggesting what current educational reforms are not doing or how the reforms are not producing what was envisioned or promised.

In these three educational reform texts, the authors argue through assumptions that numbers are more persuasive (read more readily accepted in implementing change) than the use of ethical arguments. Is this a result of educational reform texts being written for general consumers as opposed to educational policy analysts or educational researchers? This dissertation is not in a

¹³ Ravitch, *The Death and Life of the Great American School System*, 223.

¹⁴ Ibid., 224.

position to consider that question fully. However, referring back to education's audit culture discussed in Chapter Two, I would suppose that this trend of valuing quantified arguments over ethical ones would be readily considered and documentable.

I do believe that educational writing and rhetoric is using quantification in attempts to argue for an education that the authors consider better, arguing for what they considering education to be. However, in this deployment of quantification in the arguments I consider how the authors are using quantifications as proxies for what is really of value to them. I can see Biesta's point that the values become those things that can be measured instead of measuring what is valued. I also think that this rhetoric is historically shaped. In educational reform discourse, quantification has been a common tool for comparison and evaluation. It has been deployed as a tool for studying the issues of educational inequality under the assumption that quantities are objective. Quantification has become a common rhetorical tool, one that is almost expected in reformation rhetoric.

However, to quote again from Biesta

Given that the question of good education is a normative question that requires value judgments, it can never be answered by the outcomes of measurement, by research evidence or through managerial forms of accountability—even though...such developments have contributed and are continuing to contribute to the displacement of the question of good education and try to present themselves as being able to set the direction for education.¹⁵

Biesta suggests that quantified evidence of education are enacted within the cultures of outcomes-based educational evaluation, which displace educational experiences that are experiences of qualities for those that can be measured, foreclosing what Biesta considers as good education. The use of quantification is goal steering education in pseudo-productive directions.

¹⁵ Biesta, *Good Education*, 128.

A Sibling Rhetoric

I conclude this dissertation by returning to the question at the beginning of this chapter: What kind of evidence for what kind of reform? In writing this chapter, I considered how educational reforms that use quantification are doing so as possible proxies of what is really valued, valuing (and using) these measurements rhetorically because the measurements can be made. In reporting these measurements as rhetorical evidence, the qualities of desirable education become obscured by the quantities used as these proxies. I do not think it is difficult to consider how schools recently have become obsessed with changing curricula, changing teaching time, and changing teaching evaluations to parallel the changes in standardized testing in the United States. These changes are proxies for what could be considered the purposes of education. People are more interested in the results of test scores than in evaluations of the purposes of education. Yet, the changes effected are changes that mirror the calibrated measures, conflating the ideas of good test scores with the qualities of a good education, which include content learning and mastery but cannot be summarized only to that point.

I can see some objection to this dissertation because I do not offer a concrete replacement rhetoric for the use of quantification. I do not think such a replacement exists; neither should quantification be extinguished from educational discourse. I do, however, suggest that in order to understand the purposes of this dissertation, I consider a critical reading of quantification by offering a single alternative. Are there others? Yes, of course. But in providing a sibling rhetorical deployment, I hope to conclude this dissertation by considering that there are alternatives available. I quote again David Labaree, as mentioned in the epigraph of Chapter Two

In many ways, statistical analysis is compellingly attractive to us...in education. It is a magnet for grant money, since policymakers are eager for the kind of apparently objective data that they think they can trust....*The path of least resistance is to continue in the quantitative vein*, looking

around for new issues you can address with these methods. When you are holding a hammer, everything looks like a nail.¹⁶

I emphasize the line about the path of least resistance as a consideration for what types of evidence are deployed in educational reform arguments made. This path may be the easiest path to obtain funding for change or the path easiest to publish or the path most readily expected as objective. It may be the path that supports the research question and the argument ultimately being made. However, it cannot be the only rhetorical tool used to construct educational reform because quantification obscures the questions that are truly important: What is education for, and what does good education look like? The evidence from these three books suggests that readers have become obsessed by numbers and have lost track of basic questions of value and fairness.

In considering how to offer this sibling rhetoric, I have been drawn to some of the influences in how I construct knowledge and understanding the world. I recognize that this is subjective and biased. I also recognize that I will be mentioning qualities that I find beneficial as rhetorical tools. I begin by a statement about thinking and writing and knowing found in a text about learning to write.

“Thinking is trying to think the unthinkable: thinking the thinkable is not worth the effort. Painting is trying to paint what you cannot paint and writing is writing what you cannot know before you have written: it is preknowing and not knowing, blindly, with words.”¹⁷ A colleague of mine has had part of this statement from Hélène Cixous in the signature line of her emails for some time—opening me to some of Cixous’ writings and ideas. In this statement I believe that Cixous is suggesting that there is a relationship before what can be known beforehand and what is developed

¹⁶ Labaree, “The Lure of Statistics for Educational Researchers,” 21 emphasis mine.

¹⁷ Hélène Cixous, *Three Steps on the Ladder of Writing*, trans. Sarah Cornell and Susan Sellers (Columbia University Press, 1993), 38.

through a creative or learning process. I am drawn to her notion that we cannot know before we write what we will write. Yes, there are general ideas and thoughts and directions and arguments that we may desire to publish and present. However, the difference can be found in the process of writing and seeing the evolution of ideas and arguments as one writes and puts them to words.

I mention Cixous' statement here as a parallel type of rhetoric in educational reform, a rhetoric where educational insight cannot be known before one engages in it. I see a parallel in the processes of education in which one does not have clearly (pre)defined parameters before one engages in the acts of education. An education where the outcomes of such an education cannot be predetermined. I do not think of this as a random education, in the sense of probability, because the possible outcomes are not known beforehand.

This type of education depends on outcomes for evaluation, but the outcomes of this education are not predetermined. In this education the outcomes are part of the discovery, not set in advance (as happens in outcomes-based education). In this notion of discovery, the educational arena shifts. No longer is there the player and director and lighting manager (or taken more anciently—no longer gladiators fighting for permission of the emperor to live). Instead, there is a shift in what education means.¹⁸

This education is not often the portrayed education of the United States. The focus of educational evaluation in the United States is on outcomes; these outcomes, however, are not outcomes that come through discovery but are outcomes that have been determined previously and evaluated by some form of assessment or rubric. In the current rhetoric of educational evaluation, the evidence of knowledge or learning comes from quantified outcomes that measure educational

¹⁸ Gert J. J. Biesta, "Against Learning. Reclaiming a Language for Education in an Age of Learning," *Nordisk Pedagogik* 25, no. 1 (2004): 54–66; Gert J. J. Biesta, "Learner, Speaker, Student: Why It Matters How We Call Those We Teach," *Educational Philosophy and Theory* 42, no. 5–6 (2010): 540–552; Bingham and Biesta, *Jacques Rancière*.

success through demonstration on a predetermined standard. Learning has become associated with solutions that have been determined *a priori*.

This type of education cannot be evaluated through standardized measures and quantifications. This type of education does not claim to *know* beforehand what is being learned or what is being experienced. Again, will there be certain readings and directions to travel? Of course there will be readings determined before actually meeting the students, but the point of this type of education is that the goals of the readings, what will be individually learned from the readings and the discussions and the interpretations cannot be defined by some outside body and standardized through tests that assume a transmutation of qualities into calibrated quantities. This type of educational reform rhetoric cannot solely value arguments based on quantification over ethical arguments, because this type of rhetoric recognizes qualia within the rhetorical construction.

It is in this type of rhetorical structure that the work of equality considered by Jacques Rancière can be most helpful. In his *The Ignorant Schoolmaster*, Rancière considers equality through an experience of Jean Joseph Jacotot, a French-born educator who began teaching in Belgium in the 1800s.¹⁹ In Belgium, Jacotot had Flemish speaking students but was not able to speak the language himself. He did have copies of *Télémanque* that were written in both French and Flemish. He had the students work by themselves to read, translate, and interpret the text themselves, creating an equality that was based on the assumption that all could learn without the aid of a teacher's explanation. The schoolmaster in this case was ignorant, not suggesting that the schoolmaster was incompetent or unintelligent but ignorant. This ignorant schoolmaster "teaches that which is unknown to him or

¹⁹ Rancière, *The Ignorant Schoolmaster*.

her” and in a complementary way, teaching by “a dissociation between mastery of the schoolmaster and his or her knowledge.”²⁰

In Rancière’s construction of an ignorant schoolmaster, there is no way for the schoolmaster to know beforehand what will be learned and how it will be learned, as the schoolmaster cannot predetermine what will be learned from the shared experiences of context, history, and language. It is in ignorance, I believe, that the values of education are shifted from the quantified to the aesthetic, the ethical, and the humane. Earlier in the chapter, I mentioned a question by Gert Biesta asking what is education for. This type of educational reform rhetoric focuses instead on the auditable measures of meeting standards to the qualities and ethics of education.

Peter Taubman suggests that the current perceptions of outcomes-based education in the United States has been previously defined by some institution or system or person who is absent from the classroom. He suggests that educational success is limited to successes that are seen meeting predetermined goals, predetermined by those governing players in the educational arena. From *Teaching by Numbers*

Since learning has already been defined as the achievement of learning outcomes and the ability to monitor and control or manipulate one’s thinking, then content, as we usually think of it, is hollowed out. If lectures, discussions, and learning activities are all directed to developing predefined skills, dispositions, and knowledge as defined by *precisely articulated outcomes*, then there is nothing to explore, there is only something to repeat, there is nothing to question, *there are only answers*, there is nothing to create, there is only reproduction.²¹

²⁰ Bingham and Biesta, *Jacques Rancière*, 1–2.

²¹ Peter Maas Taubman, *Teaching By Numbers: Deconstructing the Discourse of Standards and Accountability in Education* (Routledge, 2009), 192, emphasis mine.

Taubman suggests that within the current notions of outcomes-based education there is only repetitions and mimicry, trying to demonstrate that one has learned what has been stated before the learning occurs. Taubman suggests that educational content is stripped of depth in favor of teaching to standards and objectives. Teaching becomes more about the answers than the creation. I agree that education has become outcomes driven in the United States; it has also become a system that holds to the easy path that was suggested by Labaree above—that of quantification.

This chapter has considered different perspectives of what good education might look like. I wish that it were easy to suggest an answer to that question. It seems, however, that the simplest path in answering that question has become (through historical confluences) that of measuring and comparing qualities through quantification. In this dissertation, I have considered how quantification requires calibrated measurements, in an effort to convert qualities of good education into these comparable quantities. In the current educational discourse, this description of good education through quantification comes from assumptions that goals of good education can be determined *a priori* and declared as standards. These goals then steer the direction that education takes in order to demonstrate the meeting of these goals through technical measures. This goal steering through quantified means has become known as outcomes-based education. In Chapter One of this dissertation, I considered how rhetoric was used to change opinions. What does good education look like? Since this does not have an answer accepted by all involved in education, the purposes of educational reform texts become rhetorical, inviting changes of opinions (and ultimately a change in educational performance). This current educational discourse rhetorically values quantified evidence instead of ethical evidence because quantified measures have become proxies for the qualities of education. Educational reform rhetoric assumes that numbers which measure outcomes are more persuasive than ethical arguments.

I admit that this sibling rhetoric is not considered by the scientific community as objective, in fact, most would consider it completely subjective, basing argument and evaluation on the personal growth instead of focusing on the collective generalizations. The National Research Council's report on scientific research in education would challenge this notion of a rhetoric that is subjective because it considers evidence beyond that which is empirically found and generalizable across populations. However, I recognize that there are arguments that would benefit from these type of evidences and considerations, considerations that cannot be predefined but promote consideration of qualities instead of promoting quantitative improvements, such as with rhetorically articulating learning outcomes.

In a sense, this type of rhetoric creates arguments that are “about imperfect information where incompleteness and indeterminacy are assets,” which “position the absence of foundation as enabling, opening us to the other.”²² What type of educational reform do we want to continue to pursue? What kinds of language and evidence will be used as the vehicle for talking about and evaluating education? This dissertation considered one type of educational reform, reflected in three different books by three different authors. This reform is one that values as rhetorical evidence the transmutation of qualities into quantities, considering the ends of an improved educational system through improvements in calibrated measurements, and replacing questions of value with questions of technique. I consider such a rhetoric as closing and based on historical assumptions of objectivity and efficiency. Although quantification is useful in exploring certain questions, it is very limiting when it defines possibilities for educational reform rhetoric.

²² Lather, *Engaging Science Policy*, 86–87.

APPENDIX

A Quantitative Trope: Quantidoche

What is quantidoche?

In seeking information about quantification as a trope, I cannot find such a reference. However, rhetorical scholars have consistently considered relationships between numbers and arguments. Aristotle, in both *Rhetoric* and *Topics*, considers the probabilities that something is true in constructing syllogisms.¹ Chaïm Perelman and Lucie Olbrechts-Tyteca consider the “increasing use of statistics and the calculus of probabilities” in section 59 of *The New Rhetoric*.² Stephen Toulmin considers how rhetoricians weigh possibilities of audiences accepting claims.³ However, the conveyance of the unfamiliar with the familiar (and comparable) quantification is not discussed. It is easy to see that this trope, quantidoche, relates to the traditional considerations of probability in rhetoric. However, I am considering this beyond just the realm of possibility but in consideration as trope has impact in considering the figured/figurative language used by (educational) texts in shaping perceptions, potentially offering language for how policy and perception are enacted.

Within the art of rhetoric are different moves and language constructions that help persuades audiences of the points being made; one such tool is the trope. Ancient tropes were considered distinct from figures of speech and figures of thought, which were all included in the use of ornamental style.⁴ However, the distinctions between these categories have become blurred in contemporary rhetoric. For this dissertation I provide a definition of trope offered by Daniel

¹ Edward H. Madden, “Aristotle’s Treatment of Probability and Signs,” *Philosophy of Science* 24, no. 2 (April 1957): 167–172.

² Chaïm Perelman and Lucie Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*, trans. John Wilkinson and Purcell Weaver (Notre Dame, Ind: University of Notre Dame Press, 1971), 255–260.

³ Stephen Edelston Toulmin, *The Uses of Argument* (Cambridge: Cambridge University Press, 2003).

⁴ Crowley and Hawhee, *Ancient Rhetorics for Contemporary Students*, 334.

Chandler that tropes are “rendering the unfamiliar more familiar” through the use of language.⁵ This dissertation constructs a quantitative rhetorical trope, *quantidoche*, taken from the Latin *quam* meaning “how much” and the Greek *dechomai* meaning “to receive.” In part, I have based this portmanteau on the rhetorical trope synecdoche, which is a trope that considers the whole from a given part or the part from the given whole. The use of quantification as a tool for generalization and validity is in essence representing the whole by the part. Since this dissertation considers how quantifications are received by public audiences, *dechomai* is an appropriate verb. I consider the blending of Latin and Greek to be fortuitous to connotations of this trope.

Although a rhetorical trope can be used in diverse fields and domains, for the purposes of this dissertation, I limit quantidoche to educational texts. In order to consider the use of this trope in educational texts, I consider a statement offered by The American Educational Research Association (AERA) relating to quantification through measurement. “Measurement is the process by which behavior or observation is converted into quantities, which may, in turn, then be subjected to some kind of quantitative analysis.”⁶ It is strange to me that behavior and observations are considered in the same form of conversion for quantitative analysis. This strangeness could be examined through different methodologies, but for this dissertation it will be considered through quantidoche. The ideas of converting qualities into quantities is important in understanding the work of this dissertation as the quantifications being analyzed do not begin as countable things, but qualities are put through a process of stripping away the individual for the general, representing the whole by figurative part, a move that I call quantidoche.

⁵ Daniel Chandler, “Rhetorical Tropes,” *Semiotics for Beginners*, 2001, <http://www.aber.ac.uk/media/Documents/S4B/sem07.html>.

⁶ AERA, “Standards for Reporting on Empirical Social Science Research in AERA Publications,” Educational Researcher, 35.6, 36. http://www.sagepub.com/upm-data/13127_Standards_from_AERA.pdf

Why analyze quantidoche?

Education affects more than just the students and teachers in the classrooms; it is considered and discussed by everyone. It is beyond just an academic discipline but is considered by parents who desire certain outcomes for their children, and by politicians who consider economic and monetary educational products. Because education has garnered opinions from diverse sources, there are many depictions written to explore, exemplify, or condemn activities associated with education, often written toward a (fractured) truth of what education *is* or what it *should be*. These writings range from practitioner research to the opinions pages in local newspapers and magazines to books to popular cinema. Among all these choices of potential texts, this dissertation explores the shaping of public perceptions through the ornamental use of quantification.

What advantages come from conceiving quantification as a trope? Rhetorical analysis provides glimpses into potential beliefs or actions associated with a given text from audiences. Rhetorical tropes consider the figurative language used to illuminate the uncommon by the common. Considering tropes allows for rhetorical analysis to consider not only the arguments of debated issues but the language of those arguments and critique of how those issues are framed in language. As Thomas Farrell considers, rhetoric is a real-life art, an art that considers issues of communal interest.⁷ Education is not only concerned with the real-life, but also addresses issues that are addressed by educational professionals and laypersons alike. Quantification is common rhetorical evidence seen in education, used as a comparative tool explaining issues that are for some unknown. Looking at quantification as a trope allows looking beyond the arguments addressing inequalities in attempts to understand the messages portrayed through these arguments.

⁷ Thomas Farrell, "Practicing the Art of Rhetoric: Tradition and Invention," in *Contemporary Rhetorical Theory: A Reader*, ed. John Louis Lucaites, Celeste Michelle Condit, and Sally Caudill (New York: The Guilford Press, 1999), 79–100.

Because education affects everyone, it is important to consider the messages that are being constructed about education. This dissertation considers a social-life aspect of education affecting the populace, and as such considers the issue of quantification through humanities-oriented research. This is supported by the standards of humanities-oriented research outlined by the American Educational Research Association

Humanities-oriented research in education attempts to gain an understanding of the explicit and implicit *messages* and *meanings* of education, to point out the tensions and contradictions among them, and to compare and critique them on ethical or other value-oriented grounds. A prominent feature of humanities-oriented research in education is its use of interpretive methods, broadly construed, which investigate the history, meanings, beliefs, values, and discourses that human beings employ in the production of *social life*.⁸

This statement considers why rhetorical studies of educational text might be of benefit as it allows for a consideration of the messages and meanings of education. Education is an issue that is considered by community as a part of the construction of social life. In considering the rhetorical nature of educational texts, there is opportunity to consider tensions that are part of the social community.

⁸ AERA, “Standards for Reporting on Humanities-Oriented Research in AERA Publications” (2009, 482, emphasis mine) http://www.aera.net/uploadedFiles/Journals_and_Publications/Journals/481-486_09EDR09.pdf

REFERENCES

REFERENCES

- “2011 National School Climate Survey: LGBT Youth Face Pervasive, But Decreasing Levels of Harassment | GLSEN: Gay, Lesbian and Straight Education Network.” Accessed April 11, 2013. <http://www.glsen.org/cgi-bin/iowa/all/news/record/2897.html>.
- AERA. “Standards for Reporting on Empirical Social Science Research in AERA Publications.” *Educational Researcher* 35, no. 6 (September 2006): 33–40.
- . “Standards for Reporting on Humanities-Oriented Research in AERA Publications.” *Educational Researcher* 38, no. 6 (September 2009): 481–486.
- Apple, Michael W. “Education, Markets, and an Audit Culture.” *Critical Quarterly* 47, no. 1–2 (2005): 11–29. doi:10.1111/j.0011-1562.2005.00611.x.
- Aristotle. *Rhetoric*. Translated by W. Rhys Roberts, 1954.
- Bercovitch, Sacvan. *The American Jeremiad*. Univ of Wisconsin Press, 1980.
- Biesta, Gert J. J. “Against Learning. Reclaiming a Language for Education in an Age of Learning.” *Nordisk Pedagogik* 25, no. 1 (2004): 54–66.
- . *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Boulder, CO: Paradigm Publishing, 2010.
- . “Learner, Speaker, Student: Why It Matters How We Call Those We Teach.” *Educational Philosophy and Theory* 42, no. 5–6 (2010): 540–552.
- Bingham, Charles, and Gert J. J. Biesta. *Jacques Rancière: Education, Truth, Emancipation*. London: Continuum International Publishing Group, 2010.
- Chandler, Daniel. “Rhetorical Tropes.” *Semiotics for Beginners*, 2001. <http://www.aber.ac.uk/media/Documents/S4B/sem07.html>.
- Cherryholmes, Cleo H. “Construct Validity and the Discourses of Research.” *American Journal of Education* 96, no. 3 (May 1, 1988): 421–457. doi:10.2307/1084999.
- . “Theory and Practice: On the Role of Empirically Based Theory for Critical Practice.” *American Journal of Education* 94, no. 1 (November 1, 1985): 39–70. doi:10.2307/1085291.
- Cixous, Hélène. *Three Steps on the Ladder of Writing*. Translated by Sarah Cornell and Susan Sellers. Columbia University Press, 1993.
- Committee on Scientific Principles for Education Research. *Scientific Research in Education*. Edited by Richard J. Shavelson and Lisa Towne. Washington, D.C.: The National Academies Press, 2002.

- Cronbach, Lee J., and Paul E. Meehl. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52 (1955): 281–302.
- Crowley, Sharon, and Debra Hawhee. *Ancient Rhetorics for Contemporary Students*. 4th ed. White Plains, NY: Longman, 2008.
- Derrida, Jacques. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore, MD: The Johns Hopkins University Press, 1997.
- Dewey, John. *The Quest for Certainty: A Study of the Relation of Knowledge And Action*. Lightning Source Incorporated, 2005.
- Farrell, Thomas. "Practicing the Art of Rhetoric: Tradition and Invention." In *Contemporary Rhetorical Theory: A Reader*, 79–100. edited by John Louis Lucaites, Celeste Michelle Condit, and Sally Caudill. New York: The Guilford Press, 1999.
- Fendler, Lynn. "Why Generalisability Is Not Generalisable." *Journal of Philosophy of Education* 40, no. 4 (November 2006): 437–449.
- Fendler, Lynn, and Irfan Muzaffar. "The History of the Bell Curve: Sorting and Idea of Normal." *Educational Theory* 58, no. 1 (2008): 63–82.
- Foucault, Michel. "Of Other Spaces: Heterotopias." Translated by Jay Miskowiec. *Architecture / Mouvement / Continuité* (1984).
<http://foucault.info/documents/heteroTopia/foucault.heteroTopia.en.html>.
- . *The Order of Things: An Archaeology of the Human Sciences*. Routledge, 1970.
- . "The Subject and Power." In *Michel Foucault: Beyond Structuralism and Hermeneutics*, 208–226. edited by Hubert Dreyfus and Paul Rabinow. 2nd ed. Chicago: The University of Chicago Press, 1983. <http://foucault.info/documents/foucault.power.en.html>.
- Gaiman, Neil. *American Gods: The Tenth Anniversary Edition: A Novel*. HarperCollins, 2011.
- Gay, Lesbian and Straight Education Network. *2009 Nation School Climate Survey*. New York: GLSEN, 2010. www.glsen.org/research.
- Gould, Stephen Jay. *The Mismeasure of Man*. W. W. Norton & Company, 1996.
- Grek, Sotiria. "Governing by Numbers: The PISA 'effect' in Europe." *Journal of Education Policy* 24, no. 1 (2009): 23–37. doi:10.1080/02680930802412669.
- Hacking, Ian. *Logic of Statistical Inference*. London: Cambridge University Press, 1976.
- . *The Taming of Chance*. Cambridge: Cambridge University Press, 1999.

- Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24, no. 3 (September 1, 1986): 1141–1177. doi:10.2307/2725865.
- . "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 18, no. 4 (May 1, 1989): 45–62. doi:10.3102/0013189X018004045.
- Hawhee, Debra. *Bodily Arts: Rhetoric and Athletics in Ancient Greece*. University of Texas Press, 2004.
- Hopmann, Stefan Thomas. "No Child, No School, No State Left Behind: Schooling in the Age of Accountability 1." *Journal of Curriculum Studies* 40, no. 4 (2008): 417–456. doi:10.1080/00220270801989818.
- Ingebretsen, Edward J. *At Stake: Monsters and the Rhetoric of Fear in Public Culture*. Chicago: University of Chicago Press, 2001.
- Labaree, David F. "Public Goods, Private Goods: The American Struggle Over Education." *American Educational Research Journal* 34, no. 1 (1997): 39–81.
- . "The Lure of Statistics for Educational Researchers." In *Educational Research: The Ethics and Aesthetics of Statistics*, 13–25. edited by Paul Smeyers and Marc Depaepe. Dordrecht: Springer, 2010.
- Lather, Patti. *Engaging Science Policy: From the Side of the Messy*. New York: Peter Lang Publishing, Inc., 2010.
- Leff, Michael C. "The Habitation of Rhetoric." In *Contemporary Rhetorical Theory: A Reader*, 52–64. New York: The Guilford Press, 1999.
- Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company, 2004.
- Lindblad, Sverker, and Thomas S. Popkewitz. *Educational Restructuring: International Perspectives On Traveling Policies*. Charlotte, NC: Information Age Publishing, 2004.
- MacLure, Maggie. *Discourse in Educational and Social Research*. 1st ed. Maidenhead: Open University Press, 2003.
- Madden, Edward H. "Aristotle's Treatment of Probability and Signs." *Philosophy of Science* 24, no. 2 (April 1957): 167–172.
- Miller, Bennett. *Moneyball*. Film, Biography, Drama, Sport, 2011.
- National Commission on Excellence in Education. *A Nation at Risk*, April 1983. <http://www2.ed.gov/pubs/NatAtRisk/risk.html>.
- Perelman, Chaïm, and Lucie Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. Translated by John Wilkinson and Purcell Weaver. Notre Dame, Ind: University of Notre Dame Press, 1971.

- Popkewitz, Thomas S. *Changing Patterns of Power: Social Regulation and Teacher Education Reform*. SUNY Press, 1993.
- . *Cosmopolitanism and the Age of School Reform: Science, Education, and Making Society by Making the Child*. Routledge, 2008.
- . *Educational Knowledge: Changing Relationships Between the State, Civil Society, and the Educational Community*. SUNY Press, 2000.
- . “PISA: Numbers, Standardizing Conduct, and the Alchemy of School Subjects.” In *Pisa Under Examination*, 31–46. edited by Miguel A. Pereyra, Hans-Georg Kotthoff, Robert Cowen, Allan Pitman, Vandra Masemann, and Miguel A. Pereyra. Comparative and International Education. SensePublishers, 2011.
<http://www.springerlink.com/content/pn502k54w313q435/abstract/>.
- Porter, Theodore. *The Rise of Statistical Thinking: 1820 – 1900*. Princeton, NJ: Princeton University Press, 1988.
- . *Trust in Numbers*. Princeton, NJ: Princeton University Press, 1995.
- Power, Michael. “Evaluating the Audit Explosion.” *Law & Policy* 25, no. 3 (2003): 185–202.
doi:10.1111/j.1467-9930.2003.00147.x.
- . *The Audit Society: Rituals of Verification*. 2 Sub. Oxford University Press, USA, 1999.
- Rancière, Jacques. *The Ignorant Schoolmaster: Five Lessons in Intellectual Emancipation*. Stanford University Press, 1991.
- . *The Politics of Aesthetics: The Distribution of the Sensible*. Translated by Gabriel Rockhill. Continuum International Publishing Group, 2006.
- Ranciere, Jacques, and Steven Corcoran. *Dissensus: On Politics and Aesthetics*. Continuum International Publishing Group, 2010.
- Ravitch, Diane. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books, 2010.
- Rose, Nikolas S. *Powers of Freedom: Reframing Political Thought*. Cambridge University Press, 1999.
- Shore, Cris, and Susan Wright. “Audit Culture and Anthropology: Neo-Liberalism in British Higher Education.” *The Journal of the Royal Anthropological Institute* 5, no. 4 (1999): 557–575.
- Silver, Nate. *The Signal and the Noise: Why So Many Predictions Fail--but Some Don't*. New York: The Penguin Press, 2012.
- Smeyers, Paul, and Marc Depaepe. “Representation or Hard Evidence? The Use of Statistics in Education and Educational Research.” In *Educational Research: The Ethics and Aesthetics of*

- Statistics*, edited by Paul Smeyers and Marc Depaepe. Vol. 5. Educational Research. Dordrecht: Springer, 2010.
- Spivak, Gayatri Chakravorty. *In Other Worlds: Essays in Cultural Politics*. New York: Routledge, 2006.
- . *Other Asias*. Malden, MA: Blackwell Publishing, 2008.
- Squire, Peverill. “Why the 1936 Literary Digest Poll Failed.” *Public Opinion Quarterly* 52, no. 1 (1988): 125–133.
- Star, Jon R., Sharon Strickland, and Amanda Hawkins. “What Is Mathematical Literacy? Exploring the Relationship Between Content-Area Literacy and Content Learning in Middle and High School Mathematics.” In *Meeting the Challenge of Adolescent Literacy: Research We Have, Research We Need*, 104–111. edited by Mark W. Conley, Joseph R. Freidhoff, Michael B. Sherry, and Steven F. Tuckey. New York: Guilford Press, 2008.
- Stewart, Susan. *On Longing: Narratives of the Miniature, the Gigantic, the Souvenir, the Collection*. Durham, NC: Duke University Press, 2005.
- Taubman, Peter Maas. *Teaching By Numbers: Deconstructing the Discourse of Standards and Accountability in Education*. Routledge, 2009.
- Thernstrom, Abigail, and Stephan Thernstrom. *No Excuses: Closing the Racial Gap in Learning*. New York: Simon & Schuster, 2003.
- Toulmin, Stephen Edelston. *The Uses of Argument*. Cambridge: Cambridge University Press, 2003.
- Tufte, Edward R. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press, 2001.
- . *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press, 1997.
- Wagner, Tony. *The Global Achievement Gap: Why Even Our Best Schools Don’t Teach the New Survival Skills Our Children Need--and What We Can Do About It*. New York: Basic Books, 2010.