Title

Talking Our Way Around Expert Caution: A Rhetorical Analysis of Value-Added Modeling in Teacher Evaluation

Permalink

https://escholarship.org/uc/item/97q9r1kp

Author

Tobiason, Glory

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Talking Our Way Around Expert Caution: A Rhetorical Analysis of Value-Added Modeling in

Teacher Evaluation

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Glory Amanda Tobiason

2017

ABSTRACT OF THE DISSERTATION

Talking Our Way Around Expert Caution: A Rhetorical Analysis of Value-Added Modeling in

Teacher Evaluation

by

Glory Amanda Tobiason

Doctor of Philosophy in Education

University of California, Los Angeles, 2017

Professor Mark P. Hansen, Chair

This dissertation proceeds in two steps, a theoretical one followed by an empirical one. Considered together, they help us understand how persuasive language may be used to endorse or justify scientifically shaky policies.

In the first step, I introduce and define a troubling phenomenon that may occur in public decision making, Audacious Use of an Information Tool (AUIT). AUIT is curious because it involves a decision not to defer to the scientific community on a question that solicits their expertise: whether a tool of science is methodologically capable doing what it is being asked to do. AUIT is potentially dangerous because it involves high-stakes decisions or plans based on possibly faulty information. I identify two cases of AUIT: the use of value-added modeling (VAM) in teacher evaluation and the use of predictive genetic testing to make personal health-care decisions.

In the second step, I map the language of AUIT using a case-study design. I focus on the first example mentioned above, the use of VAM as a central component in high-stakes personnel

decisions about individual teachers. My rhetorical analysis considers a corpus of texts from 2008

to 2015, drawn from popular media, advocacy discourse, and policy messaging from the

Department of Education. I identify four rhetorical moves that rationalize the Audacious Use of

VAM in teacher evaluation and conjecture about whether they are likely to be deployed in other

cases of AUIT.

The dissertation of Glory Amanda Tobiason is approved.

Leah Ceccarelli

Aaron Panofsky

Mike Seltzer

Noreen Webb

Mark P. Hansen, Committee Chair

University of California, Los Angeles

2017

To my dad, whose love

and love of teaching

inspire my work.

To my mom, whose love

and strength

make me strong.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEYWORDS

Education Policy

Education Research

Expertise

Measurement

Public Understanding of Science

Rhetoric

Science Communication

Science Policy

Social Science

Sociology of Science

Teacher Evaluation

Teacher Quality

Validity

Value-Added Modeling

ACRONYMS

Audacious Use of an Information Tool (AUIT)

Direct-to-Consumer (DTC)

Dynamic Stochastic General Equilibrium (DSGE)

Eclipsing Technical Concerns (ETC)

Frequently Asked Questions (FAQs)

Ignoring Technical Concerns (ITC)

Genetically Modified Organism (GMO)

Los Angeles Unified School District (LAUSD)

Manufacturing Resolution of Technical Concerns (MRTC)

Questioning Motives of Technical Concerns (QMTC)

Race to the Top (RTT)

Science, Technology, and Society (STS)

Scientific Certainty Argumentation Methods (SCAMS)

Single Nucleotide Polymorphism (SNP)

Sociology of Scientific Knowledge (SSK)

United States Department of Education (USDOE)

Value-Added Modeling (VAM)

GLOSSARY

Certified Experts – Members of the socially-recognized scientific community (Turner, 2001), whose expertise has been certified (Collins & Evans, 2002) and who possess specialized, technical knowledge or skills; used interchangeably here with "Scientists."

Contributory Expertise – The type of expertise necessary to contribute to a scientific domain. One with such expertise could, for example, "reasonably apply for a job in the science in question or at least publish papers in the professional journals or perhaps be let loose in the laboratory" (Collins, 2004, p. 128).

Deficit Model – A conception of public understanding of science that takes the primary source of conflict in technical decision-making contexts to be public ignorance, attributed to journalistic obfuscation or the public's irrational beliefs, limited experience, or insufficient cognitive capacities (Gross, 1994; Nisbet & Scheufele, 2009). Scholars note that this conception "plays into the hands of technocratic attitudes among decision makers: a de facto ignorant public is disqualified from participating in science policy decisions" (Bauer, Allum, & Miller, 2007, p. 80).

Experience-Based Experts – "Members of the public who have special technical expertise by virtue of experience that is not recognized by degrees or other certificates" (Collins & Evans, 2002, p. 238).

Information Tool – An analytical method or device developed and used by scientists, that outputs information from which an inference is made about the state of the natural or social world.

Interactional Expertise – The type of expertise that involves fluency in the technical language of a scientific domain, without the ability to contribute to the domain. Examples might include "sociologists [of science], high class journalists, or certain types of science administrators" (Collins, 2004, p. 129).

Linear Model of Science-Policy Interaction – An account of science-policy interaction. "Scientists… discover how nature work[s], and then this information [is] 'handed off' to the civic arena… where citizens… decide how to apply the scientists' discoveries to technology and policy development" (Walsh, 2010, p. 38).

Measurement – The use of a test to generate a score (i.e., a number) that gives information about the quantity of a particular psychological construct present in an individual (Crocker & Algina, 1986).

Non-Scientist – Everyone who's not a scientist.

Positivist – One of two main perspectives (along with "relativist") about the nature of scientific truth in the Science Wars. A robust form of positivism takes the natural world to be a "unique truth and the current state of scientific knowledge is assumed to be the best available approximation to that truth" (Martin & Richards, 1995, p. 3). It is assumed that the social factors

intervening between nature and scientific truth are minimal, so "there is no need to examine why scientists believe what they believe" (Martin & Richards, 1995, p. 3).

Post Normal Science – "Those inquiries that occur at the interfaces of science and policy where uncertainties and value-loadings are critical." The term encompasses the entire problem-solving endeavor, of which "science" (as traditionally understood) is but one part of a complex system of interrelated natural, technical and societal elements. "Depending on the particular context, the task may be more like policy-related research, or science-related decision making, or creative technical-social innovation" (Funtowicz & Ravetz, 2003, p. 4).

Problem of Extension – An important theoretical problem in science studies related to authority and participation in technical decision-making contexts. The problem begins with a disciplinary tendency to expose the value-laden-ness of science and thus weaken the distinction between scientists and nonscientists. We encounter the Problem of Extension when "there are no longer any grounds for limiting the indefinite extension of technical decision-making rights" (Collins & Evans, 2002, p. 235).

Relativist – One of two main perspectives (along with "positivist") about the nature of scientific truth in the Science Wars. A robust form of relativism takes scientific reality to be "the set of statements considered too costly to modify" (Latour & Woolgar, 1979, p. 243). Scientists' accounts of the natural world "are not directly given by nature, but may be approached as the products of social processes and negotiations" (Martin & Richards, 1995, p. 4) within the scientific community.

Rhetorical Ingredient – Any element of the rhetorical situation that might be exploited in order to build an argument, including a particular stakeholder or group; the historical or economic context for the discourse; a feature of the media landscape; a political, social, or ethical exigency; etc.

Science – A problem-solving endeavor (Funtowicz & Ravetz, 2003) that (1) involves the work of certified experts, (2) has as its purpose the systematic pursuit of knowledge (Pielke, 2007), and (3) involves observation, testing, theory choice or development, and logical and mathematical reasoning (Williams, 2000).

Science Wars – A series of heated, intellectual exchanges that began in the early 1990s concerning the nature of scientific truth. A central tension existed between the positivist and relativist perspectives: the extent to which scientific understanding is an apprehension of natural truth and the extent to which it is a product of social processes and negotiations.

Science-Regulating Policy – Policy that regulates the practice of certified experts; one of two directions of the science-policy relationship, that where science is the object of policy (Funtowicz & Strand, 2007).

Science-Related Policy – Policies that regulate social relations and involve scientific knowledge or tools in their formation (Pew Research Center & American Association for the Advancement of Science, 2015); one of two directions of the science-policy relationship, that where products of science are used in policy (Funtowicz & Strand, 2007).

Scientists – Members of the socially-recognized scientific community (Turner, 2001), whose expertise has been certified (Collins & Evans, 2002) and who possess specialized, technical knowledge or skills; used interchangeably here with "Certified Experts."

Strong Programme – A school of thought in the Sociology of Scientific Knowledge (SSK) that rejects so-called "weak" sociologies of science: those that only apply social analysis to *failed* scientific theories. Scholars in the Strong Programme "attempt to explain adherence to *all* beliefs about the natural world, whether they be perceived to be true or false, rational or irrational, successful or failed, in an equivalent or symmetrical way [emphasis added]" (Martin & Richards, 1995, p. 5).

Technical Decision-Making Context – A point at which "science and technology intersect with the political domain over issues that are of visible relevance to the public" (Collins & Evans, 2002, p. 236).

Tool – See "Information Tool."

Validity – "The overall degree of justification for test interpretation and use… [which is] concerned with social context and social purpose [and] scientifically and philosophically grounded in both evidence and ethics" (Messick, 1981).

VAM-Based Teacher Evaluation – The central use of VAM in the evaluation of individual teachers for high-stakes purposes.

ACKNOWLEDGEMENTS

I am deeply grateful to the members of my dissertation committee, who were absurdly giving of their time, attention, and expertise. Aaron Panofsky, thank you for helping me see my discipline from the outside. Mike Seltzer, thank you for your steadfast insistence—from the very beginning—that this sort of work is worth doing. Leah Ceccarelli, thank you for taking a chance on what must have looked like a dubious introductory email from Southern California two years ago, and for your invaluable feedback ever since. Noreen Webb, thank you for being a role model, advocate, keen editor, and mentor for the entirety of my time at UCLA; I look up to you. Mark Hansen, I could not have done this without you. Thank you for keeping me on track, reassuring me during the panicky parts, and being so fully engaged in the process.

This work was made possible by the generous intellectual and financial support of the Social Research Methodology faculty in UCLA's Graduate School of Education and Information Studies. In particular, I am indebted to Felipe for the frowny eyebrows and fierce critiques and to Mike Rose for helping me think like a writer. I would also like to thank Susan Losh at Florida State University and Christina Haas at the University of Minnesota for their encouragement and willingness to provide detailed feedback on some of the conceptual groundwork for this study. And I am grateful to Audrey Amrein-Beardsley at Arizona State University as well, for sharing her knowledge of VAM implementation.

My peers have been invaluable in this effort. Thank you especially Kevin Schaff and Megan Kuhfeld, for your patient and thoughtful responses to non-sequitur emails about

quantitative esoterica. Thank you also Sebastian Lemire and Nicole Mancevice, for our little writers' group, which evolved so delightfully over the years.

On a more personal note… Thank you, David, for being right beside me throughout this thrilling, teary, joyous, exhausting, all-absorbing process. Thank you for your reassurance on so many occasions that "the diss is not broken." You are my rock. Finally, Amelia. Thank you for driving what must be thousands of miles, in total, for our weekly sister-dates. Thank you for listening, caring about this work, and lending your fearsome editing powers. Thank you for keeping me sane and honest. I believe in me because you believe in me.

VITA

EDUCATION

2008	TESOL Certification LanguageCorps Phnom Penh, Cambodia
2003	M.S., Pure Mathematics Western Washington University Bellingham, WA
2002	B.S., Mathematics/French Western Washington University Bellingham, WA

PROFESSIONAL EXPERIENCE

2015-2017	Graduate Teaching Assistant University of California Los Angeles Los Angeles, CA
2014-2015	Content Developer / Course Facilitator WestEd San Francisco, CA
2012-2015	Graduate Student Researcher National Center for Research on Evaluation, Standards, and Student Testing - UCLA Los Angeles, CA
2010-2012	ESL Specialist Center City Public Charter Schools Washington, DC
2009-2010	Kindergarten EFL Teacher Terakki Vakfı Okulları Istanbul, Turkey
2008-2009	Pre-School EFL Teacher American Pacific School Phnom Penh, Cambodia
2006-2008	High School Math Teacher The Next Step Public Charter School Washington, DC

2003-2005	Teacher / Community Development Specialist United States Peace Corps Dar es Salaam, Tanzania
2002-2003	Graduate Teaching Assistant Western Washington University Bellingham, WA

SELECTED PUBLICATIONS

Tobiason, G. A. (2015). Expertise gaps in value-added modeling: Are we consulting the right experts? *Bulletin of Science, Technology and Society, 34*(5-6), 183-191.

Tobiason, G. A. (2014). The semantics of measuring teacher effectiveness: How word choice shapes public perception, policy, and practice. *InterActions: UCLA Journal of Education and Information Studies, 10*(1).

Scheriber, S. J. & Tobiason, G. A. (2003). The evolution of resource use. *Journal of Mathematical Biology, 47*(1), 56-78.

HONORS AND AWARDS

2016	Dissertation Year Fellowship, UCLA Graduate Division
2015	Merit Scholarship, UCLA Graduate School of Education and Information Studies
2014	Merit Scholarship, UCLA Graduate School of Education and Information Studies Graduate Summer Research Mentorship Program, UCLA Graduate Division
2013	Merit Scholarship, UCLA Graduate School of Education and Information Studies Louise Tyler Fellowship, UCLA
2012	Merit Scholarship, UCLA Graduate School of Education and Information Studies

I. Chapter One: Introduction

A. *How This Dissertation Came to Be and What to Expect*

This is a strange dissertation. It doesn't follow a traditional structure and it sits at the intersection of three disciplines: educational research; Science, Technology, and Society (STS); and rhetorical criticism. To orient the reader to the logic of what's to come, I'll begin with a few paragraphs describing how my interest in this work developed and how the different pieces of the dissertation fit together.

When I began my career as a classroom teacher, neither I nor any of my colleagues had heard of value-added modeling (VAM); ten years later, when I left the classroom for graduate school, VAM was a part of nearly every conversation about our professional evaluations. I was both intrigued by and skeptical of the tool, asking myself, *what's the relationship between my VAM score and what I do every day with my students?* During my training as an education researcher, I developed a more scientific perspective on VAM, informed by what I was learning about research design, validity, measurement, statistical techniques, etc. I came to appreciate the boundaries of the tool's technical capacity, including the limited precision, stability, and accuracy of VAM scores for individual teachers.

I was struck by the odd disconnect between the technical limitations of VAM (which are common knowledge in the scientific community) and the way the tool was being used in high-stakes teacher evaluations. It puzzled me that widespread expert concern about the technical capacity of the tool seemed not to constrain its policy use. Though my coursework focused my attention on expert discourse about VAM, I also kept an ear to the nonexpert side of the

conversation, where I noticed something interesting: patterns of rhetoric that seemed related in some way to the puzzling disconnect between expert caution and tool use.

At this point, my curiosity branched into what would eventually become the two primary tasks of this dissertation. First, I set out to generalize, define, and (if possible) locate other examples of this odd disconnect, which I eventually named "Audacious Use." This work is described in Chapters 2 and 3. Second, I devised and carried out a rhetorical analysis of nonexpert discourse around VAM. This work is described in Chapters 4 and 5. Eventually these two curiosity-branches joined up again to reveal some of the rhetorical underpinnings of Audacious Use. This work is described in Chapter 6.

B. Roadmap for the Reader

This chapter begins with background information about teacher evaluation, including where VAM fits, logically, into the endeavor. After this, I describe how the use of VAM in teacher evaluation was incentivized by the federal government in the late 2000's and how this played out in individual states. By detailing the policies implemented in Florida, I exemplify the phenomenon at the heart of the dissertation. This is followed by a statement of the problem and the specific research questions that address it. I discuss the importance and limitations of the study and close with a preview of what the reader can expect in the subsequent chapters.

C. Teacher Evaluation

1. Widespread Interest in Teacher Quality

In the early 1980's, a sobering narrative began to shape Americans' perception of our education system: schools are in crisis, this jeopardizes our economic security, and so reform is urgently needed (McIntush, 2000). For some, this sense of crisis and urgency is as strong now as

2

it was three decades ago. Arne Duncan (Secretary of Education under Obama) described our nation's "urgent educational problems" as "morally unacceptable and economically unsustainable" (Duncan, 2012, para. 21) and three-quarters of Americans would assign a grade of C or worse to the nation's public schools (Phi Delta Kappa International, 2016).

Contemporary popular media coverage of education reform tends to center on four topics: school choice, common academic standards, standardized testing, and teacher quality. For many, the last of these has a powerful emotional proximity. Most of us spent our formative years in close, daily contact with teachers. While we may be unsure about how charter schools work, or how standards differ from curricula, our perception of the nature and impact of good teaching draws on extensive subjective experience. Not surprisingly, the vast majority of us believe teacher quality is the most important factor in improving education (Phi Delta Kappa International, 2015).

The gut appeal of "teacher quality" is often intensified with reference to a powerful archetype, the "heroic teacher" (Steudeman, 2014) who "overcome[s] challenges of austerity, poverty, and racism without outside assistance" (p. 502). This myth attributes student success to good teachers and draws attention away from other contributing factors, including economic advantage, racial or geographic privilege, or family support. Portraying the work of teachers in this way—as a sufficient remedy for all that ails education—channels a sense of urgency about education reform *in general* into a focused interest in teacher quality.

2. A Tangle of Technical and Nontechnical Issues

Increasingly over the past decade, national and state lawmakers have come to view teacher-focused education reform as part of their job description, and many of their policy proposals follow a simple narrative, sketched here in italics. *Teachers matter a lot, so we should focus our*

reform efforts getting more good ones and fewer bad ones in our classrooms. Personnel

decisions to accomplish this depend on information about the quality of individual teachers, and

so we need a way to gauge this. The logic of this narrative may sound simple and

uncontroversial, but it elides several complex questions about the potential impact of teachers

(relative to other policy levers), the nuances of what good teaching is and how to recognize it,

and the logistics of how to get more good teachers. Below, I list some of the more significant

and controversial of these questions.

- How much responsibility for student learning lies with teachers, schools, family, the

  community, etc.?

- How far into the future and in what ways do teachers influence their students?

- What do we believe teachers should be doing?

- How will we know if they are doing it?

- If a teacher's performance is not acceptable, what are the pros and cons of various

  personnel actions we might take (e.g., retraining her, replacing her, etc.)?

- How important is staff consistency to a school community? To student learning?

What the reader should notice about these questions is that most can be answered from either

a nontechnical perspective (e.g., drawing on values or personal experience) or a technical

perspective (e.g., drawing on empirical research). For example, we might approach the first

question based on what we believe about how learning occurs; what we believe about student

agency; our experience as students; our experience as parents of students; etc. At the same time,

data analysts and statisticians can shed a different kind of light on the question: researchers who

study sources of variance in student outcomes have found, for instance, that out-of-school factors

have much greater influence on student success than in-school factors, including teachers.

This two-sided-ness means that public decision making around teacher evaluation can be described as a *technical decision-making context*: a point "where science and technology intersect with the political domain because the issues are of visible relevance to the public" (Collins & Evans, 2002, p. 236). The key questions in these scenarios don't lie neatly on one side or the other of an expert-nonexpert divide. Thus decision makers and those seeking an informed perspective face the challenges of integrating nontechnical and technical insights and deciding—in the case that these two perspectives don't align—which to prioritize. These challenges have become particularly salient since value-added modeling (VAM) arrived on the scene as a strategy to gauge the quality of individual teachers.

3. Enter: Value-Added Modeling (VAM)

VAM is a collection of statistical techniques that aims to link-year-to-year changes in student test scores to various factors that might influence these changes. Typically, researchers use VAM to study *large-scale* factors; they make inferences about, for example, the effectiveness of a new state curriculum or the introduction of a district-wide professional development program. But in the last decade, the use of VAM to make inferences about *individual* teachers has gained popularity.

When VAM[1] is used in teacher evaluation, the basic logic involves comparing a student's *actual* test score to *what she would have scored* if she had been taught by an average teacher. A student's predicted test score is calculated from her prior tests scores and other background information. For example, the predicted test score of a struggling English-language learner will

---

[1] For the purposes of this discussion, it is unnecessary to differentiate between the methodological variants of VAM or even from "student growth models," which are used similarly in teacher evaluation. The technical concerns and complexities of interpretation that arise in the expert caution I will analyze are the same regardless of how the model is specified or

be lower than the predicted test score of typically-high-scoring native speaker. If the student

scores higher than what the model predicted, this difference is attributed to her teacher, who

receives a higher VAM score (which is calculated by aggregating over students in the teacher's

classroom). This logic is illustrated in Figure 1-1 below.

The logic of VAM

Figure 1-1. This figure illustrates how a hypothetical student's predicted 7th grade test
score is used to compute a VAM score for her teacher.
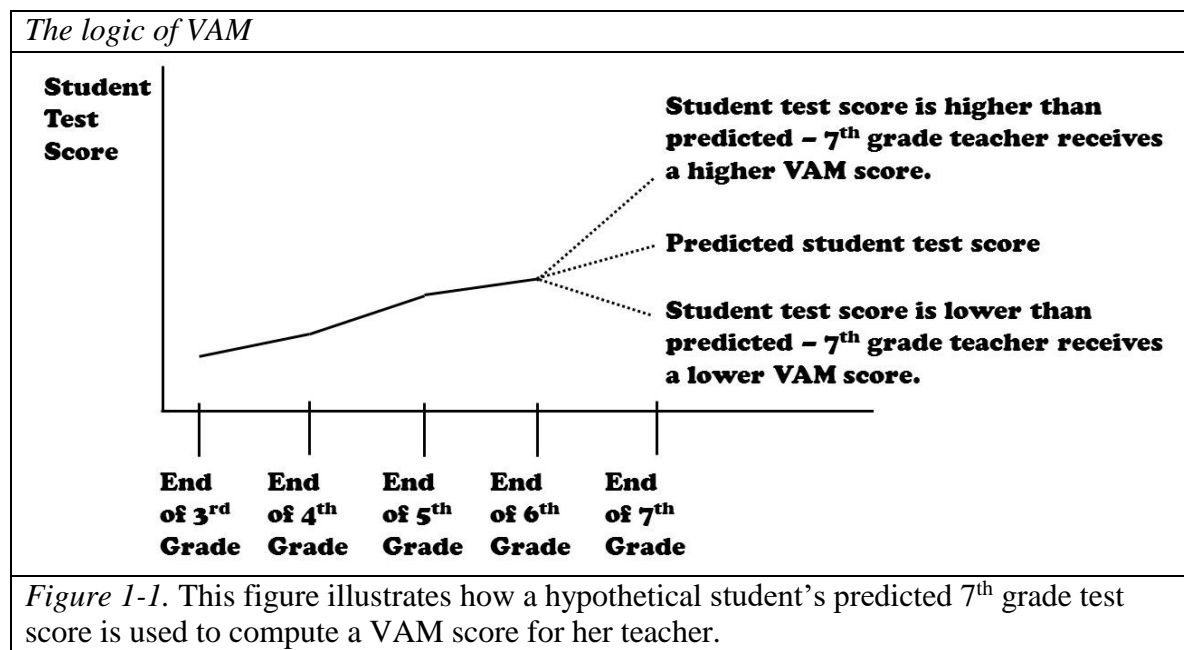
The idea here is to make realistic predictions so that teachers of more challenging students do

not unfairly receive lower ratings *simply because they teach more challenging students*. The

equity behind this growth-rather-than-proficiency logic is clear: it would be unfair to evaluate

teachers based on their students' end-of-year test scores, without taking into account where the

students began the year. This is the sense in which VAM attempts to "level the playing field" for

teachers.

A VAM score is interpreted as an estimate of the teacher's effectiveness, which becomes part

of the teacher's overall rating that is usually expressed on a four- or five-point scale. Washington

which covariates (e.g., socioeconomic status, race, special-education status, English language
proficiency) are included.

DC, for example, classifies teachers as ineffective, minimally effective, developing, effective, or highly effective (District of Columbia Public Schools, 2016).

VAM is an example of what I'll later refer to as an "information tool," a method or device that outputs information from which an inference is made about the natural or social world. Building value-added models and interpreting their results requires a level of statistical expertise beyond that of most nonscientists (including, importantly, most policymakers), for whom the methodological innards of the tool are opaque. This seems to suggest that—when it comes to technical questions about the capacity of the tool—the logical place to turn for answers is the experts who develop and use the models.

To be clear, for many (perhaps most) of the questions that arise around the use of VAM in teacher evaluation, it may be difficult to decide which stakeholder perspectives to listen to or which kinds of expertise to defer to. It is certainly not the case that those with expertise relative to VAM should have a sort of carte-blanche authority in this technical decision-making context. But the perspective of these experts is (I'll argue) the one we ought to listen to if we want to understand whether VAM is methodologically up to snuff. Put differently, these experts are uniquely qualified to answer the question, *can the tool deliver the kind of high-quality information necessary to make high-stakes decisions about individual teachers?* What I'll do in the next section is show the reader that federal and state policymakers who have been keen to use VAM in teacher evaluation do not seem to have heeded the perspective of these very experts, on this very question.

D. Using VAM in Teacher Evaluation: Policy and Expert Caution

1. Federal Education Policy and VAM

The United States Department of Education (USDOE) began to promote VAM as a gauge of teacher quality with Race to the Top (RTT), an enormous program of federal education grants. The introduction of this program in 2009 turned out to be a watershed moment in the history of teacher evaluation: "nothing in the past compares with the wave of value-added-based teacher accountability brought on by President Obama's Race to the Top" (Harris & Herrington, 2015, p. 71).

RTT was introduced as part of the American Recovery and Reinvestment Act after the economic downturn of 2007, a time when budgets were tight and states were eager to secure federal education funding. The program's goal was to spark innovation and reform in school improvement, and it contained suggestions in several key policy areas, including the adoption of common standards and the expansion of charter schools. Under the program, funding eligibility was contingent on there being no "legal, statutory, or regulatory barriers at the State level to linking data on … student growth … to teachers and principals for the purpose of teacher and principal evaluation" (USDOE, 2009, p. 4). Additionally, in their applications, states earned points if their teacher evaluation systems "take into account data on student growth … as a significant factor" (USDOE, 2009, p. 9) in determining compensation, promotion, retention, tenure, and dismissal.

While there is no explicit mention of VAM in RTT, there are two reasons to believe Secretary Duncan had this particular information tool in mind. First, the legislation mirrors, point-for-point, the reform agenda laid out in an influential policy blueprint published in 2006 titled, *Identifying Effective Teachers Using Performance on the Job* (Gordon, Kane, & Staiger).

This blueprint calls for "measures of outputs and performance rather than credentials" and stipulates that "some measure of 'value-added,' or the average gain in performance for students assigned to each teacher, would need to be a significant component of [teacher evaluation]" (p. 6). In 2014, a lead author of the blueprint (Robert Gordon) was appointed the Assistant Secretary for Planning, Evaluation, and Policy Development in the US Department of Education.

The second reason to believe Secretary Duncan intended states to include VAM in their teacher evaluation systems is implicit in the policy. To locate it, we need to do a little semantic unpacking of the term "student growth." The legislation stipulates that calculations of "student growth" *may* include changes in school- or district-level test scores or measures "that are rigorous and comparable across classrooms," but *must* include changes in "a student's score on the State's assessments under the ESEA [Elementary and Secondary Education Act]" (USDOE, 2009, p. 14).[2] Statistical best practice and common sense tell us that simply subtracting scores from subsequent years (without taking context or student characteristics into account) blurs too many important details and leads to unfair comparisons between teachers who teach different types of students. Researchers concur that the statistical machinery of VAM is preferable in this case to the simple subtraction of scores. What this means is that high-stakes decisions are to be made based on a teacher's evaluation, which must include outcomes in the form of a significant "student growth factor,"[3] which—in turn—includes a VAM score.

State Education Policy and VAM

---

[2] This is the definition that applies to students in grades and subjects where a statewide, annual, standardized assessment is used. In other cases, the USDOE suggests alternatives, most of which are also test scores: "alternative measures of student learning and performance such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments; and other measures of student achievement that are rigorous and comparable across classrooms" (USDOE, 2009, p. 14).
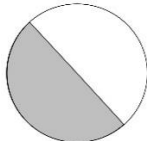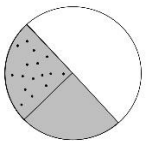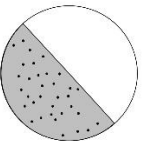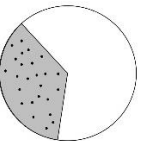
Different states have enacted the policy directives of RTT differently, due in large part to two

choices that must be made in implementation. These choices are illustrated in Figure 1-2 and

articulated here: (1) How much weight should VAM have in the "student growth factor" (i.e.,

how important should it be, relative to other RTT-approved components of "student growth,"

mentioned above)? (2) How much weight should the "student growth factor" have in order to be

considered a significant part of the overall evaluation?

The inclusion of outcomes in teacher evaluation: Choices facing RTT states

Figure 1-2. This figure illustrates two connected decisions that states must make in adopting the policies laid out in RTT.
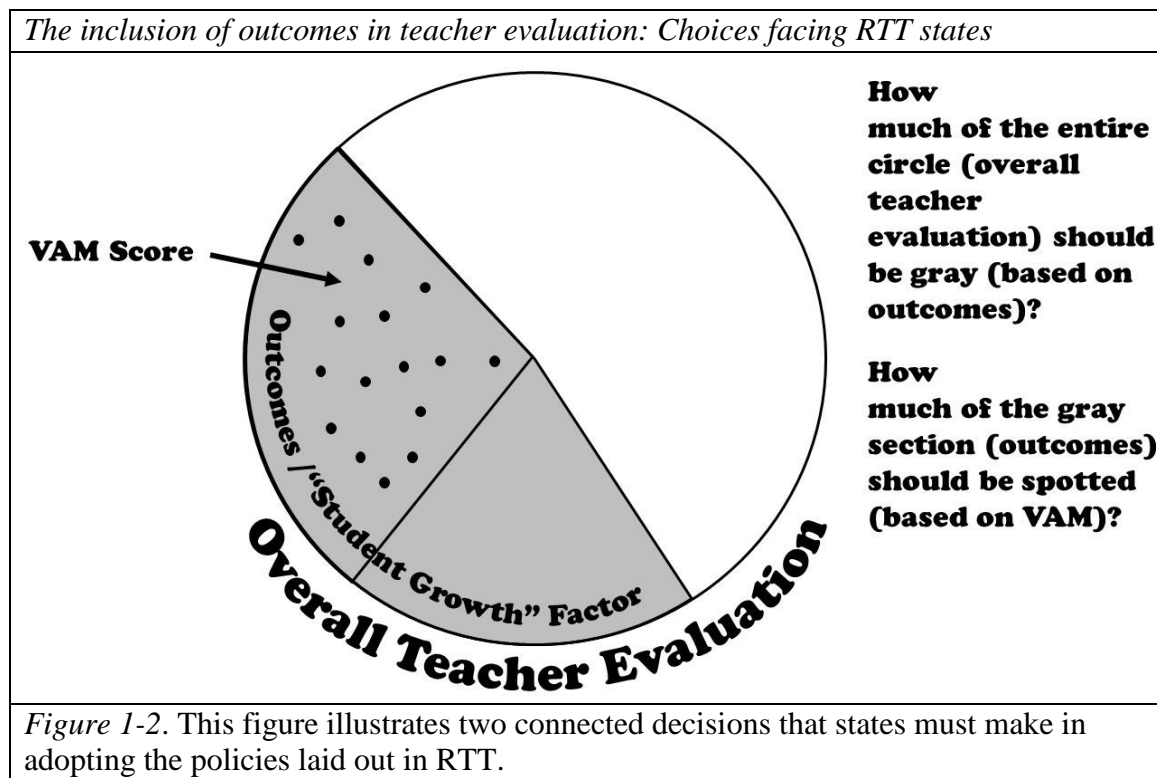
A comprehensive picture of how different states have answered these questions would be

nearly impossible to compile, as would an account of the particular stakes that have been

attached to teacher evaluations. Why? In addition to a great deal of state-to-state variation,

teacher evaluation policies are a moving target, changing drastically and quickly since the roll out

---

[3] The inclusion of "supplemental measures" is recommended as well; this "may include, for example, multiple observation-based assessments of teacher performance" (USDOE, 2009, p. 12) but student growth must be a "significant factor."

of RTT[4] (NCTQ, 2013). Table 1-1 below uses the graphical convention established in Figure 1-2 to give the reader a snapshot of the range of teacher evaluation systems that had been adopted in 2015.

State-to-state variation in the use of VAM in teacher evaluation				
Wisconsin	Hawaii	Louisiana	Idaho	Washington
Outcomes account for 50% of overall evaluation; VAM accounts for 0% of outcomes.	Outcomes account for 50% of overall evaluation; VAM accounts for 50% of outcomes.	Outcomes account for 50% of overall evaluation; VAM accounts for 100% of outcomes.	Outcomes account for 30% of overall evaluation; VAM accounts for 100% of outcomes.	Outcomes account for a "substantial factor" of overall evaluation; VAM accounts for 0% of outcomes.

Table 1-1. To illustrate how the role of VAM in teacher evaluation varies from state to state, policies from five states are summarized here. (NCTQ, 2015)

The many moving parts of RTT and its state-by-state interpretation are complex, but three simple points summarize what's important for my purposes:

(1) Over time, more states are requiring that outcomes (of which "student growth" is a part) be included in teacher evaluations. In 2009, there were 35 states that didn't include outcomes in their teacher evaluations; in 2015, this had shrunk to 7 (NCTQ, 2015). When states do include outcomes, most require or recommend that they count for half of a teacher's overall evaluation (Hull, 2013).

(2) Some states are including VAM in the "student growth" portion of their teacher evaluation systems. It has been suggested that this is most often done in such a way that

---

[4] Changes in state policy may also be attributed to the 2011 announcement by the USDOE that states would be eligible to receive waivers from the onerous sanctions of the No Child Left

VAM counts for between 30% and 50% of a teacher's overall evaluation (Goldhaber, 2015).

(3) In 2009, tying "decisions of consequence, such as tenure, dismissal or licensure advancement" (NCTQ, 2013, p. 1) to teacher evaluation was not a feature of *any* state's policy, but in 2015, evaluations were a part of dismissal decisions in 24 states (NCTQ, 2015).

2. The Case of Florida

Let's consider how VAM use evolved in Florida, a state that implemented RTT's policy suggestions in a strikingly ambitious way (Aldeman & Chuong, 2014). My purpose is not to suggest that this is a model for what's happening in most states, but rather to illustrate a use of VAM that typifies the central problem addressed in the proposed dissertation.

In 2009, VAM was not part of teacher evaluation in Florida (NCTQ, 2010). The state received a RTT grant in 2010, and by 2014, the two questions mentioned in Figure 1-2 were answered in the following way for about a third of Florida teachers (Florida Department of Education, 2014), those whose students take a statewide, standardized test for which a value-added model has been developed (Florida Department of Education, n.d.):

- How much weight should VAM have in the "student growth factor" of a teacher's evaluation? *VAM should make up the entire "student growth factor." In other words, a teacher's impact on "student growth" is synonymous with that teacher's VAM score.*

- How much weight should the "student growth factor" have in order to be considered a significant part of the overall evaluation? *The "student growth factor" should make up*

Behind Act (Polikoff & Porter, 2014), provided they used student test scores in the same ways stipulated in RTT (USDOE, 2012).

half[5] of the overall evaluation. The other half should be based on an "instructional

practice factor," measured—in part—by observations aligned to a district-determined

framework of instructional practice.

Decisions about how to use both halves of the overall evaluation to classify teachers as "highly-effective," "effective," "needs improvement / developing," and "unsatisfactory" were left up to districts. Once teachers were so classified, however, all districts were required to take the following actions (Mead, 2012):

- Notify parents if their child's teacher has been classified as "unsatisfactory" or "needs improvement" a certain number of times during the previous three years.

- Dismiss teachers if they have been classified as "unsatisfactory" or "needs improvement" a certain number of times during the previous three years.

- Base workforce reductions on evaluations, and specifically *not* base workforce reductions on seniority.

- Give principals the authority to refuse the placement or transfer of a teacher who is not classified as "effective" or "highly-effective."

- Establish a salary schedule with the greatest salary increments provided to teachers classified as "highly-effective" and no increments provided to teachers classified as "needs improvement / developing" or "unsatisfactory."

Personnel actions like this are serious. It is imperative they be based on good information, for the consequences of a mistake are grave, and could threaten the job security, earnings, and professional and social reputation of hundreds of thousands of teachers, including the

---

[5] When a teacher has less than 3 years of available data to calculate a VAM score, this is reduced to 40%, in recognition of the fact that VAM scores become less reliable with fewer years of data.

approximately 60,000 (NCES, 2015) Florida teachers whose evaluations depended on VAM scores in the years following RTT.

The reason I've sketched this example from Florida is to illustrate a particular use of the information tool of VAM, one that largely defines the professional competence – and could determine the professional trajectory – of teachers in Florida. While this use of VAM (as a central part of the evaluation of individual teachers for high-stakes purposes) is critically important to those teachers, it also has broader significance. It illustrates a national trend noted by many researchers. For example:

- "We suspect that such reforms [the development of educator evaluation systems in which one component is student performance on standardized tests] are here to stay and that test-based measures of teacher performance will be incorporated into teacher evaluation systems with increasing frequency" (Ballou & Springer, 2015, p. 77).

- "[S]ince the 2010 legislative session, over 20 states have enacted legislation focusing on educator effectiveness … An important feature of most of these reforms is the use of student achievement growth on standardized tests— henceforth I refer to this as "value added"—as one of multiple components that factor into a teacher's summative performance evaluation" (Goldhaber, 2015, p. 87).

The weight that VAM scores carry in Florida is high, compared to other states, but some believe the scores should have even more weight. A policy brief from the Center for Education Reform, an advocacy group that supports VAM-based teacher evaluation proposed in 2010 that principals be allowed

> to set individual teachers' salaries based on evaluations that are primarily (75 percent or more) based on a teacher's demonstrated impact on student achievement growth… Other factors that should be included in a teacher's evaluation and subsequent contract—but should not eclipse more than 25 percent of the outcome—are: a teacher's skills and knowledge, a

teacher's advanced responsibilities, [and] a teacher's willingness to mentor new teachers. (Center for Education Reform, 2010, p. 2)

3. Caution from the Scientific Community

We're now ready for the final piece of this narrative, the one that makes clear the substantive problem at the heart of the proposed dissertation. The scientific community has expressed strong caution about the central use of VAM in the evaluation of individual teachers for high-stakes purposes.[6] There is general agreement among educational researchers that the limited precision, consistency, and accuracy of VAM scores is cause for concern. See, for example, Baker et al. (2010); Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein (2011); Haertel (2013); Harris (2011); McCaffrey, Koretz, Lockwood, & Hamilton (2003); McCaffrey, Sass, Lockwood, & Mihaly (2009); National Research Council (2009); and National Research Council (2010). Moreover, professional associations have released statements opposing VAM-based teacher evaluation: American Educational Research Association (2015) and American Statistical Association (2014). Taken together, it is clear that the prevailing view within the scientific community is that VAM does not have the capacity to deliver the kind of high-quality information that would support the serious decisions being made with it.

E. Statement of the Problem and Research Plan

In light of this caution, the federal and state policy trends described in the previous section point towards something that might be described as "audacious use" – an information tool is being used to make important decisions or plans in a way that is unendorsed by the scientific community.[7] My selection of the adjective "audacious" plays on its dual meaning. First, a "lack

---

[6] Throughout the dissertation, I will refer to this policy as "VAM-based teacher evaluation."
[7] It may occur that relevant experts opine publicly on various other questions that surround the use of the tool (e.g., whether tool use is ethically palatable, whether tool use advances a

of respect" is shown for the methodological caution of those qualified to give it; second, "surprisingly bold risks" are taken when dubious information is used in a high-stakes way (Audacious, n.d.).

Audacious Use of an Information Tool (AUIT) is a curious phenomenon. Generally, we listen to experts when they speak on issues outside our technical expertise but within theirs. If an engineer tells us an elevator can carry 20 people, we don't put 30 people in it. If a doctor tells us that drug X and drug Y will produce effect Z (and if we want to avoid effect Z), we don't combine the drugs. To be sure, the boundaries of legitimate expertise are not always this clear, especially in science-society interactions where uncertainty is high and values consensus is low. But when it comes to technical questions like *what is the margin of error for this estimate?* or *what inferences are supported by the results of this analysis?* a decision not to heed scientific caution is odd.

The goal of the dissertation is to better understand this oddness, including its rhetorical dimensions, and four research questions orient this work. They're listed in Table 1-2 below, along with a brief description of what's involved in answering them.

---

political agenda, etc.). I want to be clear that I am not labeling as "audacious" a choice to disregard expert perspectives on these various other issues. This sort of "selective deference" to

Research questions and work of the dissertation	
1. What is AUIT?	The work involved in answering this question is conceptual. I synthesize the relevant STS literature on expertise and public decision-making and craft a definition that is consistent with current understanding and theory.
2. What are examples of AUIT?	The work involved in answering this question is empirical. After identifying a second potential case of AUIT (DTC genetic testing), I verify whether this case and the case of VAM satisfy my definition from above. That is, I check whether the cases involve information tools, high-stakes decisions or plans, and widespread expert concern about the methodological capacity of the tool.
3. What rhetorical moves rationalize the Audacious Use of VAM?	The work involved in answering this question is empirical. I construct a corpus of texts and, through rhetorical analysis, identify key rhetorical moves that may rationalize the Audacious Use of VAM.
4. Do these moves generalize to other cases of AUIT?	I approach this question in two ways. One way (involving logical and conceptual work) is to consider what a rhetor needs to deploy each move and to evaluate whether it is likely to be present in other cases of AUIT. The other approach (involving empirical work) is a preliminary study—not a comprehensive analysis—of the case of DTC genetic testing, where I look for examples of each move.

Table 1-2. The research questions that frame the dissertation are listed and briefly described.

F. Importance of the Dissertation

1. Importance in Academia

Much scholarship exists around the dynamics of public understanding and the ceding of

expert authority in technical decision-making contexts involving hard sciences. Relatively few

studies, however, consider social sciences and even fewer do so in an empirical way, as noted by

Mair, Greiffenhagen, and Sharrock (2013) in their survey of the literature. This gap in the

scientists will be discussed in detail in Chapter 2.

research points to the need for in-depth case studies like this one. The findings from case-study designs (which prioritize depth over breadth) are an essential part of abductive theory-building: they foster a level of precision and detail that, once established, may be applied, tested, or modified in other contexts and cases.

2. Importance in Education Reform

The technical decision-making context I've been describing for the past several pages (surrounding the central use of VAM to evaluate individual teachers) was the subject of a recent special issue of a prominent education research journal. From the introduction:

> The development of value-added methodologies and, in particular, their use to evaluate teacher effectiveness, has become an issue of intense interest and concern within the educational community. This interest is well placed… [T]he use of teacher value-added measures could have a greater influence on classroom instruction than perhaps any single reform in decades—for good and for ill. (Harris & Herrington, 2015, p. 71)

This observation from the editors captures the exceptional significance of this "policy moment" and underscores the importance of understanding and avoiding problematic or inappropriate uses of VAM.

3. Importance, More Broadly

AUIT is a potentially dangerous phenomenon. Economist Douglas Harris, who helped develop the value-added models used in New York City, underscored this in 2010 when he remarked (in reference to the use of these models in teacher evaluation), "as a general rule, you should be worried when the people who are producing something are the ones who are most worried about using it" (Otterman, para. 27). Nonexpert users of an information tool may overestimate what the tool can tell us. They may misinterpret its output, or assume the technical quality of the information is higher than it actually is. When high-stakes decision or plans are

made based on a faulty understanding of the inferences that are supported by an information tool, unintended, negative consequences may ensue.

AUIT is not only risky, it may be quite widespread. Information tools are commonly used by those unfamiliar with their methodological capacity. Often this use is endorsed by experts (e.g., personal breathalyzers, forensic DNA-matching techniques, over-the-counter pregnancy tests, etc.), but sometimes it's not and it elicits expert concern about low information quality and high decision stakes (e.g., forensic fingerprint-matching techniques, over-the-counter HIV tests, stock market prediction software, direct-to-consumer genetic tests, etc.). Furthermore, societal conversations about the appropriate use of these tools are often factious and involve a diversity of stakeholders and perspectives… meaning a voice of scientific caution could understandably be overlooked.

Given the danger and possible prevalence of AUIT, studying the phenomenon is important for two main reasons. First, doing so sensitizes us (as nonexpert users of tools) to the possibility of Audacious Use. If we're aware of AUIT, we're primed to listen for a voice of methodological caution from scientists. Second, if we understand how AUIT is legitimized, we may also learn ways to expedite its resolution or prevent it from occurring in the first place.

G. Limitations

1. Single-Case Research Design for Empirical Work

The empirical work on the language of AUIT is grounded in a single case, that of VAM-based teacher evaluation. When I consider the extent to which my findings might extend to other cases of AUIT, I draw examples from the case of DTC genetic testing, but the dissertation doesn't include a thorough rhetorical analysis of this case or any other. This is primarily because I'm writing as a single author, with training in a single science. So while I'm able to recognize

technically subtle rhetorical moves in the "science surrounding VAM" (i.e., quantitative education research), I don't bring the same scientific expertise to other potential cases of AUIT. This points to the need for future research analyzing the language of other cases of AUIT. And I believe that research will be best carried out by scientists who are willing to "jump the fence" (as I have done in this dissertation) from the scientific discipline in which they were trained to science-studies-esque disciplines like STS or science communication.

2. Rhetorical Analysis: Interpretation not Quantification

The purpose of rhetorical analysis is interpretive insight. The method reveals persuasive tactics (arguments, structures, topoi, etc.) that shape audience perception and understanding. One of the limitations of the method, however, is that it doesn't tell us about the frequency or prevalence of the particular tactics it identifies. For this reason, the methodology pairs well with a subsequent content analysis, where the primary purpose is quantification, not interpretation. This pairing was outside the scope of the dissertation, but it is a logical next step that could be used to investigate questions like, *which of the rhetorical moves identified in the study are deployed most frequently?* or *can we trace the origins of certain moves (e.g., to a particular speaker or event)?*

H. Organization of the Remaining Chapters

Chapter 2 addresses my first research question (What is AUIT?). In this chapter, I develop and define the concept of AUIT, based on (a) a survey of contemporary science-society interactions that feature a disconnect between expert and nonexpert perspectives and (b) a review of the Science, Technology, and Society literature and theory.

Chapter 3 addresses my second research question (What are examples of AUIT?). I establish that VAM-based teacher evaluation satisfies my definition of AUIT, and I locate another case: direct-to-consumer genetic testing.

Chapter 4 outlines the methodology and data I use to study the language of Audacious Use. I explain and justify the type of rhetorical analysis I conduct. I describe and justify the construction of my textual corpus from popular media, advocacy discourse, and policy messaging about VAM.

Chapter 5 presents the findings from my rhetorical analysis and thus answers my third research question (What rhetorical moves rationalize the Audacious Use of VAM?). I introduce four moves that rhetors deploy in arguments about VAM and describe their likely effect on the reading or listening audience.

Chapter 6 concludes the dissertation with a reflection on its findings. I address my final research question (What rhetorical moves rationalize AUIT?) by discussing (1) whether the moves identified in the case of VAM seem to be present in the case of DTC genetic testing and (2) whether we should expect to see these moves in cases of AUIT, in general. I discuss the implications of Audacious Use for scientists and nonscientists, and suggest directions for future research.

II. Chapter Two: Audacious Use of an Information Tool

A. Roadmap for the Reader

A great deal of Chapter 1 was spent discussing Value-Added Modeling (VAM), and this was in order to motivate—by way of example—the central concept of the dissertation, Audacious Use of an Information Tool (AUIT). This chapter will take us away from VAM in order to define AUIT as a "type" and situate it in the STS literature. After giving the reader a sense of what to expect in this chapter, I provide formal definitions of "Information Tool" and "AUIT." The remainder of the chapter maps AUIT onto the theoretical STS landscape, explaining its relationship to classic conceptual issues like the problem of extension, deference to expertise, and science's embedded subjectivity.

B. The Kind of Scholarship This Is

In order that the reader understand where this chapter belongs in the literature, let me explain the *kind* of research I'll present here. A common move in STS scholarship is to describe and interpret case studies, where the unit of analysis is a particular interaction between science or technology and society. Examples include Paroske's (2009) examination of AIDS policy in South Africa or Banning's (2009) study of public sense-making about global warming. Empirical studies like these give us rich, thick descriptions of real-world technical decision-making contexts. This chapter engages in a different kind of STS scholarship, "type-naming." That is, describing a commonality across cases, naming it, and thus generating a new, aggregate unit of analysis that we might call a "type." Ceccarelli (2011), for example, noted the common pattern of consensus, controversy, and public understanding in cases like Paroske's and

Banning's and gave us a name for it: Manufactured Scientific Controversy (MSC).[8] The type I

will name in the following pages, AUIT, is another curious phenomenon of public decision

making.

C. What I Actually Did to Write This Chapter

Like much type-naming work, this study began with an exploratory survey of cases of public

decision making. In particular, I surveyed cases where an application of science was

accompanied by a disconnect between expert and nonexpert perspectives. I considered, for

example, the use of mammograms in presymptomatic breast cancer screening (e.g., Lerner,

2003), the presence of genetically modified organisms (GMOs) or nanotechnologies in the food

industry (Schurman & Munro, 2010), and the forensic use of fingerprint- or DNA-matching

techniques (Byers & Johnson, 2009).

In most of the cases I surveyed, expert and nonexpert perspectives seemed to differ on

questions that were value-laden and multi-dimensional (e.g., do the benefits of early breast

cancer screening outweigh the potential risks?), making a rationale for deference to one

perspective or the other predictably problematic. But a few cases stood out. They seemed to

involve a choice not to defer to expert caution when there was a clear rationale to do so: *those

familiar with the tool's methodological capacity were questioning whether it could actually do

what it was being asked to do*.

This research followed the principles analogical theorizing (Vaughan, 2014) and abductive

analysis (Tavory & Timmermans, 2014). As I moved from case to case, I toggled constantly

between emerging theory and what I saw in the data, iteratively refining my understanding of

---

[8] Other examples of type-naming work include Proctor's (2008) "Agnotology" or
Freudenburg, Gramling, and Davidson's (2008) "Scientific Certainty Argumentation Methods."

Audacious Use. The analysis concluded when I arrived at a definition that adequately described what was essential and noteworthy about the phenomenon.

D. Information Tools

Information Tool: NOUN

an analytical method or device, developed and used by scientists, that outputs information from which an inference is made about the state of the natural or social world

Information tools are everywhere: thermometers, litmus paper, breathalyzers, glucose monitors, pregnancy tests, statistical models used in economic projections, etc. What is important about the above definition is that information tools *provide information about* (rather than *change*) the state of the world. Technical decision making around "change tools" (e.g., genetic modification of food organisms, vaccinations, fluoridation of drinking water, etc.) is, of course, interesting and worth studying, but it's outside the scope of this dissertation.

An information tool is like a little conceptual factory. It takes in data or materials and spits out information about them. The user of the tool examines the output and concludes something about the state of the world. In some cases, this inference follows immediately from the tool output and is very straightforward: the litmus paper turns red and the user concludes that her material is acidic. In other cases, the inference follows immediately from the tool output, but interpreting that output requires some expertise: most laypeople don't know what to make of mammogram film, but a technician can quickly conclude the presence or absence of suspicious breast tissue. In still other cases (and these will turn out to be the interesting ones), the inference is predicated on more than just tool output; it requires additional knowledge or skills (e.g., which regression coefficients in this table are meaningful and how are they related to one another?) or

contextual information (e.g., what do we know about the correlation between this particular pattern of genes and this disease?). These more interesting cases mirror the way scientists typically use information tools.

When a scientist wants to describe, explain, or predict, she engages in what we might call a "research endeavor," where her information tool is just one of many moving parts. Other moving parts include research questions (which bound and define the problem), theory and existing literature (which surround and orient the problem), research designs (which specify what data to collect), and strategies for collecting those data. This is illustrated below, first in Figure 2-1 and then with a hypothetical example from educational research.

The role of information tools in generating scientific statements

Figure 2-1. A scientist observes the information outputted by an information tool, considers the context of the information (i.e., the overall research endeavor), and infers something about the state of the world.
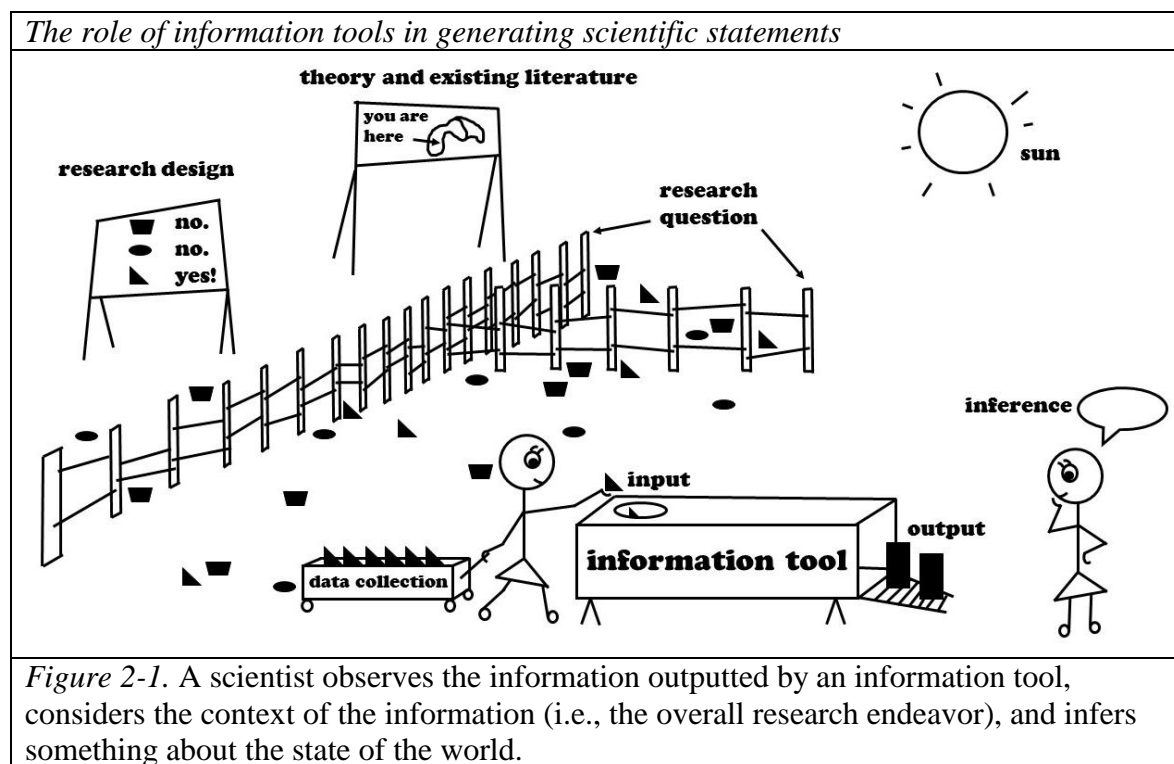
A mean is a very simple information tool. Thought of as a factory, it "takes in" a set of numbers (let's say, the number of correctly answered questions on a history test for a group of students) and "spits out" a single number (say, X). The inferential distance between this number

and the following conclusion is negligible: "on average, these students answered X% of the questions correctly."

Now, if our researcher is aiming for a substantively more interesting description of the world (perhaps, "on average, these students have mastered X percent of grade-level history content"), then the inferential distance increases, as she must consider other elements of the research endeavor like how the test was scored ("data collection," in the figure above) and the alignment between the test and the grade-level curriculum ("existing literature," in the figure above).

If our researcher is interested in explaining something about *how* history students learn, then she might venture into the realm of causal inference. Doing so will increase yet again the inferential distance between tool output and her conclusion because valid causal claims depend not only on tool output, not only on information about measurement and curriculum, but on additional issues of data collection and research design.

Consider if our simple information tool is used in the context of a study investigating the relative effectiveness of virtual and traditional history textbooks. We can imagine splitting a group of students and having some of them use a virtual text and the others use a traditional one. In this case, our researcher would examine what the factory spits out: X *and* Y this time, for the "virtual group" and the "traditional group," respectively. If X is less than Y, she may infer something like "virtual textbooks are less effective learning supports than traditional ones." The tool output seems to support this conclusion, but consider two other elements of the research endeavor that lie, epistemologically speaking, in the distance between the output and the inference.

The first is an issue of data collection: which students participated in the experiment? This matters because it indicates how far we ought to generalize our results. If, for example, all the students in the study were non-native English-speakers, then we've learned something about

textbook effectiveness for these students, but perhaps not for native speakers. The second issue is one of research design: how were students assigned to the "virtual group" and the "traditional group"? What's important here is whether these groups were different in some respect other than the type of textbook they used. If, for example, the "virtual group" was taught by a first-year teacher and the "traditional group" was taught by an experienced teacher, then the difference between X and Y may reflect the difference in teachers, not the difference in textbooks.

The thing to notice here (and what's conveyed in the figure above) is that often, output from an information tool is just one part of a scientist's inference. The statement she ultimately makes is predicated on a holistic understanding and synthesis of *all* the elements of the research endeavor.

E. Audacious Use of an Information Tool

Audacious Use of an Information Tool: NOUN

use of an information tool to make high-stakes decisions or plans against the advice of relevant experts, among whom there is agreement concerning the technical inadequacies of the tool for the stated use

Notice that AUIT is a description of what was (or is) happening at a particular historical moment. The audacity of a certain use of a certain tool might change over time, if decision stakes change or if caution from experts changes. For this reason, I will specify a date for the two cases of AUIT I will analyze in the next section. First, however, let me address some questions this definition may have raised for the thoughtful reader.

1. What do you mean by "technical inadequacies"?

In this dissertation, "technical inadequacies" or "technical concerns" will refer to one or more of the following issues:

- The **precision** of tool output (i.e., does the tool return exact information, as opposed to a ballpark estimate?)

- The **consistency** of tool output (i.e., if we use different versions of the tool, or use the tool on different occasions, do we get the same information?)

- The **accuracy** of tool output (i.e., does the tool return correct information?)

2. Who are these "relevant experts"?

The term, "relevant experts" begs the question, *relevant to what?*, so let me explain how I'm using this term. The use of an information tool is logically (though perhaps not explicitly) preceded by the question, *can this tool deliver the information it is being asked to deliver?* By "relevant experts," I mean those whose answer to this question would draw on an understanding of the methodological details of the tool, including its technical properties, the types of research questions and inferences it supports, the research design in which it is being used, etc. These relevant experts are familiar with the tool's development, typical performance, and limitations. Collins's (2004) concept of "contributory expertise" is helpful here: those who could, for example, "reasonably apply for a job in the science in question or at least publish papers in the professional journals or perhaps be let loose in the laboratory" (p. 128).

Notice that in naming AUIT, I am foregrounding the perspective of these relevant experts *only when it comes to the methodological capacity of the tool relative to the decision stakes*. I am describing a choice to disregard this expert caution as "audacious," harkening to the dual meaning of this adjective. First, a "lack of respect" is shown for the methodological caution of

those qualified to give it; second, "surprisingly bold risks" are taken when dubious information is used in a high-stakes way (Audacious, n.d.). Now, it may occur that relevant experts opine publicly on various other questions that surround the use of the tool (e.g., whether tool use is ethically palatable, whether the pros of tool use outweigh the cons, etc.). I want to make it clear that I am not labeling as "audacious" the choice to disregard experts' perspectives on these various other issues. To my mind, there is nothing audacious (or particularly interesting) about this choice, in the same way that there's nothing audacious about tuning out your mechanic when she opines about the stock market.

3. When you say, "advice of relevant experts," how much of the scientific community does this entail? Must there be consensus that tool use is inadvisable?

The relevant experts advising against tool use should include a significant contingent of the scientific community, though there might not be complete consensus. This is because expressing caution involves making a judgement call about how serious the technical concerns are relative to the decision stakes, and some amount of expert disagreement here (i.e., a set of experts who acknowledge technical concerns but still endorse use) does not jeopardize what's interesting about AUIT: a central reaction from the scientific community is one of caution, and yet use proceeds anyway.

Different AUIT researchers might operationalize the phenomenon in different ways. One approach would be a content analysis of the peer-reviewed literature, similar to what has been done to establish the extent of disciplinary agreement about anthropogenic global warming (e.g., Powell, 2015). Another approach (the one I will take in the next chapter) is to review relevant position statements issued in response to tool use by professional organizations of relevant experts. I have intentionally not stipulated a threshold for what constitutes "enough" expert

29

advice against tool use (e.g., caution must be expressed in "more than one position statement" or "more than a third of the peer-reviewed articles that mention tool use"). What constitutes a meaningful threshold will vary from case to case. Generally speaking, though, we might say that a case of AUIT becomes more "robust" as the caution approaches a level of consensus among relevant experts.

F. Audacious Use among the Pigeons of Science Studies

1. The Problem of Extension

My definition of AUIT brings up issues of authority and participation in technical decision-making contexts, and these are the issues "where the pigeons of much recent social science are coming home to roost" (Collins & Evans, 2002, p. 235). Decades of disciplinary effort focused on the "inside" of science, examining the nature of scientific knowledge and asking *what do scientists know and how do they come to know it?* One of the things this work revealed was the value-laden-ness of science, and this served to weaken the distinction between scientific and nonscientific knowledge. Collins and Evans (2002) explain that this has raised a fundamental question about when to defer to scientists. If white-coated objectivity is a myth, if a scientific perspective is no longer the epistemological gold standard it was once taken to be, if participation in technical decision making should extend to include a nonscientist perspective… just how far should it extend? The Problem of Extension is that, absent a clear boundary between science and nonscience, "there are no longer any grounds for limiting the indefinite extension of technical decision-making rights" (Collins & Evans, 2002, p. 235).

This is an enormous problem. I've sketched it not because I plan to solve it, but rather because it's a useful way to locate AUIT in the science studies literature. AUIT is a very peculiar kind of technical decision-making context, one where we can see the Problem of Extension in

action and where—I'll argue—we can actually make some headway figuring out whether, when, and why to listen to scientists.

2. "Audacious"?

The principle audacity in Audacious Use is committed by nonscientist[9] policymakers or users of the tool: they show a "lack of respect" for the caution from the scientific community. In selecting this wording, I am advocating (albeit vaguely) a certain esteem for the scientific perspective. In a disciplinarily unfashionable way, I am advocating for the scientist, not the downtrodden sheep farmer (Wynne, 1989) (more on this to come).

How can I do this without ignoring decades of work in the Sociology of Scientific Knowledge (SSK) that have documented grave trouble with such deference to science? In the next section, I'll acknowledge three sources of this trouble, but let me first briefly preview for the reader where we'll go after this: my argument for deference to science hinges, ultimately, on the *very specific type of scientific statement* involved in AUIT. My scientists don't have carte blanche to ride roughshod over my nonscientists, and my nonscientists act audaciously only when they disregard *this particular type of scientific statement*. There are certain circumstances, I'll argue, (i.e., those involved in AUIT) where a deference to science is appropriate. Before discussing these circumstances, however, let's recall why a great deal of what scientists say may *not* warrant deference and ought to be taken with an SSK-approved grain of salt.

---

[9] In principle, there is no reason that information tools couldn't be used audaciously *by scientists*, in the face of caution from *other* scientists. If both the using-scientists and the cautioning-scientists possess contributory expertise, we might describe this as "scientific controversy." If the using-scientists have relatively less expertise about the tool, we might describe this as a sort of "within-science AUIT." Instances like these may be important flashpoints for the study of expertise, but they don't directly involve the public (i.e., they're not "technical decision-making contexts"), so they're outside the scope of this dissertation.

3. A Habit of Deference: Three Troubles

If we divide the participants in technical decision-making contexts into two groups, scientists and nonscientists, then roughly speaking, the Problem of Extension is about which group we should listen to. For a long time, societal norms commonsensically suggested that when science is involved, we should defer to scientists (Briggle, 2008), reflecting an implicit trust in the superiority of their understanding: "we stare, the scientists see; we gawk, they gaze. We guess; they know" (Gopnik, 2015). Hardwig (2006) argues that, given our society's increasing technological sophistication and the wealth of specialized knowledge that bears on public decision making, deference to experts can make rational sense: "one can have good reasons for believing a proposition if one has good reasons to believe that *others* have good reasons to believe it" (p. 328). This "epistemic dependence," Hardwig suggests, is a key way of avoiding what Frankfurt (2005) refers to as "the production of bullshit," which "is stimulated whenever a person's obligations or opportunities to speak about some topic exceed his knowledge of the facts that are relevant to that topic" (p. 63).

There are three potential sources of trouble with this habit of deference. The first (discussed in the next section about sheep) is concerned with missing out on pertinent information from nonscientists. The second and third (discussed in the subsequent sections about the Science Wars and "big-P politics") are concerned with a misguided faith in the objectivity or purity of information from scientists.

4. Trouble One: Not Listening to Sheep Farmers

In his canonical case study, Wynne (1989) examined the interactions between scientists and Cumbrian sheep farmers and their different approaches to problem solving in the environmental aftermath of the Chernobyl disaster. One of the things Wynne showed was how information

from nonscientists can be more relevant and accurate than information from scientists. The

sheep farmers, drawing on their knowledge of local geography and grazing habits, pointed out a

fundamental flaw in the design of an experiment to measure levels of radioactivity in the soil.

The scientists disregarded this information, proceeded with the experiment, and then quietly

abandoned it when the farmers' concerns proved correct (Wynne, 1992). This and similar studies

underscore a potential problem with a habit of deference to the scientific perspective: limiting

authority and participation in technical decision making to those with certified expertise (i.e.,

scientists) means we risk overlooking the "lay expertise" of nonscientists.

5. Trouble Two: Listening to Scientists and the Science Wars

The second problem with a habit of deference to experts in public decision-making lies in the

answer to the question, *what is a scientist really doing, epistemologically speaking, in the*

conduct of her science? This question lies at the heart of the Science Wars between, roughly

speaking, positivists on one side and relativists on the other. They're divided about the presence

(and / or importance) of the "human element" in the work of scientists (i.e., the social, ethical,

and political values of the scientist and of the community of scientists in which she works).[10]

A staunch positivist would hold that the scientist is *discovering an objective reality* and her

work reflects that reality. Of course there are more and less principled ways to do this, and "the

classic positivist ideal is that researchers … should be neutral, nonpartisan students and

commentators on the issue under study" (Martin & Richards, 1995, p. 9). But at the end of the

day, assuming she keeps the human element at bay (i.e., she sets aside her own values and she

---

[10] I'm painting with intentionally broad brush strokes here, not because the nuances of the Science Wars are unimportant, but because this degree of nuance allows me to make key points about the epistemology of AUIT while still maintaining momentum in the overall argument of the chapter.

shields her work from social and political influences of the scientific community), the scientist is discovering and giving an account of "the world out there."

A relativist would maintain that a scientific account is less a direct reflection of nature, and more a product of social processes in the scientific community and subjectivities in the scientist. Philip Kitcher (1993) summarizes this perspective when he writes, "the deep point of the sociological critique is that the social forces that operate in this modification of practice – the rules for consensus shaping, the conversations with peers, the training process and broader socialization within a larger community – may be sufficiently powerful that the effects of nature are negligible" (p. 162). In other words, the scientist is not discovering an objective reality, but rather *creating a social reality* within the structures, customs, and discourse norms of the scientific community. SSK scholars tend to embrace this perspective, and one of the canonical pieces of SSK scholarship is the work Latour (1987) did opening the "black box" of biochemical research and helping us see the human element inside of science.

Let me be more precise about what I mean by the "human element." It makes for a dramatically satisfying story to think of the human element as a concerted, nefarious effort on the part of scientists to smuggle their personal agendas into their work and thereby exert their secret will on an unsuspecting public. Sure, this can happen, but it's overly simplistic: a straw man account of the problem that concerns relativists. Another part of the problem—a more interesting and conceptually challenging one—is that the human element can shape science in much more subtle ways, *often unbeknownst to the scientist*. It is present in the very fabric of the scientific process: in the language and framing of research questions (Longino, 1990); in the description and interpretation of data (Heider, 1988); in the cognition of evaluative judgements (Wilson, DePaulo, Mook, & Klaaren, 1993); and in the way scientific tools shape conceptualization of and intervention in the subject of research (Levin, 2014). If one finds the

relativist perspective at all compelling, then deference to a scientific perspective is potentially troublesome, for even when a scientist speaks on scientific topics, her social, ethical, and political values (and those of the community of scientists in which she works) are embedded in what she says.

6. Trouble Three: Listening to Scientists and Big-P Politics

I'll discuss one final problem with a habit of deference to a scientific perspective: members of the scientific community are also members of the public. They have opinions and beliefs about social issues, about wise and unwise policy moves. Shapin (1979) documents how these "big-P politics" (Collins & Evans, 2002, p. 245) can shape research agendas, questions, and findings, and Walsh (2010) notes just how harshly the public judges those scientists who use their scientific identity to further their political proclivities. The challenge for a scientist is how to reconcile her political identity (i.e., her concern with big-P politics) with the norms of her discipline that prize cooperation, disinterestedness, skepticism, universality, and humility (Merton, 1973). Pielke (2007) suggests four ways to act as both a scientist and a citizen, four archetypes for integrating these identities, (e.g., "the stealth issue advocate" or "the honest broker of policy alternatives"). I mention all of this work to underscore the fact that separating science from big-P politics is not straightforward, and so when we defer to scientists, we run the risk of deferring (perhaps inadvertently) to their big-P politics as well.

7. Deference to Science in the Particular Scenario of AUIT

We ought to listen to Wynne's sheep farmers. We ought not listen blindly to scientists. This appears to leave my call for deference to scientists (implicit in my choice of the word, "audacious") on shaky ground. For two reasons, however, the ground beneath AUIT is actually quite solid. First, I'm calling for selective listening. A great deal (perhaps most) of what

35

scientists say *doesn't* warrant deference; I only suggest that we heed those statements that pertain to a tool's methodological capacity. A caution against tool use based on low information quality (i.e., "what the tool can and can't tell us") is very different from a caution on political or ethical grounds. The former, arguably, contains much less of the human element than the latter. The second reason the ground beneath AUIT is solid is that it doesn't propose that we defer to anyone waving a "science badge," just those who are particularly close to the tool.

8. Close to the Tool: Scientists with Contributory Expertise

The definition of AUIT refers to "relevant experts," by which I mean those scientists with "contributory expertise." This term is taken from the work of Collins and Evans (2002) differentiating types of expertise (relative, of course, to a particular chunk of science). They describe those with *no expertise*, whose understanding of the science at hand is negligible or nonexistent. There are also those with *interactional expertise*, who cannot conduct the science themselves, but have "enough expertise to interact interestingly" (p. 254) with "those who actually do [the science]" (p. 244). This last group is said to have *contributory expertise*.[11] In order to describe the expertise types of various participants in a science-related policy context, it's necessary to be clear about the particular science relative to which expertise will be considered: someone with contributory expertise in the science of hierarchical linear modeling, for example, could have no expertise in developmental social psychology.

For the purpose of illustration, let's apply these types to the case of VAM-based teacher evaluation described in Chapter 1. The science at hand is the one to which the tool of VAM belongs, namely the science of statistically accounting for variation between classrooms in

---

[11] It's important to note that "these three categories are ideal types and, as with most such classifications, there will be boundary problems" (Collins & Evans, 2002, p. 255). I do not

patterns of change on standardized tests. A geophysicist with no interest in education policy would have *no expertise*. A district administrator charged with setting teacher-evaluation policy might have *interactional expertise*. A quantitative education researcher trained in the development and interpretation of student growth percentile models (which may feature in VAM) would likely have *contributory expertise*.

At the risk of sounding tautological, questions about the workings of an information tool cannot be answered in a meaningful way by those who are unfamiliar with the tool. Put differently, it only makes sense to pose these questions to someone who understands what the tool does, what it takes as inputs, and what it returns as outputs… i.e., someone with contributory expertise. This isn't a case of disregarding the views of sheep farmers in a discussion about grazing patterns because of an institutionalized or hegemonic refusal to recognize their "experience-based expertise." Rather it's a case of disregarding their views in a discussion about the reliability of results from Southern blotting (a method used frequently in cloning research for detecting particular DNA sequences) because they don't have anything meaningful to say on the topic. Should the discussion turn to, say, the implications of sheep cloning for the sheep farming trade or the ethical complexities of cloning in general, then overlooking the voice of farmers would be problematic. But AUIT isn't about such potential turns in the conversation; it concerns the very specific scenario when the methodological capacity of an information tool is in question.

9. The Challenge of Demarcation

A great deal of the chapter thus far could be summarized like this: "we ought to listen to scientists *when they talk about the tools of science*; if we don't, we're acting audaciously."

---

intend, in this dissertation, to work on these boundary problems. I will simply take as a starting point that these three types are a useful way to talk about expertise.

Ultimately, my argument hinges on efforts at "demarcation,"[12] a term I'll use to encompass attempts to distinguish between "facts and values" (Weber, 2002)… between "the *is* and the *ought*" (Hume, 2004) that might be present in the scientific endeavor. This work involves brokering an "uneasy divorce" (Kasanmoentalib, 1996, p. 44) between objectivity and subjectivity in science. Broadly speaking, those disinclined to do this work tend to cluster around one of two conceptual poles (Kitcher, 1993), summarized below and illustrated in Figure 2-2.
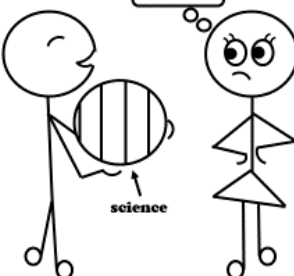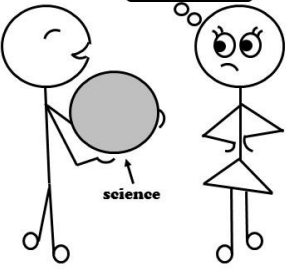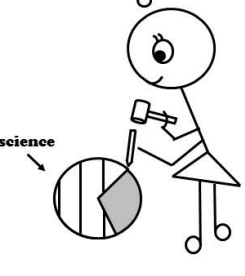
Locating the human element in science		
The old image of scientific rationality	**The Strong Programme**	**The challenge of demarcation**

Key: ▭▭▭▭ = objectivity ▬▬▬ = subjectivity

Figure 2-2. Two classic responses to the challenge of demarcation (along with their rationale) are presented. The author's response is depicted in the final column.

The "old image of scientific rationality" (Mayo, 1991, p. 252) holds that science comprises impartial algorithms used to adjudicate between competing hypotheses, and that metaphysical beliefs, goals, and subjective interests lie securely *outside* the process of science. In other words,

---

[12] This is my term. Others have described this as "boundary work" (e.g., Jasanoff, 1987) or "appraising objectivity" (Mayo, 1991). I find the first term too vague for my purposes, and I'm uncomfortable with the way the second implicitly privileges objectivity over subjectivity, which may be appropriate in some decision-making contexts, but certainly not all. "Demarcation" connotes an equal concern with both objectivity and subjectivity.

demarcation work is largely unnecessary because science consists of facts: there's no need to ferret out the embedded human element because it's negligible.

The second (and, in a sense, completely opposite) rationale for eschewing demarcation work can be seen in the Strong Programme. This particularly salient form of relativism, which treats scientific understanding exactly the same as other ways of understanding (Martin & Richards, 1995), holds that science consists primarily (if not exclusively) of the human element. Thus demarcation efforts to cordon off the human element are trivial: if we separate out the human element, there would be precious little "science" left over.

As a trained scientist working in science studies, I find both of these conceptual poles unsatisfying. The first is naïve. In the practice of my discipline, I see first-hand the ways that social, ethical, and political values shape what I and my colleagues call "facts" or "findings." On the other hand, I find the capitulation of the Strong Programme a regrettable alternative. While this approach protects us from values that might be nestled into the work of scientists, the price is too high: we throw out the baby of scientific insight with the bathwater of the scientist's subjectivity. Finding footing in the "vast, unexplored middle ground between these extremes" (Kitcher, 1993, p. 164) is the challenge of demarcation. By defining AUIT in this chapter, I am asserting that the challenge is not insurmountable.

10. What Sort of a Thing is AUIT?

Technical decision-making is fascinating. It pulls together a variety of stakeholders, with a variety of ways of knowing about the world, working towards a variety of purposes, using a variety of discourses. These societal phenomena are like flashpoints for a great deal of research in science studies, public understanding of science, science communication, rhetoric of science, etc. What I've done in this chapter is a common move in this body of research: I've established a

type. That is, I've pointed out and named a particular thing that might happen in public decision making involving a particular chunk of science (i.e., an information tool). In this sense, AUIT sits alongside "Manufactured Scientific Controversy" (Ceccarelli, 2011) over climate change or "Fear Stoking" (Ferdman, 2015) in the case of GMOs. These types are useful in and of themselves because they can deepen our understanding of the recurring patterns of conflict, understanding, language use, etc. that characterize public decision making. They're also natural starting points for research: if these patterns are problematic, then describing and studying them sets us on a path towards improving them.

III. Chapter Three: Two Cases of AUIT

A. Roadmap for the Reader

In Chapter 2, I defined AUIT as a "type" and situated it in the STS literature. This chapter applies that definitional work and establishes its utility by identifying two instances of AUIT: the case of direct-to-consumer (DTC) genetic testing from 2007 to 2012 and the case of VAM from 2008 to 2015.

B. The Use of Predictive Genetic Testing to Make Personal Health-Care Decisions (2007 - 2012)

Using an information tool called predictive genetic testing, genetic epidemiologists study the relationship between disease risk and patterns of genes. In 2007, a handful of private companies had begun to use this tool to produce commercially available "genetic risk profiles" (Hunter, Khoury, & Drazen, 2009), which purportedly told customers how likely they were to develop serious health conditions like heart disease, asthma, diabetes, and cancer. The idea was that, empowered with this information about their future health, customers could make important decisions about their present health (e.g., changing diet or exercise, seeking follow-up medical tests or interventions, limiting environmental factors that contribute to disease risk, etc.). In 2008, DTC genetic tests were hailed by *Time Magazine* as the "Invention of the Year" (Hamilton, 2008).

As more companies began marketing these genetic risk profiles, scientists began to express serious reservations about this use of predictive genetic testing. They were concerned, in particular, with instances when the low information quality of the tests wasn't conveyed to consumers or when consumers were left to their own devices to interpret the reliability,

41

precision, or accuracy of risk estimates. By 2012, these reservations had coalesced into position statements (issued by prominent professional organizations), all expressing concerns about the tests (Uhlmann & Sharp, 2012). To verify that this was a case of AUIT, I will establish that these concerns pertained to the tool's technical inadequacy and that they were unrefuted. The following analysis requires a rudimentary familiarity with the methodological details of predictive genetic testing, so I will begin with a one-paragraph primer.

A single nucleotide polymorphism (SNP) is a well-known genomic address: a location in our DNA where different individuals might have different alleles. Predictive genetic testing begins with a sample of an individual's DNA. The researcher decides which SNPs to consider and then uses a microarray to determine the alleles that are present at those SNPs, for that individual. Next, the researcher refers to published association studies, which compare the prevalence of particular alleles (at particular SNPs) between individuals with a genetic condition and individuals without it. Based on these studies, a risk estimate (the tool's output) is computed: the tested individual's likelihood, relative to the general population, of developing the genetic condition. At this point, the tool output is handed off to the consumer, who makes an inference (about future health) and decisions (about present health care).

In order to establish that in 2012 DCT genetic testing constituted a case of AUIT, I will draw on the work of Skirton, Goldsmith, Jackson, and O'Connor (2012), who conducted a systematic review of position statements issued about DTC genetic testing. The authors found eight such statements, issued by professional organizations whose membership includes professionals familiar with the methodological details of predictive genetic testing: the American College of Clinical Pharmacology (Ameer & Krivoy, 2009), the American College of Medical Genetics (American College of Medical Genetics [ACMG], 2008), the American College of Obstetricians and Gynecologists (American College of Obstetricians and Gynecologists [ACOG], 2008), the

American Society of Clinical Oncology (Robson, Storm, Weitzel, Wollins, & Offit, 2010), the

American Society of Human Genetics (Hudson, Javitt, Burke, & Byers, 2007), the European

Society of Human Genetics (European Society of Human Genetics [ESHG], 2010), the

International Society of Nurses in Genetics (International Society of Nurses in Genetics

[ISONG], 2009), and the National Society of Genetic Counselors (National Society of Genetic

Counselors [NSGC], 2007). These statements reflected the following technical concerns about

the methodological suitability of the tool to the proposed use:

- Risk estimates are inconsistent over time. Estimates will likely change as more

  association studies are conducted and scientists learn more about the relationship between

  SNPs and conditions. (Ameer & Krivoy, 2009; ESHG, 2010; Hudson, Javitt, Burke, &

  Byers, 2007; Robson, Storm, Weitzel, Wollins, & Offit, 2010)

- Risk estimates are inconsistent from company to company. Different companies make

  different methodological choices (which SNPs to consider, which association studies to

  consult, which statistical techniques to use in the calculation of risk estimates), which can

  lead to different estimates. (Ameer & Krivoy, 2009; ESHG, 2010; ISONG, 2009;

  Robson, Storm, Weitzel, Wollins, & Offit, 2010)

- Risk estimates are inaccurate. Estimates are based on a very nascent body of scientific

  knowledge, and the relationship between SNPs and hereditary conditions is complex.[13]

  Variation at each SNP accounts statistically for only a tiny fraction of inherited

  predisposition, and the nature of this association (i.e., is it causal?) is often unclear; some

  SNP variations are associated with increased risk, others with decreased risk; it is hard to

  know which are the "important" SNPs to consider for a genetic condition; sometimes it is

a combination of variants that is predictive; and environmental factors are also

fundamental in the expression of genetic conditions. (ACMG, 2008; ACOG, 2008;

Ameer & Krivoy, 2009; Hudson, Javitt, Burke, & Byers, 2007; ISONG, 2009; Robson,

Storm, Weitzel, Wollins, & Offit, 2010)

When these position statements were issued, there were certainly many proponents of DTC

genetic testing and many arguments made in favor of the tool use. There was not, however, a

disciplinary debate among relevant experts about the validity of the above technical concerns,

and all of the statements (which represent a significant contingent of the scientific community)

urged caution about this use of predictive genetic testing, so this example is a case of AUIT.

C. Value-Added Modeling Used in Teacher Evaluation (2008 - 2016)

Value-added modeling (VAM) is a collection of statistical techniques that aims to link year-

to-year changes in student test scores to various factors that might influence these changes.

Typically, researchers use VAM to study *large-scale* factors; they make inferences about, for

example, the effectiveness of a new state curriculum or the introduction of a district-wide

professional development program. But in the last decade, the use of VAM to make inferences

about *individual* teachers has gained popularity. This approach to teacher evaluation has been

endorsed in federal education policy (e.g., Race to the Top; United States Department of

Education, 2009) and is often celebrated as a more efficient, objective alternative to traditional

approaches like classroom observations or review of teacher credentials (e.g., Felch, Song, &

Smith, 2010; Klein et al., 2010; or Reform Support Network, 2013). In 2009, fifteen states

required that teacher evaluations include "objective measures of student achievement" (which are

---

[13] For a few conditions (e.g., age-related macular degeneration), the association with SNP variation is well understood (Black & Clark, 2016), and accurate risk estimates are possible. This is not the case for the vast majority of conditions that DTC genetic tests purport to predict.

often—though not always—derived using VAM). At the end of 2015, after five years of RTT implementation, this number had grown to forty-three (National Council on Teacher Quality, 2015). In many states, these scores are factored into decisions about hiring, firing, retention, salary, and bonus pay; in some districts, they are made public. All of this has serious consequences for teachers, but also for students and society more broadly: VAM is being used to decide, in part, which teachers ought to teach our students. This use has sparked controversy among teachers, parents, advocacy groups, unions, philanthropists, policymakers, and others; the fact that experts have weighed in on the controversy suggests that this might be a case of AUIT.

Several professional organizations have issued position statements about the use of VAM in teacher evaluation, most from organizations of teachers or principals which I will not consider here. To be sure, teachers and principals have expertise relevant to many critical issues in the technical decision-making context surrounding the use of VAM in teacher evaluation (e.g., the day-to-day workings of classrooms and schools, pedagogical best practice, etc.). But the work at hand is to establish a case of AUIT, so I will consider only those statements from professionals with *methodological* expertise relative to VAM.

To date, two such statements have been issued, both expressing caution about the use of VAM in the evaluation of individual teachers. The American Statistical Association introduces its statement as a clarification "as to what can and cannot reasonably be expected from the use of VAMs" (American Statistical Association [ASA], 2014, p. 1), and the American Educational Research Association concludes that "there are considerable risks of misclassification and misinterpretation in the use of VAM to inform [teacher] evaluations" (American Educational Research Association [AERA], 2015, p. 4). Both statements express more detailed caution that:

- VAM scores are imprecise for many teachers. A VAM score is a single number, but, like all statistical point estimates, there is error in its estimation. Each score is associated with

a confidence interval, a *range* of possible values for the score. The concern is that this range of values is frequently wide enough to include VAM scores that would lead to different teacher ratings (e.g., "developing" instead of "effective"), which would lead to different professional consequences. In some scenarios, this concern is less problematic: scores tend to be more precise when they are computed with more years of data, for teachers with more students, and for teachers at the very top and very bottom of the score distribution. (AERA, 2015; ASA, 2014)

- VAM scores are inconsistent—from model to model—for the great majority of teachers in the center of the score distribution. A teacher's VAM score can change if the model is specified (a) with data from different classrooms, (b) with scores from different standardized tests, (c) using different student background characteristics, or (d) using more or less data (e.g., many or few years, classes, students). One concern this presents is that a district's model specifications could change, resulting in the reclassification of a teacher whose practice did not change. Another concern is that as more years of data become available, a teacher's VAM score could change, reflecting a change in available data, not a change in practice. In some scenarios, this concern is less problematic: scores tend to be more consistent as they are generated with more years of data and for teachers at the very top and very bottom of the score distribution. (AERA, 2015; ASA, 2014)

- VAM scores can be inaccurate (i.e., biased) for teachers of certain types of students. Computing VAM scores is essentially about partitioning responsibility for student test scores among the different factors that help or hinder learning (e.g., prior learning, background characteristics, the students' teacher, etc.). This partitioning is much easier in research designs where random assignment is possible, which is not the case with VAM (i.e., students are not randomly assigned to teachers.) The concern is that a VAM

score might expect "too much" of some teachers (and unfairly penalize them for teaching

hard-to-teach students) or "too little" of others. While researchers disagree about the

magnitude of bias possible in VAM scores, they do not dispute that there *is* bias. (AERA,

2015; ASA, 2014)

While there is widespread policy support of the use of VAM in the evaluation of individual

teachers, there is not a disciplinary debate among relevant experts about the validity of the above

concerns. Both statements (which represent the views of a significant contingent of the scientific

community) express caution about VAM-based teacher evaluation, so this is another case of

AUIT.

IV. Chapter Four: Rhetorical Analysis Methodology

A. Roadmap for the Reader

In Chapter 2, I motivated and defined AUIT, and in Chapter 3, I located two instances of the phenomenon: DTC genetic testing and VAM-based teacher evaluation. This chapter describes the methodology and data I used to answer the research question, *what rhetorical moves rationalize the Audacious Use of VAM?* I begin with an overview of rhetorical analysis and justify this methodological choice along two lines: (a) the epistemological nature of public understanding and (b) the long-standing and central role of rhetoric in educational and psychological measurement. Following this, I describe the remaining details of the study: the textual corpus I considered, the analytical lenses I used, and the phases of the analysis. Chapter 5 describes the results of the study.

B. Rhetorical Analysis: A Primer

Some research methods are relatively unambiguous. Researchers who identify with them generally enjoy a shared understanding of their processes and boundaries, so definitional groundwork is largely unnecessary. For example, a study that proposes to use inferential statistics to investigate the relationship between two constructs or the relative effectiveness of two treatments typically doesn't begin with a primer on inferential statistics. Other methods, however, are much roomier in definition. In particular, researchers who describe their work as rhetorical analysis may, in practice, do very different sorts of things.

A central purpose of this chapter is to describe the rhetorical analysis I carried out in this dissertation, and I'll do this primarily by making explicit my understanding of the method. I'll use a "Frequently Asked Questions" (FAQs) format to outline some important features of

48

rhetorical analysis, including its practices and tools; its typical objects of analysis; what makes it different from other, similar methods; and how this method is used to study public understanding of science.

1. FAQ 1: What is rhetorical analysis?

This method of analysis is used to examine the ways language works in a particular context, for a particular audience, at a particular time. Researchers focus on how rhetorical moves elicit certain responses from readers and what makes particular discourses compelling or effective.

2. FAQ 2: What are the practices of rhetorical analysis?

A researcher who uses this method does four things.[14] First, she **notices**. Her eye is caught by a particular rhetorical phenomenon that she believes merits study. Then, she **describes**, helping others to perceive the phenomenon. In the process of describing the phenomenon, she also **interprets** it, gives it meaning, makes sense of it; here we see the qualitative nature of the method. The fourth practice of rhetorical analysis involves judgement. As the researcher **evaluates** the phenomenon, "in some way or another, implicitly or explicitly, [she] says that the rhetoric, product, or process is well done or ill" (Brock, Scott, & Chesebro, 1990, p. 16). The first three of these steps are pillars of most (if not all) research endeavors; the fourth step sets rhetorical analysis apart in that it explicitly invites the researcher's unique perspective.

3. FAQ 3: What are the tools of rhetorical analysis?

Researchers have at their disposal an entire analytical toolbox, filled to overflowing: frameworks, typologies, categories, models… myriad ways to understand language use. In the

---

[14] These practices of rhetorical analysis are a modification of the work of Brock, Scott, and Chesebro (1990). Their presentation includes description, interpretation, and evaluation. I

introduction to her comprehensive manual, "Rhetorical Style," noted rhetorician Jeanne

Fahnestock explains that

> Rhetorical theory presents a rich array of methods for understanding actual arguments, spoken or written – from the genres of persuasion and taxonomy of issues to the special and common lines of arguments or *topoi*, and from the overall arrangement of a case and the potential parts of a full oration to its management in small-scale units or moves. (Fahnestock, 2011, p. 8)

This embarrassment of vocational riches is both a blessing and a curse: because of the

emergent nature of the method, tools must be selected *during* analysis, depending on what the

data call for. To make this selection prior to analysis would constrain (or even cripple) the

exploratory process. To apply every tool at every turn "would be unwieldy, and probably

uninteresting. Thus, the rhetorical tools that one uses to analyze any one text can differ from

those used in analyzing another text" (Leach, 2000, p. 218). The researcher is required to listen

to her data and select the most useful tools based on what she finds as she proceeds. This means,

of course, that she must be intimately familiar with the contents of her toolbox.

4. FAQ 4: What forms of data can be analyzed with this method?

In this study, I limited myself to the study of texts (including reported speech), but rhetorical

analysis can also be used to study spoken discourse. Some scholars even use the term *rhetorical*

to describe the analysis of images (Olson, Finnegan, & Hope, 2008), non-verbal communication

(Hawhee, 2009), or the arrangement of physical spaces (Dickinson, 2002).

Rhetorical analysis is not typically applied to data elicited by social science research, like the

transcript of an interview, because use of the method is traditionally predicated on a deep

appreciation for the context of communication. Interviews occur in a very specific, intentionally

contrived context, and a rhetorical analysis of language used in this context would yield a

---

believe it's worthwhile to appreciate the initial insight that inspires a researcher to undertake this

research finding like, "here's a particular way that language works *in interviews with psychological researchers*." While this information is potentially useful to psychological researchers who design, conduct, and interpret interviews, it refers to a relatively rare communicative context. Much more typical is the use of rhetorical analysis to investigate "found" or "natural" language, allowing researchers to draw conclusions about how language works *out in the world* (Leach, 2000).

For a great deal of its history, the method was applied to overtly persuasive discourse: "politicians' speeches, newspaper editorials and lawyers' orations are traditional sources for rhetorical analysis" (Leach, 2000, p. 218). But, as we've come to understand over the last century, the idea of neutral, disinterested discourse is profoundly problematic. All discourse is interested, all discourse is persuasive… in fact, an author or speaker's claim otherwise is itself a rhetorical move. Acknowledging that there is no such thing as "just the facts, ma'am," contemporary researchers also use this method to study what might once have been considered impartial or objective discourse about science.

5. FAQ 5: How does a researcher using this method approach data collection?

It depends on the nature of the research question. Many questions in social science suggest a form of statistical sampling, and procedures for this type of evidence selection are well established. For other questions (often those that arise in textual or qualitative research), statistical sampling is inappropriate, and corpus construction is a more fitting principle for data collection. As Bauer and Aarts (2000) contrast the rationales for these two selection procedures, they explain that "corpus construction typifies unknown attributes while statistical random sampling describes the distribution of already known attributes in social space" (p. 20). Research

---

work, and so I have added the prerequisite practice of "noticing."

that seeks to identify as-yet-unidentified patterns of language that may surround AUIT is clearly inquiry of the former type, not the latter.

Corpus construction is functionally equivalent to representative sampling: it "maintains the efficiency that is gained from selecting some material to characterize the whole" (Bauer & Aarts, 2000, p. 20), while avoiding the ethical and conceptual problems of unsystematic (or merely convenient) selection. A corpus may evolve during the research process, as the researcher refines her understanding of the rhetorical phenomena under study, but the fundamental rationale for data selection will not change.

6. FAQ 6: How does rhetorical analysis compare to content analysis?

Like rhetorical analysis, content analysis can be tricky to define, but generally speaking, it's "a set of techniques for systematically identifying message characteristics for the purpose of making inferences (often formal statistical inferences) about the contours of our symbolic environment" (Evans & Hornig-Priest, 1995, p. 327). Put more simply, it's a way of drawing quantitative conclusions about qualitative features of text (Gregory & Miller, 1998), and it is this quantitative element that is absent from rhetorical analysis. Both methods begin with "considerable thought … given to the 'kinds,' 'qualities,' and 'distinctions' in the text" (Bauer, 2000a, p. 132), but after this initial descriptive step, the methods part ways: content analysis counts these text features and rhetorical analysis interprets and evaluates them (Brock, Scott, & Chesebro, 1990).

7. FAQ 7: How does rhetorical analysis compare to literary criticism?

While researchers from both traditions may attend to the same features of language (e.g., patterns of word choice, functional categories, text structure, etc.), they tend to analyze different types of data and make different sorts of inferences. Literary critics analyze texts that are

noteworthy for their aesthetic value and uniqueness. After identifying "the signature language manipulations of the literary artist" (Fahnestock, 2011, p. 12), a literary critic makes inferences about the artist's perspective, experience, or ideology. Rhetorical analysts, however, tend to focus on different data: texts that have persuasive value. Identification of the "language manipulations" in the text leads to inferences, *not about the author, but about the language itself*: its persuasive contours and its potential to influence audience attitudes and actions. Rhetorical analysis is – in a sense – much more pragmatic than literary criticism; it's concerned with the mechanics of texts, how and why they work, rather than what is beautiful or unique about them.

8. FAQ 8: Okay, wait. Rhetorical analysis has no quantitative element and it overlaps a lot with literary criticism. Is this really a legit method of social science research?

Some handbooks of social science research include entire sections on rhetorical analysis (Bauer, 2000b), some bless it implicitly (Hartas, 2010; Merriam, 2009), and others ignore it altogether (Krathwohl, 2009). This points to a lack of disciplinary consensus about the status of rhetorical analysis in the social sciences. Methodologically thoughtful rhetorician Joan Leach confirms this, remarking that "there is a tradition to rhetoric that sits uneasily in the social sciences" (Leach, 2000, p. 211).

This unease is neither fatal to the present study nor should it be particularly surprising: gate-keeping (about theories, methods, literatures, ethics, etc.) is an inevitable and perpetual part of the process of defining disciplinary boundaries. Perhaps for our present purposes it is less useful to ask whether rhetorical analysis—generally speaking—constitutes social science scholarship and more useful to ask if it is possible to use this method to say something rigorous and useful about information tools, Audacious Use, and how language might rationalize the co7ption of the

former into the latter. I believe it *is* possible and that the present study demonstrates the rigorous application of the method and its usefulness.

9. FAQ 9: How do researchers who use this method think about external validity of their findings?

Surrounding rhetorical analysis is a tradition of profound conservatism and caution about the generalizability of research findings. The method isn't used to seek out timeless qualities of *discourse, in general*. Instead, it's "very much concerned with particularities – of texts, of writers, of readers, of cultures, of historical moments" and has "little inclination towards grand scientific explanations of the workings of language" (Scott, 2008, p. 300). In fact, one of the defining characteristics of the method is its reverence for the relationship between text and context, and language is never analyzed blind to the situation in which it was produced.

10. FAQ 10: Woah. Hold on. If scholars who use this method are unconcerned with extending their findings beyond the data they study, then how is their work even remotely useful?

First of all, I take issue with the phrasing "unconcerned with extending their findings." A sense of restraint about the transferability of research results does not imply a disinclination to extend those results, but rather an appreciation for how difficult it is to do well. It's difficult, primarily because of the nature of rhetorical findings: they're always about *language in context*. And ambitious generalization from highly contextualized findings is just not methodologically smart. So rhetorical analysts build theory laterally, moving from case to case (Leff, 1980; Vaughan, 2014), helping us see how cases overlap (or don't) and why. Their goal *is* to extend their findings, they *do* hope their research illuminates more than just an isolated case, but their method is analogic, not inductive.

54

Second, if you're using the word "useful," to mean "able to help us explain and predict social phenomena," then rhetorical analysis is extremely useful. Predictable patterns in audience response are an assumption of the method (Fahnestock, 2011; Scott, 2008), but – this is the important part – this response is highly dependent on the context of the rhetorical situation. Yes, people will respond to text in predictable ways. But accurate prediction doesn't come from strong, universal theories about language use in the main, but rather from knowing enough about the audience, the historical moment, the reason for the text, the writer's goal, analogous cases, and whatever else has been said about the subject at hand (Bitzer, 1968; Leff, 1980; Zarefsky, 2008).

11. FAQ 11: Could you given an example illustrating how rhetorical findings may be generalized beyond the studied data?

Sure. Scholars have identified a set of persuasive tactics called Scientific Certainty Argumentation Methods (SCAMS) that are a regular feature of certain science-society interactions. SCAMS involve accentuating the inevitable uncertainty of scientific findings, even in instances where a finding is widely accepted among experts. Rhetors who use these tactics suggest that regulation is premature and should be delayed until it is "unambiguously justified" by science (Freudenburg, Gramling, & Davidson, 2008, p. 2). Scholars recognize that, in general, SCAMS are deployed in technical decision-making contexts where particular stakeholders seek to avoid regulation of profitable but potentially risky actions.

12. FAQ 12: This dissertation uses rhetorical analysis to focus on public sense-making about science. How has the method been used in this pursuit by other researchers?

A recent, 25-year review of the literature (Condit, Lynch, & Winderman, 2012) found that studies like this tended to pursue one (or sometimes more) of four primary purposes. The four are listed and illustrated below.

- To understand science-public interactions. In a pair of studies, Jordan (2004; 2009) examines medical and popular discourses of plastic surgery. He finds that this class of surgery is often justified by invoking cultural ideals of "health" and "normalcy," instead of traditional strategies for justifying medical intervention that appeal to a scientific rationale. This justification strategy can be understood in the context of a widespread "Plastic Body" conceptualization of the permanence vs. malleability of human corporeal identity.

- To challenge scientific rhetorics that are seen to be problematic or unjustified. Koerber, Arnett, and Cumbie (2008) examine both the original text of a scientific article on labor pain medication published in the *New England Journal of Medicine* and subsequent media representation of the article. They find that public misperceptions about the drug originated in the article itself – not in the media's translation – and had their source in novel word choices and conceptual categories.

- To improve scientific rhetorics. In the first portion of her study of "Manufactured Scientific Controversies," Ceccarelli (2011) is concerned with understanding science-public interactions (the first purpose, listed above), but in the second portion, she focuses her efforts on improving scientific rhetorics. After analyzing the

ineffectiveness of common arguments from defenders of mainstream science, she

suggests alternative strategies to engage more effectively in public debate.

- <u>To build rhetorical theory (often with a particular focus on its application to public understanding of science)</u>. McClure's (2009) research refines the traditional theoretical construct of the "narrative paradigm." He begins with a critique of the original construct as overly conservative and unable to account for what scholars have actually found about the role of narrative in public sense making. A modification to the theory is proposed and its utility is illustrated in the case of public discourse on "Young Earth Creationism."

13. FAQ 13: Into which of the above categories does this dissertation fit?

Using Condit, Lynch, and Winderman's (2012) categorization, this dissertation aims to use rhetorical analysis "to understand science-public interactions." More specifically, my research is a search for particular language behaviors that appear to accompany AUIT. An example of this type of work is Ceccarelli's (2011) identification of two language behaviors – the exploitation of balancing norms and the appeal to democratic values – that seem to be present in cases of "Manufactured Scientific Controversy."

C. Methodological Justification

There are two reasons for my choice of rhetorical analysis as a research approach. The first reason is related to the nature of the theoretical problem at the heart of the dissertation: a particular version of public understanding has led to a particular kind of use of an information tool (i.e., audacious). The second reason is related to the disciplinary heritage of the substantive problem at the heart of the dissertation: the public interpretation and use of an information tool

from the field of educational and psychological measurement. Along these two lines, in the

following two sections, I elaborate and justify my methodological choice.

1. Beyond the Deficit Model of Public Understanding

Many scholars in science studies (including rhetoricians) scrutinize particular intersections of

science and society, and much of their research focuses on questioning the long-unquestioned,

oft-default political legitimacy of scientists. The goal is often to "broaden the circle" and make

room in the conversation for the perspectives of nonscientists (Farrell & Goodnight, 1981;

Harding, 1998; Keller, 1985; Pielke, 2007; Wynne, 1989). If this "broad circle" motif sounds

familiar, that's because it featured in Chapter 2. There it was expressed in terms of the

multiplicity of legitimate perspectives that potentially bear on much science-related policy.

I believe in broadening the circle. That is, I believe that science-related policy ought to

include the perspectives of nonscientists as well as scientists. But when it does, we run up

against the Problem of Extension, which is essentially questions of *how and how far to broaden*

the circle (Collins & Evans, 2002). Once we acknowledge the legitimacy of nonscientist voices,

we begin to face such questions as:

- Who has a legitimate right to weigh in on the various kinds of questions that pepper a

  technical decision-making context?

- How do we separate expertise from political rights?

- When should we listen and defer to scientists (because they have relevant expertise) and

  when should we not (because – for some questions – their credentials aren't relevant in

  the slightest)?

As a nonscientist public turns its attention to the capacity and appropriate use of an expertise

tool, the above questions arise. Various understandings and perceptions of the tool emerge.

Some of these coalesce into science-related policy. And sometimes, this policy takes the form of AUIT. Now, if I believed that the perceptual mechanics underlying (and perhaps leading to) AUIT were simply a matter of the public failing to understand the technical aspects of science, then this dissertation might have proceeded very differently from this point forward. Under this "deficit model" of public understanding, public ignorance is often taken to be the source of conflict in matters of science-related policy and often the proposed solution is straightforward: "educate the public about the technical details of the matter in dispute. Once citizens are brought up to speed on the science, they will be more likely to judge scientific issues as scientists do and controversy will go away" (Nisbet & Scheufele, 2009, p. 1767).

Research efforts pursued under this "deficit model" have traditionally been committed to quantitative documentation of public misunderstanding, often through surveys of the public or content analyses of the media (Gross, 1994), and with a frequent aim to "police corruption of scientific truth" (Evans & Hornig-Priest, 1995, p. 330). Had I kept with this tradition, I might have used my case of VAM to plumb the depths of statistical illiteracy in the general public, for example, and demonstrated that most nonstatisticians don't know what to make of confidence intervals or correlation coefficients. The implications and recommendations of my work would likely have included a call for different pedagogical strategies: ways to increase Jane Q. Public's statistical literacy.

But I don't believe the perceptual dynamics of AUIT amount to a deficit in public understanding of science. Or rather, I believe there is much, much more going on. As we saw in the cases identified in Chapter 3, AUIT arises amid a swirl of ethical, political, and social concerns. In seeking to use an information tool to make decisions or plans, nonscientists must not only devise an understanding of the tool, but also synthesize this understanding into a more

capacious epistemological map. My aim in this study was to unpack the nuances of both parts of this sense-making process.

To this end, I chose a methodological approach familiar to many rhetoricians of science (Condit, Lynch, & Winderman, 2012): situate the research at a particular interaction between science and society; study the broader ethical, political, and social context; consider who is communicating with whom, and how and why; identify common sense-making devices (like metaphors or euphemisms); and map out frequent arguments and analyze their persuasive logic. These research moves have the potential to collect a set of data rich enough to answer the overarching question of the dissertation, *what makes AUIT make sense?*

If I choose to build on this work in the future and suggest communicative strategies to address or avoid AUIT, this is not because of a secret hope that more accurate public understanding of technical details will increase the pull of a scientific perspective in public decision making. A scientific voice must carry or falter on its own merits. There may, however, be shifts in language that can mitigate the problem of Audacious Use, not in the old-school spirit of venerating science simply because it is science, but rather because the scientists who develop and use information tools have something important to say about the inferences those tools can support. Identifying particular uses of language that may obscure this input from scientists potentially brings clarity and awareness to *everyone* involved in public decision making, scientists and nonscientists alike.

2. VAM and Rhetorical Analysis: Likely Bedfellows

In the previous section, I argued that rhetorical analysis is a good methodological fit for this study because of (what I believe to be) the theoretical nature of "problems" in public understanding of science. Here, I'll explain another reason for my choice of method: the

disciplinary heritage of my substantive case: "If measurement is science and the use of measurement is applied (political) science, the justification and defense of measurement and its validity is and may always be a rhetorical act" (Messick, 1988, p. 43).

The problem of how to recognize teacher quality can be thought of as a *measurement* problem. Here, I am using the word *measurement* as it's understood by experts in the fields of educational and psychological measurement: the use of a test to generate a number (i.e., a score) that we interpret as information about the amount of a particular psychological construct that characterizes an individual (Crocker & Algina, 1986).

Educational and psychological measurement is a poster child for Post Normal Science: psychological constructs can be defined in a multitude of ways, assigning numbers to these constructs is essentially uncertain, and the interpretation of scores is value-laden. Eminent psychologist and measurement expert Samuel Messick points out that "ideological overlays are hard to avoid in educational and psychological measurement. Values pervasively influence test interpretation in subtle and not so subtle ways, especially for very general constructs like intelligence or competence" (1981, p. 13).

3. Validity

Furthermore, a great number of tests are developed *specifically* for high-stakes decision making outside the scientific community, and so this field cannot distance itself from issues that arise when its science becomes part of science-related policy. The disciplinary convention is to discuss these issues in terms of *validity*: "the overall degree of justification for test interpretation and use… [which is] concerned with social context and social purpose [and] scientifically and philosophically grounded in both evidence and ethics" (Messick, 1981, p. 18).

61

Validity is a central, long-standing concern in educational and psychological measurement. A pervasive tone of caution runs through much of the literature, urging scientists and nonscientists alike to pay careful attention to what test scores mean (and don't mean) and what to do (and not do) with them – in short, to think long and hard about the validity of a particular test for a particular use, especially when there are high stakes attached to this use. Well-known applied linguist and measurement expert Lyle Bachman adds that "[t]his is particularly important in situations where an assessment that was originally developed for one use may be considered for a different use" (Bachman & Palmer, 2010, p. 96). Here, the connection between validity and VAM becomes clear: we can think of VAM as a measurement tool, used to obtain a score, from which a particular inference is made, which leads to a particular policy action.

Given the disciplinary significance of validity, is there a disciplinary protocol for establishing it? How do experts in the measurement community suggest we approach validity? What would we need in order to sanction a particular use of a measurement tool?

4. Enter: Rhetorical Analysis

We need nothing more and nothing less than a convincing argument (Cronbach, 1980), a rationale, a logical narrative that marshals relevant evidence in a compelling way. "Validation combines scientific inquiry with rational argument to justify (or nullify) score interpretation and use" (Messick, 1994, p. 3). This argument should include not only the technical capacity of the tool and the inferences it supports, but also the particulars of the technical decision-making context where the tool will be used. It should address what's being measured, who's being measured, who's doing the measuring, what actions will be taken, and who might be affected by these actions (Bachman & Palmer, 2010; Crocker & Algina, 1986; Cronbach, 1980; Messick, 1981).

Experts in the measurement community make a clear case for the centrality of rhetoric in validation. We might even say there is a disciplinary precedent for a version of rhetorical analysis in the critique of validity arguments, though this version is somewhat staid. For example, when measurement experts analyze a validity argument, they typically don't attend to the nuances of word choice or make explicit the metaphors that shape the argument under study. They typically don't think about the emotions or imagery evoked in the text. They do, however, consider the context of the discourse, its assumptions and exigence, its audience and history. And their focus – as in all rhetorical analysis – is on *how* the language works and what makes (or fails to make) the argument persuasive. In this spirit, we see the disciplinary relevance of rhetorical analysis and the precedent for a form of this analysis (i.e., the critique of validity arguments) to support appropriate public interpretation and use of information tools.

5. Two Possible Concerns

In making my case for the use of rhetorical analysis in my study of the Audacious Use of VAM, I may have raised two questions for the reader, but they're quickly dispatched.

First, given that there are scores of published articles examining the validity of VAM in teacher evaluation, how is the work of this dissertation original? An exemplar of this sort of existing work is Haertel's 2013 study, "Reliability and validity of inferences about teachers based on student test scores." In it, he examines the logic of a generic, authorless argument for the use of VAM in teacher evaluation. His purpose is to contextualize the technical, conceptual, and social problems with this use of VAM in the broader technical decision-making context – a classic move in validation (or invalidation) efforts. The overlap between my work and Haertel's is that I too am examining arguments for the use of VAM in teacher evaluation. But there are critical differences in our data and purposes. I analyzed arguments made by nonscientists in the

public sphere, arguments that typically don't adhere to norms of scientific discourse and that marshal a diversity of types of evidence. And, unlike Haertel, my aim was not a validation study, but rather an exploration of nonscientist understanding of a particular interpretation and use of VAM.

The second question I may have raised for the reader pertains to whether the assignment of VAM scores to individual teachers is really an exercise in measurement. In the science-related policy I'm considering, the construct we would like to measure is "teacher quality," which pertains to what a teacher does in the classroom. A common interpretation of a VAM score is something like, "how much 'teacher quality' an individual teacher has." What is the "test" that generates a VAM score? Consider the way that "test" is defined by a consortia of three of the largest professional organizations in educational and psychological measurement: "a device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council of Measurement in Education [NCME], 2014).

A VAM-based approach looks for the effects of teacher behaviors. The assumption is that differences in teacher quality will manifest themselves in differences in student growth on standardized tests (i.e., that changes in student test scores are valid indicators of teacher quality). When VAM is thought of this way (as a kind of measurement problem) we can evaluate VAM-based teacher evaluation in terms of validity and ask whether a "clear articulation of each intended test score interpretation for a specified use" has been set forth and whether "appropriate validity evidence in support of each intended interpretation" has been provided" (AERA, APA, & NCME, 2014, p. 23).

<center>*   *   *</center>

The essentials of rhetorical analysis were covered in the first section of this chapter, which served to summarize the nature, processes, and challenges of this work. The remainder of the chapter describes how the method is applied in this specific study. It includes a description of the corpus, a summary of the rhetorical lenses that launched the research, and the phases of analysis.

D. Corpus

My goal in constructing my corpus was to collect evidence of rhetorical moves that may rationalize the Audacious Use of VAM. I considered texts from three domains where teacher evaluation policy is discussed for the benefit of an audience without technical expertise in VAM: popular media, advocacy, and policy messaging from the Department of Education. Throughout the study (i.e., for my initial corpus and for subsequent additions), I used three selection criteria for corpus material. First, the text had to mention VAM-based teacher evaluation in order to be relevant to the case. Second, it had to be published between 2008 and 2016 (when this research was conducted). In 2008, Michelle Rhee captured national attention with her highly-controversial implementation of VAM-based teacher evaluation in Washington DC. The following year, the policy appeared in RTT, championed and incentivized by Education Secretary Arne Duncan, and in the subsequent years, there was a sharp increase in the number of states using VAM in teacher evaluation (NCTQ, 2015). My third criterion was that the text be intended for an audience without technical expertise in VAM, for the goal of the study was to locate moves that may be persuasive to a nonexpert audience. Below, I describe my initial corpus and explain how I added to it throughout the study.

1. Popular Media

I considered *Time Magazine* and the *Washington Post* to be representative sources of widely-read, lay-audience, general-interest, popular media. I searched for both news content and opinion pieces (blog entries, op-eds, letters-to-the-editor, etc.). A search of the *Time Magazine* archives simultaneously using the terms "teacher" and "evaluation," yielded 317 articles, 62 of which discussed VAM. A search of the *Washington Post* archives using the phrases "teacher evaluation" and "value added" (quotation marks used in search) yielded 277 articles, 276 of which discussed VAM.[15]

2. Advocacy Discourse

I reviewed the websites of 64 advocacy groups and think-tanks[16] whose perspective on education policy has been described as reform-oriented (Ladner & Lips, 2012; Ravitch, 2014). In the context of education policy, a "reform-oriented perspective" pertains to a variety of different issues. What's important for my purposes is that such a perspective tends to support substantial revisions to the way teachers are prepared, evaluated, paid, and managed, and often these revisions call for VAM-based teacher evaluation. The organizations varied in the extent to which they discussed VAM-based teacher evaluation policy, from no discussion to extensive coverage (e.g., blog series, white papers, draft legislation, communication strategies). This domain contributed 192 texts to my initial corpus, approximately 3 texts per website.

---

[15] I used slightly different search strategies for the two publications when it became clear that the publications use different default Boolean search operators. For example, searching the Washington Post using the terms "teacher" and "evaluation" returned a total of 4,347 hits, all of which appear to mention "teachers" or "evaluation," but very few of which mention "teacher evaluation." The strategies I chose to use are the ones that seemed to cast the broadest relevant net for each publication.

[16] The distinction between advocacy groups and think-tanks is unimportant for this study, as both types of organizations produce public-facing discourse intended to inform and influence opinion and policy.

3. Policy Messaging

I focused my attention on discourse from Secretary Duncan to a variety of audiences with various interests in education but no technical expertise relative to VAM (e.g., teachers, union personnel, reporters, elected officials). A search of the Department of Education's website using the search phrase "teacher evaluation" yielded 45 speeches from Duncan, 22 of which discussed VAM.

4. Additional Texts

Through a process similar to snowball sampling, my corpus grew as my analysis progressed. For example, it often occurred that a newspaper article would link to a relevant article from a different news source. If the linked article contained evidence relevant to my study, I added it to my corpus (provided it aligned with my other selection rationale). These additions continued until I reached a sort of saturation point where it seemed I had seen and analyzed the significant, widely-occurring ways that VAM-based teacher evaluation is discussed. This additional corpus material comprised 164, 27, and 4 texts in the domains of popular media, advocacy discourse, and policy discourse, respectively. Table 4-1 summarizes the original and added materials in each domain.

Corpus of the dissertation: initial data set, additions, and in total				
	Popular Media	Advocacy Discourse	Policy Messaging	Total
Initial Texts	338	192	22	552
Added Texts	164	27	4	195
Total Texts	502	219	26	747
Table 4-1. This table summarizes the distribution of corpus texts over three domains.				

E. Analytical Lenses

As is often the case in emergent research, the tools I used to analyze my corpus evolved in response to what I found over the course of the study (Morgan, 2008). All of the texts, however, were initially considered in light of the following theoretical lenses. The first three are suggested by Leach (2000), and the fourth is my own, based on the particulars of my study.

1. Establish the Rhetorical Situation

As Bitzer (1968) defined it, there are three features of the rhetorical situation, listed below with guiding questions:

- Audience. For whom was this discourse intended and who is capable of taking action as a result of encountering this discourse?

- Exigence. Why does this discourse exist? What "needed to be done" and thus led to the production of this discourse?[17]

- Constraints. What persons, events, objects, and relations limit decisions and actions that might be taken as a result of encountering this discourse?

2. Identify the Types of Persuasive Discourse Using Stasis Theory

There are several articulations of Stasis Theory, but generally speaking, it's a tool for classifying the types of questions (and hence, arguments) that might arise about a subject. In my analysis, I used Fahnestock and Secor's (1988) 5-point articulation to categorize arguments about VAM as arising from questions of fact, definition, cause, value, or action.

---

[17] Some scholars (e.g., Vatz, 1973) have taken issue with this conceptualization of the rhetorical situation, arguing that it's problematic to think of exigence as existing prior to discourse and that, in fact, a speaker or writer often implicitly creates the exigence for his or her discourse.

3. Apply the Five Canons of Rhetoric

The canons function like a rubric or a conceptual scheme of the different components of a persuasive argument. In my analysis, I used the first three (the others pertain to spoken discourse) to simplify, compare and contrast, and evaluate the arguments that appear in my corpus texts.

- Invention. What means of persuasion does the author use? Does the argument appeal to authority or credibility? Emotion? Logic?

- Arrangement. How does the author structure the argument?

- Style. What stylistic devices (e.g., tropes or schemes) does the author use? How (and how effectively) does the author take into account the conventions of the context or topic?

- Memory. Is the speaker able to deliver the argument from memory?

- Delivery. Does the speaker make effective use of voice, timing, volume, gesture, visual aids, etc.?

4. Identify Typical Science-Related Policy Frames

Because my research is focused on science-related policy, an important part of my analysis was to consider common argument frames that have been identified in science-related policy debates. I used Nisbet and Scheufele's (2009) typology as a lens and determined whether their frames (listed here) or variants of them appeared in my corpus texts:

- Social progress,

- Economic development / competitiveness,

- Morality / ethics,

- Scientific / technical uncertainty,

69

- Pandora's box / Frankenstein's monster / runaway science,

- Public accountability / governance,

- Middle way / alternative path, and

- Conflict / strategy.

F. Phases of Analysis

Phase 1 of the analysis was an initial scan of all my corpus material. This generated a list of

fifteen potential rhetorical moves that served as the starting point for Phase 2, in which I

significantly refined the list by adding to it, deleting from it, collapsing multiple moves into one,

splitting single moves into multiple, etc.

In Phase 2, I worked with approximately a third of my corpus, tagging instances of each of

the fifteen moves. As I did this, I noticed that some of my moves didn't appear very often. For

example, I had expected to see a move I called, "Burying the Lede," in which a rhetor brings up a

technical concern in a footnote or mentions it briefly, at the end of an article, buried beneath

flashier, distracting ideas. I found very few instance of this, so I dropped it from my list. Some

of the moves on my list evolved as I watched them "in action" during Phase 2. For example, as I

tagged more and more instances of a move I called "Inoculation" (in which a rhetor presents a

weak version of an opposition argument and shows the reader or listener how to address it), I

realized the move was more complex than I initially thought. "Inoculation" was revised in

Phases 2 and 3 and eventually became "Manufacturing Resolution to Technical Concerns," a

move included in my final results. Phase 2 concluded with a revised list of eight rhetorical

moves that more accurately reflected what I had found in the first third of my corpus.

Phases 3 and 4 mirrored Phase 2. During Phase 3, I added another third of my corpus to the

data I was actively reviewing; I further revised my list to include eleven moves. During Phase 4,

70

I added the final third of my corpus to the data I was actively reviewing. I refined my list a final

time, identifying the most significant moves and conceptually organizing them to balance clarity

with parsimony.

V. Chapter Five: Findings

A. Roadmap for the Reader

This chapter presents the empirical findings of the dissertation: four rhetorical moves that rationalize the Audacious Use of VAM in teacher evaluation. In the sections below, I describe the moves, explain how they influence a reader or listener, and illustrate them with examples.

B. An Overview

My findings are listed in Table 5-1 below. Though these four moves vary in their logic, they all give the reader or listener a way to "get around" technical concerns, by which I mean they legitimate VAM-based teacher evaluation in spite of caution from the scientific community. In the discussion below, I'll use the term "rhetorical ingredient" to mean any element of the rhetorical situation that might be exploited in order to build an argument. Examples include a particular stakeholder or group; the historical or economic context for the discourse; a feature of the media landscape; a political, social, or ethical exigency; etc. I'll refer back to these rhetorical ingredients in Chapter 6, when I discuss the presence of these moves in other cases of VAM.

Rhetorical moves that rationalize the Audacious Use of VAM

1. Questioning Motives of Technical Concerns (QMTC) by
    • Associating them with interested nonexperts
    • Describing ulterior motives
2. Manufacturing Resolution of Technical Concerns (MRTC)
3. Eclipsing Technical Concerns (ETC) by suggesting that opponents
    • Are avoiding reality
    • Fear change
    • Are withholding information
4. Ignoring Technical Concerns (ITC)

Table 5-1. The four main rhetorical moves identified in the study are listed, including variants of the first and third move.

In the first two moves, a rhetor acknowledges technical concerns and then suggests (in various ways) that they do not have the decision-making heft to hinder policies of VAM-based teacher evaluation. In the second two moves, technical concerns are not mentioned. Instead, a rhetor sets up opposition arguments that are either morally suspect (the third move) or can be dispatched on nontechnical grounds (the final move).

1. A Disclaimer

This research sought to identify rhetorical moves that rationalize the Audacious Use of VAM in teacher evaluation. I set out only to identify common moves and explain how they work, not to assert that they are part of a calculated policy agenda to manipulate public opinion. Such a conclusion is the reader's to draw, if she believes it is warranted, but I make no such claim. Absent a "smoking gun" memo or similarly damning evidence, ascribing agency and intent is delicate, difficult work. I have not undertaken it here.

C. Questioning Motives of Technical Concerns

The first move I'll introduce, Questioning Motives of Technical Concerns (QMTC), works on the principle of distraction. As a rhetor discusses technical concerns, she introduces the idea of motive and invites the reader or listener to ask, *why is this technical concern being raised?* The parties whose motives are questioned (or impugned) are nonexperts who, having learned of the technical concerns, bring them up in debates about VAM-based teacher evaluation.

QMTC may curb the decision-making heft of technical concerns in two ways. First, it draws attention away from the substantive content of the concerns, inviting the reader or listener to focus on the political aspects of the debate. Second, by associating the concerns with nonexperts, the move strips them of an aura of objectivity they might have otherwise had. To be clear, there is no direct assertion in QMTC that the *experts* who express concern have a partisan interest in

73

the debate. That is, the move doesn't directly question the rigor or impartiality of the concerns.

But in an indirect way, through a sort of "guilt by association," the rhetor mitigates the concerns:

perhaps they can be taken with a grain of salt, considering that they align with the motives of

partisan groups.

I'll distinguish two variants of the move below. The second is a more aggressive version of

the first, but they use the same fundamental strategy to shape audience perception. The reader

should note that these variants are often deployed together.

1. QMTC – Associating Technical Concerns with Interested Nonexperts

In this gentler variant of QMTC, a rhetor associates technical concerns with a specific

nonexpert group or individual who has a presumed political or economic interest in VAM-based

teacher evaluation. The rhetor mentions the group, but leaves it up to the audience to infer their

interest in the use of the tool. This is illustrated in a *New York Times* article that discusses

teacher quality, how to measure it, and its long-term effect on students:

> Supporters [of VAM-based teacher evaluation] argue that such metrics hold teachers
> accountable and can help improve the educational outcomes of millions of children.
> Detractors, most notably a number of teachers unions, say that isolating the effect of a given
> teacher is harder than it seems, and might unfairly penalize some instructors. Critics
> particularly point to the high margin of error with many value-added ratings… (Lowry, 2012,
> para. 6-7)

The implicit message here is that teachers' unions are only raising technical concerns because

it serves their political agenda. If a reader tends to oppose the political priorities of teachers'

unions, this move may be particularly effective. But the move is potentially persuasive

regardless of the reader's personal beliefs. By linking the concern with *any* interested group of

nonexperts, the rhetor casts it in a partisan light. The examples in Table 5-2 further illustrate this

variant.

Language used to associate technical concerns with political interests	
Source	**Suggestive Language (emphasis added)**
Article in *The New York Times* titled, "Grading Teachers by the Test"	"**Critics** have questioned the Harvard scholars' findings. **Teachers** argue there is no way they could isolate the impact of teaching itself from other factors affecting children's learning, particularly such things as the family background of the students, the impact of poverty, racial segregation, even class size" (Porter, 2015, para. 16).
Editorial in *The New York Daily News* titled, "Release of Teacher Test Data in English and Math is a Big Step Forward"	"**Teachers union boss** Michael Mulgrew has demagogued the ratings effort. Nitpicking the methodology and mocking the complex formula, he says the reports should have been withheld" ("Release of Teacher Test," 2012, para. 10).
Table 5-2. Examples illustrate how rhetors may suggest that technical concerns about VAM are raised for political reasons.	

2. QMTC – Describing Ulterior Motives

The second variant of QMTC works on exactly the same principle as the first (i.e., associate the concern with partisan, nonexpert interests), so this section will consist primarily of examples. The difference between this variant and the first is that here, instead of linking the technical concern to a particular nonexpert group, the rhetor posits an ulterior motive for which the technical concern has been raised. This variant is deployed in the following excerpt from a policy paper published by the advocacy organization Bellwether Education Partners:

> Evaluations by RAND, The Gates Foundation's MET project, and other researchers show that while value-added models should be used with caution, they can help responsibly inform some personnel decisions and are not the lottery their critics make them out to be. Yet critics have seized on technical elements of value-added data as a way to undercut teacher evaluations generally. (Rotherham & Mitchel, 2014, p. 15)

The final sentence in the passage makes it clear that the reader should not take technical concerns at face value. Because they are raised to further a political agenda, they should invite skepticism, and their authority in decisions about teacher evaluation policy is dubious. The same effect is produced in the examples in Table 5-3 below.

Language used to suggest that technical concerns are raised for an ulterior motive	
Source	**Suggestive Language (emphasis added)**
Article in *The Washington Post* titled, "Contentious Teacher-Related Policies Moving from Legislatures to the Courts."	"Critics say that the unions are exaggerating both the problems associated with value-added scores and the weight that they carry in evaluations. Value-added scores account for up to 50 percent of evaluations in some states, and a smaller portion in many others, with the remainder of teachers' ratings comprised of classroom observations and other measures. 'Essentially teacher unions **don't want any evaluation**,' said Eric Hanushek, an economist at Stanford University's conservative Hoover Institute and a supporter of value-added measures. '**That's what they're angling for**'" (Brown, 2015, para. 8-9).
Article in *The Los Angeles Times* titled, "D.C. Schools May Hold Lesson for L.A."	"Some educational experts and union leaders say that value-added is not reliable enough for high-stakes decisions on firing, tenure or pay… [S]enior program officer Steve Cantrell said concerns that the method may inaccurately assess some teachers must be balanced against the likelihood that it will improve the chances for children to have an effective instructor. 'If you **shift the perspective from what is best for adults to what is best for students**, then it's super clear that value-added can improve the system over time,' he said" (Watanabe, 2010, para. 13-15).
Table 5-3. Examples illustrate how rhetors may suggest that those who raise technical concerns about VAM have an ulterior motive.	

In summary, while both variants of QMTC may steer the reader towards questions of motive, they differ in how explicitly they do this. In both cases, the rhetorical ingredient needed to deploy the move is the same: a nonexpert group with known political interests that raises technical concerns about VAM.

D. *Manufacturing Resolution of Technical Concerns (MRTC)*

The second move I'll present, Manufacturing Resolution of Technical Concerns (MRTC), gets around technical concerns by suggesting that they can be or have been resolved and thus need not hinder adoption of VAM-based teacher evaluation. This move draws its persuasive

potential from a mismatch between a technical concern and what I will refer to as a "misaligned finding": a scientific finding that is spuriously presented as a mitigating counterpoint to the technical concern.

A key difference between this move and the last is the extent to which their deployment involves the details of technical concerns. In this respect, QMTC is a blunt rhetorical tactic: a rhetor can suggest that concerns about the precision of VAM estimates are raised for ulterior motives without knowing what a lack of precision means or when, why, and for whom it might be a problem. Deploying MRTC, on the other hand, is more delicate work: it requires that a rhetor understand the details of technical concerns well enough to muddy them in a way that suggests they've been resolved.

The essential rhetorical ingredient in MRTC is the finding that a rhetor misaligns. This finding must be similar enough (e.g., substantively, methodologically, etc.) to the technical concern with which it's juxtaposed that a reader or listener doesn't perceive the mismatch. The canonical form of the move is as follows. A rhetor opens the move with a technical concern. This is followed by an adversative (e.g., "but," "however," "others say"). She closes the move with the misaligned finding. This move distorts the truth, but not by fabricating research findings. Everything the rhetor reports is accepted in the scientific community. Rather, the crux of the truth distortion is that the misaligned scientific finding is presented as if it resolves the technical concern when, in fact, it does not.

Consider the use of MRTC in the same *New York Times* article from which I excerpted the in-text example of QMTC in the previous section. The author attenuates the technical concern that VAM scores are inconsistent for the majority of teachers in the middle of the distribution. She opens the move with the technical concern, which she links (using the adversative, "but") to two misaligned findings about special cases where the concern does not apply: "[VAM scores]

77

tend to bounce around for a given teacher from year to year and class to class. But looking at an individual's value-added score for three or four classes, the researchers found that some consistently outperformed their peers" (Lowry, 2012, para. 7).

In this example, the findings that follow the adversative do not align with the concern, in the sense that they apply only to teachers for whom a lot of data are available or teachers at the top of the score distribution. Now, a perceptive reader or one with some background knowledge about VAM may notice this misalignment, in which case the move will lose its persuasive power. But for many readers, the take-away message may be defined less by the details of the passage, and more by its overall cadence: it moves from doubt to confidence, and this order is key. The author begins with a "yellow light" (i.e., the technical caution) and ends with what many lay readers will interpret as a "green light" (i.e., the special-case findings). Consider how differently the passage would read with a green-to-yellow construction: "Looking at an individual's value-added score for three or four classes, the researchers found that some consistently outperformed their peers. But VAM scores tend to bounce around for a given teacher from year to year and class to class." Had the author ordered the passage like this, the take-away message would be one of caution instead of resolved caution.

1. Terrifically Misaligned Findings

Even when rhetors deploy MRTC with terrifically misaligned findings, this yellow-to-green construction may be effective. Consider this excerpt from a *New York Times* article about the 2012 teachers' strike in Chicago:

> Several studies have shown that teachers who receive high value-added scores—the term for the effect that teachers have on student test performance—in one year can score poorly a year later. "There are big swings from year to year," said Jesse Rothstein, associate professor of public policy and economics at the University of California, Berkeley. But other studies have shown that students taught by teachers who achieve high value-added scores go on to have

78

lower teenage pregnancy rates, are more likely to go to college and earn higher incomes as adults. (Rich, 2012, para. 16)

To be sure, an astute reader may note that evidence of long-term teacher effects does not resolve a concern about unstable VAM scores. But to a casual or hurried reader, the adversative ("but other studies have shown") and the yellow-to-green construction may be sufficient to manufacture resolution of Rothstein's concern.

2. "Expert-Augmented" MRTC

I found a few instances of MRTC in which rhetors augmented the canonical form of the move by referring to specific experts: one linked to the technical concern and another linked to the misaligned finding. Consider the following example of this "expert-augmented" MRTC, taken from a *New York Times* article discussing controversy over VAM-based teacher evaluation:

> "If these teachers were measured in a different year, or a different model were used, the rankings might bounce around quite a bit," said Edward Haertel, a Stanford professor who was a coauthor of the report. "People are going to treat these scores as if they were reflections on the effectiveness of the teachers without any appreciation of how unstable they are." Other experts disagree. William L. Sanders, a senior research manager for a North Carolina company, SAS, that does value-added estimates for districts in North Carolina, Tennessee and other states, said that "if you use rigorous, robust methods and surround them with safeguards, you can reliably distinguish highly effective teachers from average teachers and from ineffective teachers." (Dillon, 2010, para. 11-13)

The misalignment here is similar to what we saw in the first example of this section. Haertel's concern is with the stability of scores for individual teachers; Sanders's "resolution" deals only with teachers at the top ("highly effective") and bottom ("ineffective") of the distribution. We're able to compare these "extreme" teachers (either to teachers at the other extreme or teachers in the middle) with more confidence, but comparisons between and among most teachers (i.e., the non-"extreme" ones) remain deeply problematic. Three days prior to being quoted in this article, Sanders himself emphasized the danger of this conceptual misalignment on National Public Radio:

He [Sanders] says that value-added analysis can accurately single out both star performers and ineffective teachers. But, Sanders cautions, "can you distinguish within the middle? No you can't, not even with the most distinguished, value-added process that you can bring to the problem." And Sanders worries that parents may come to the wrong conclusions about those middle-performers. (Abramson, 2010, para. 10-11)

The reader (of this chapter) may be wondering if "expert-augmented" MRTC might be better described as "Manufacturing Scientific Controversy" (Ceccarelli, 2011). In Chapter 6, I discuss why I believe this isn't the best description. For now, I'll simply point out that the primary purpose of the move above seems to be the specious resolution of technical concerns, rather than the inflation of expert disagreement.

E. Eclipsing Technical Concerns (ETC)

The third rhetorical move eclipses technical concerns by framing the debate as a moral one and locating support of VAM-based teacher evaluation on the high ground. Rather than explicitly address technical concerns, a rhetor implies that opposition arguments (be they technical or otherwise) arise from unprincipled motives, inclining a reader or listener away from these arguments and towards support of the policy. Three variants of this move were identified; each corresponds to a different moral lapse that rhetors suggest leads to opposition arguments.

1. ETC – Those Who Oppose VAM-Based Teacher Evaluation are Avoiding Reality

The first variant of ETC suggests that opposition arises from a desire to avoid an unpleasant but critical reality. Secretary Duncan relied heavily on this move to champion his Race to the Top (RTT) program, particularly the parts that pertained to teacher evaluation. In a speech to education journalists shortly after RTT was introduced, he asserted that policies like VAM-based teacher evaluation "expos[e] the good, the bad and the ugly around issues like teacher effectiveness… This is not always fun. No one wants to admit their flaws—let alone do something about them" (U.S. Department of Education, 2009, April 30, para. 29-30). On his

annual "Back to School" tour the next year—as he discusses the publication of VAM scores and their use in high-stakes teacher evaluation—we see the same moral framing: "The truth is always hard to swallow but it can only make us better, stronger and smarter. That's what accountability is all about—facing the truth and taking responsibility and then taking action" (U.S. Department of Education, 2010, August 25, para. 80).

By constructing the debate in this way, Secretary Duncan signals to the listener that proponents of VAM-based teacher evaluation are champions of courage and integrity; by implication, opponents are skirting both the truth and their responsibility for it. The examples in Table 5-4 further illustrate this tactic.

Language used to suggest that opponents are avoiding reality	
Source	**Suggestive Language (emphasis added)**
Article in *Time Magazine* titled, "Blame Game: Let's Talk Honestly About Bad Teachers"	"And the politics won't change until we can at least talk about teacher effectiveness and the broader problems of educational management in **an intellectually honest way**" (Rotherham, 2011, para. 12).
Open letter about education reform, published by *The Washington Post* and penned by several leaders of large school districts	"[A] poorly performing teacher can hold back hundreds, maybe thousands, of students over the course of a career. Each day that we **ignore this reality** is precious time lost for children…" (Klein et al., 2010, para. 8). "Of course, we must also do a better job of providing meaningful training for teachers who seek to improve, but **let's stop pretending** that everyone who goes into the classroom has the ability and temperament to lift our children to excellence" (Klein et al., 2010, para. 12).
One in a series of articles that accompanied the release of VAM scores by the *Los Angeles Times*	"But across the country, parents have no access to objective information about teacher effectiveness, and many districts have opted to **ignore the data"** (Felch & Song, 2010, para. 14).
Speech by Secretary Duncan to Arkansas teachers, administrators, and policymakers to promote reform policies of Race to the Top (RTT)	"We cannot **shrink from the truth**" (U.S. Department of Education, 2010, August 25, para. 108).
Speech by Secretary Duncan to state governors, shortly after introduction of RTT	"Some states actually have laws creating a firewall between teacher evaluation and student achievement. This isn't fair to kids or to teachers. Worse yet, **it's not honest**" (U.S. Department of Education, 2009, June 14, para. 31).
Table 5-4. Examples illustrate how rhetors may suggest that opposition to VAM-based teacher evaluation is motivated by a desire to avoid an unpleasant reality.	

2. ETC – Those Who Oppose VAM-Based Teacher Evaluation Fear Change

The second variant of ETC suggests that fear of change underlies opposition to VAM-based

teacher evaluation. Rhetors who deploy this strategy frequently depict VAM as modern or

sophisticated, while depicting other methods of measuring teacher quality as simple or obsolete.

For example, in a *New York Times* article, "traditional criteria like evaluations from principles"

are contrasted with the "nationwide experiment" (Porter, 2015, para. 11) and "sophisticated

research" (para. 13) that constitute VAM-based teacher evaluation. In the article's penultimate

paragraph, we see the suggestion that a fear of technological change motivates opponents of the

policy:

> Brad Jupp, a special adviser to Secretary Arne Duncan, compares the anxiety about the
> adoption of new evaluation tools to the uncertainty in the 1940s over what would happen if
> the sound barrier was broken. Some people thought it would destroy the plane. Others
> thought the plane would accelerate to a million miles per hour. When Chuck Yeager finally
> broke it in 1947, neither happened. (Porter, 2015, para. 29)

Rhetors may also emphasize the shift in culture or thinking that accompanies VAM-based

teacher evaluation, and then suggest that opponents are uncomfortable with this philosophical

change. We see this strategy in a report published by the advocacy organization Students First,

titled, "Elevating the Teaching Profession: Increasing Teacher Quality": "…state and school

leaders must work concurrently to change the culture around evaluations and how they are used.

We must as a nation become comfortable with differentiating performance among educators…"

(Students First, 2013, p. 6).

This variant of ETC presents readers with a choice: brave support of VAM-based teacher

evaluation on the one hand and fearful inflexibility on the other. The tactic is further illustrated

in Table 5-5.

Language used to suggest that opponents fear change	
Source	**Suggestive Language (emphasis added)**
Speech by Secretary Duncan to the National Education Association, shortly after introduction of Race to the Top (RTT)	"Now, let's talk about data. I understand that word can make people **nervous**…" (U.S. Department of Education, 2009, July 2, para. 61).
Guidebook of rhetorical strategies (sponsored by the Department of Education) for communicating about VAM-based teacher evaluation and other RTT-mandated reforms to the profession	"The concept is a new one, and represents a shift from **past** systems. Teacher evaluations **traditionally** have not been tied to student performance, and many educators and researchers are understandably **anxious** about this **unfamiliar and unexplored** territory" (Reform Support Network, 2013, p. 11).
Speech by Secretary Duncan to education leaders in Los Angeles; the excerpted text precedes an entreaty for VAM-based teacher evaluation	"The time is now to get off the **well-trod road** in favor of the **road less traveled**. The time is now to confront challenges—and leave the **comfort zone** behind" (U.S. Department of Education, 2011, March 22, para. 20).
Article in *The New York Times* titled, "Principals Protest Role of Testing in Evaluations"	"Asked if he was surprised by the number of principals who had signed [a statement of opposition to VAM-based teacher evaluation], he wrote, 'It's not at all surprising' that the introduction of a new evaluation system 'would produce **anxiety**'" (Winerip 2011, para. 13).
Table 5-5. Examples illustrate how rhetors may suggest that opposition to VAM-based teacher evaluation arises from fear of technological or cultural change.	

3. ETC – Those Who Oppose VAM-Based Teacher Evaluation are Withholding Information

The third variant of ETC is used in the debate over whether the VAM scores of individual teachers should be available to parents or the public. Technical concerns bear directly on this debate, for publishing imprecise, inconsistent, or inaccurate information about an individual is understood by some to be a form of libel. Rhetors eclipse these concerns by constructing the debate around topoi like right-to-information (i.e., citizens should have access to information held by public bodies) and political-censorship (i.e., political power can be maintained by withholding information). This move is often deployed by news outlets defending their choice to

publish VAM scores. For example, after expressing sympathy to the family of a teacher who

committed suicide following the publication of his VAM score, the *Los Angeles Times* continued

their condolence statement with:

> The Times published the database, which is based on seven years of state test scores in the
> L.A.U.S.D. schools, because it bears directly on the performance of public employees who
> provide an important service, and in the belief that parents and the public have a right to
> judge the data for themselves. (Lovett, 2010, para. 11)

When the debate is framed in this way, the implicit message is that those who oppose making

VAM scores public are trying to withhold or hide information that ought to be publicly available.

This move likely resonates with a reader who values principles of transparent governance,

inclining her to support the release of scores. The examples in Table 5-6 further illustrate this

tactic.

Language used to suggest that opponents are withholding information	
Source	**Suggestive Language (emphasis added)**
One in a series of articles that accompanied the release of VAM scores by the *Los Angeles Times*	"For now, **parents remain mostly in the dark**… Which instructor a child gets is usually **decided behind closed doors** by principals and teachers, whose criteria vary widely" (Felch, Song, & Smith, 2010, para. 103-105).
One in a series of articles that accompanied the release of VAM scores by National Public Radio and the *Cleveland Plain Dealer*	"One [reason to publish VAM scores] is that state lawmakers created the value-added system to come up with a better way to assess teachers, to **give the residents of the state better accountability**… Another is that tax dollars are used to compile the ratings, meaning **the people of Ohio have paid for this**" (IdeaStream, 2013, para. 14).
Article in *The Washington Post* in which a parent discusses teachers whose VAM scores are at the bottom of the distribution	"Any kid who is stuck with one of those teachers is going to have a problem… Why would you want to **hide that information** from parents?" (Brown & Balingit, 2015, para. 13)
Article in *The New York Times* in which the chancellor of New York City schools defends his release of VAM scores	"For too long parents have been **left out of the equation**, left to pray each year that the teacher greeting their children on the first day of school is truly great, but **with no real knowledge** of whether that is the case, and with no recourse if it's not" (Otterman, 2010, para. 10).
Table 5-6. Examples illustrate how rhetors may suggest that opposition to the publication of VAM scores is motivated by a desire to withhold information from parents, state taxpayers, or the public.	

Most of the time, technical concerns are absent from arguments that use ETC. But I noted a few instances where the move was deployed after acknowledging technical concerns. For example, editors from several Ohio news outlets who collaborated to release individuals' VAM scores "decided it was more important to provide information—even if flawed—to help parents understand their children's education" (IdeaStream, 2013, para. 12). One of these editors offers further justification, in which we see the topoi mentioned above: "If public information exists about the quality of a teacher, who are we to deny that information to the parent?" (IdeaStream, 2013, para. 16).

In summary, the required rhetorical ingredient in ETC is a moral dimension in the controversy about tool use, and a way to situate opposition arguments on the moral low ground. The persuasive potential of any given moral dimension will depend, of course, on the political, historical, and social context of tool use, and a variant of ETC that is effective in one context may not be in another. What's important is that the moral dimension resonates strongly enough with a reader or listener that her inclination to consider technical aspects of the tool fades in comparison.

F. Ignoring Technical Concerns (ITC)

The final rhetorical move identified, Ignoring Technical Concerns (ITC), is the most straightforward. The required rhetorical ingredient is a collection of nontechnical arguments against tool use. The rhetor presents only these opposition arguments in her discussion of VAM-based teacher evaluation. This allows her to appear balanced (because she has mentioned both pros and cons), while keeping expert caution out of the discussion. As we saw in the introduction to the study, this move is possible because several points of controversy that arise around teacher quality have both a technical and a nontechnical dimension. Rhetors simply confine their presentation to the latter, setting up and responding to value- or belief-oriented concerns about VAM.

ITC is similar to the use of a "red herring" to change the course of an argument or a "straw man" to set up a weak opposition argument. In all cases, a rhetor positions the conversation on grounds where it is easier for her to deliver a strong argument, while avoiding issues where her position is weaker.

This move is illustrated in a *Washington Post* article about states' use of student test scores. The discussion of VAM comprises five paragraphs. The first three introduce VAM and relate an

anecdote about a teacher who raised the test scores of her struggling students. The fourth and

fifth paragraphs set up and resolve two opposition positions:

> Some teachers worry that an emphasis on data ignores other progress that can't be measured on a test, such as emotional and social development. Others are concerned that the data could be used against them. "That's a rational reaction when you think about how data has been used in the past," Guidera said. "We have to transform the way we think about data from a hammer that's going to hurt teachers to a flashlight that's going to help them." (Layton, 2012, para. 14-15)

Notice that both objections are belief-oriented: about what teacher are supposed to be doing

and about how personnel decisions should be made, respectively. No other concerns about

VAM-based teacher evaluation are mentioned in the article. The belief-based objections are

attributed to a faulty metaphor (i.e., data as a hammer) and resolved with an alternative one (i.e.,

data as a flashlight).

The first and most obvious way ITC rationalizes Audacious Use is by keeping readers

unaware of technical concerns about VAM. Even if a reader is already aware of these concerns,

however, the move can still be effective. By leaving expert caution out of the discussion, the

author implies that it is not a key factor in decision making about VAM-based teacher evaluation.

In this way, the author gives the reader license to opine about the policy without consulting the

scientific community. The examples in Table 5-7 further illustrate this variant.

Focusing on nontechnical concerns while ignoring technical ones	
Source	**Nontechnical Concerns (emphasis added)**
Article in *Time Magazine* about Chancellor Michelle Rhee's reforms to teacher evaluation in Washington DC	"'People say, "Well, you know, **test scores don't take into account creativity and the love of learning**,"' [Michelle Rhee] says with a drippy, grating voice, lowering her eyelids halfway. Then she snaps back to herself. 'I'm like, "You know what? I don't give a crap." Don't get me wrong. Creativity is good and whatever. But if the children don't know how to read, I don't care how creative you are. You're not doing your job'" (Ripley, 2008, para. 21).
Speech by Secretary Duncan at the Harvard Graduate School of Education	• "It is absolutely true that **many of today's tests are flawed**. They don't measure critical thinking across a range of content areas. They are not always aligned to college and career-ready standards. They don't always accurately measure individual student growth" (U.S. Department of Education, 2012, February 7, para. 68). • "And [standardized tests] certainly **don't measure qualities of great teaching that we know make a difference**—things like classroom management, teamwork, collaboration, and individualized instruction. They don't measure the invaluable ability to inspire a love of learning" (para. 69). • Now, some folks will point out, correctly, that **most teachers don't teach in tested subjects**. So, how can student achievement be factored in to teacher evaluation in nontested subjects? It's a great question. But I have every faith that teachers themselves can come up with solutions (para. 84).

Table 5-7. Technical concerns do not appear in the articles from which these quotations are excerpted. Instead, rhetors focus on nontechnical concerns about VAM-based teacher evaluation, like those exemplified here.

Although simple, this move is important because it appears to have been endorsed by the Department of Education. In the early 2010's, many states were rolling out new systems of teacher evaluation in exchange for RTT funds. A Department-sponsored guidance document, the "Educator Evaluation Communications Toolkit" was published in 2013. It provided state, district, and school leaders with rhetorical strategies to "abate criticism, the spread of

misinformation and unnecessary concerns" (Reform Support Network, p. 3) among teachers

whose professional evaluations had just been overhauled.

The document recommends ways to talk about teacher evaluation *in general*, but one chapter

is devoted specifically to VAM-based teacher evaluation. It includes seven "lessons for

communicating about value-added data and other measures of student learning" (p. 12). These

are listed in Table 5-8 below.

Suggestions (from the Department of Education) for talking about VAM
1. Hold realistic expectations.   2. Emphasize function within the larger evaluation system.   3. Acknowledge shortcomings.   4. Anticipate and be prepared to respond to misinformation.   5. Stay out of the weeds.   6. Plan for glitches.   7. Move quickly to implementation.
Table 5-8. Seven recommendations are listed; they are intended to help rhetors frame conversations with teachers about VAM-based teacher evaluation. (Reform Support Network, 2013, p. 12)

In the explication of the fifth lesson, "Stay out of the weeds," rhetors are instructed to ignore

two of the three major technical concerns about VAM:

> Offer a detailed explanation to those who are interested, but stay focused on the big picture
> for everyone else. If you find yourself having a conversation about year-to-year instability or
> margins of error with a roomful of teachers, chances are most of them will not find the
> information either useful or helpful to their practice. (p. 12)

Recall that to deploy ITC, a rhetor also needs a collection of *nontechnical* concerns about tool

use that she can raise and address (in lieu of discussing technical concerns). The guidance

document provides these, enumerating ten "fears and anxieties that many teachers have about the

current trajectory of education reform and the future of their profession" (p. 11), none of which is

technical. The list includes, for example, concerns about "an unhealthy focus on standardized

testing," and "the seeming 'mechanization' of teaching" (p. 12).

<center>*   *   *</center>

Each of these moves provides a "way around" technical concerns about VAM, but these paths differ in interesting ways. Some moves explicitly address the concerns, others don't. ETC elicits an emotional or aesthetic audience response, whereas MRTC takes a logical (if misleading) approach. The first variant of QMTC leaves the audience to "connect the dots," while the rhetor does this work explicitly in the second variant. In the next chapter, I'll consider which of these varied moves we should expect to see in other cases of AUIT, and why.
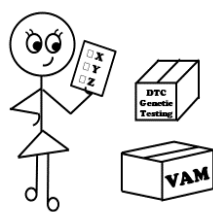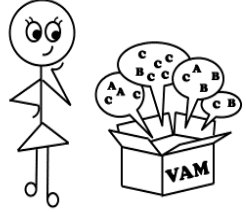
VI. Chapter Six: Discussion

A. Roadmap for the Reader

I begin this chapter with a reminder of the overall logic of the dissertation, a summary of its primary findings, and a discussion of the challenges and surprises I encountered during the research. Next, I discuss whether and why my rhetorical findings in the case of VAM might extend to other cases of AUIT. After pointing to some strategies nonscientists and scientists might use to combat Audacious Use, I comment briefly on the notion of Manufactured Scientific Controversy in the case of VAM. I then point out five questions the study raises and suggest how they might be addressed in future research. I conclude by considering the importance of the study, both inside and outside the scholarly literature.

B. Recapping and Reflecting on the Preceding Five Chapters

This research began by noticing a particular phenomenon of public decision making. As I peered at it more and more closely and from different angles, I became increasingly convinced that it warranted further study… and ultimately an entire dissertation. At its most basic, the logic of this research was first to build theory about the phenomenon, and then to verify and expand this theory. To this end, the study was organized into a series of four research questions. The reader is reminded of these questions and the work involved in answering them in Figure 6-1 below.

Developing theory about AUIT by studying the case of VAM			
RQ 1: What is AUIT?	RQ 2: What are examples of AUIT?	RQ 3: What rhetorical moves rationalize the Audacious Use of VAM?	RQ 4: Do these moves generalize to other cases of AUIT?

Figure 6-1. Each research question of the dissertation is accompanied by an illustration of the work involved in answering it.

In Chapter 2, I answered the first research question in the figure above ("What is AUIT?") with the following two definitions:

Information Tool: NOUN

an analytical method or device developed and used by scientists, that outputs information from which an inference is made about the state of the natural or social world

Audacious Use of an Information Tool: NOUN

use of an information tool to make high-stakes decisions or plans against the advice of relevant experts, among whom there is agreement concerning the technical inadequacies of the tool for the stated use

The most difficult part of this definitional work was capturing what is important and potentially dangerous about AUIT, without resurrecting the long-dead, simplistic notion that a "science badge" should earn the wearer automatic deference in public decision making. What

my survey of science-society interactions revealed is that a great deal of the controversy that swirls in cases of Audacious Use hinges on nontechnical issues of values or beliefs. In domains like these, I believe strongly that a "science badge" should neither privilege nor disadvantage a particular perspective. But I also believe that scientists have unique and relevant insights about the methodological capacity of information tools… and these insights must not be lost in the complete relativism of, for example, the Strong Programme. My challenge was to craft a definition of AUIT that granted scientists no special authority in nontechnical matters, but insisted on their authority when it came to the technical adequacy of the tool.

I spent Chapter 3 answering the second research question from Figure 6-1, "What are examples of AUIT?" I surveyed a variety of cases of science-society interaction that involved information tools and identified two instances of AUIT: the use of predictive genetic testing to make personal health-care decisions and the use of VAM in high-stakes teacher evaluation. What I came to appreciate while doing this work was the variation that can exist across different cases of AUIT. For example, those who choose to use the tool may or may not be the ones directly affected by the decisions or plans that follow. In the case of VAM, the tool is typically used by administrative officials (e.g., district or school leaders) and the results are experienced by teachers and students. The situation is different in the case of genetic testing, where the results affect the tool user herself (and perhaps her family). Another source of variation in cases of AUIT is the position of the federal government towards tool use: support in the case of VAM, and opposition in the case of genetic testing. These observations about variants of AUIT simply confirm that there's a great deal of interesting work to be done in the interdisciplinary workspace around the phenomenon.

The next logical step of the dissertation was to use rhetorical analysis to address the third research question from Figure 6-1, "What rhetorical moves rationalize the Audacious Use of

VAM?" I began this work by assembling a corpus of 552 texts from the domains of media, advocacy, and policy discourse around VAM-based teacher evaluation. This corpus increased in size by about a third over the course of the study, through a process similar to snowball sampling. From an initial review of this corpus, I identified a list of fifteen potential rhetorical moves, which I refined over four phases of analysis. Of all the moves I identified, the four I chose to present in Chapter 5 seemed the most common, interesting, and potentially persuasive in the rationalization of AUIT:

- Questioning Motives of Technical Concerns,

- Manufacturing Resolution of Technical Concerns,

- Eclipsing Technical Concerns, and

- Ignoring Technical Concerns.

This work took longer than I expected because I found that very little has been done to map the current rhetorical landscape of teacher evaluation and teacher quality. It was difficult, therefore, to distinguish between topoi that operate in this larger, contextualizing sphere of discourse, and topoi that are *particular* to the Audacious Use of VAM. For example, marketplace narratives and factory / industrial metaphors seem to orient many discussions of teacher evaluation and teacher quality *in general*, not just those discussions that involve VAM. When I realized this, I was able to set aside these "contextualizing" topoi, and focus on identifying arguments particular to VAM.

At this point in the dissertation, I have answered my first three research questions. Now all that remains is to extend the results of Chapter 5—which pertain to VAM, in particular—to AUIT, in general. This final task addresses the fourth research question from the figure above ("Do these moves generalize to other cases of AUIT?") and is dispatched in the next section.

C. Generalizing from VAM to AUIT

What I will do in this section is logical, conceptual work, not empirical work. I consider each of the four moves presented in Chapter 5 and explain if and why we should expect to see it deployed in other cases of Audacious Use. Throughout this discussion, I will point to examples from the case of DTC genetic testing. The reader should remember that I have not conducted a comprehensive rhetorical analysis of this case. The examples I cite were handpicked to illustrate that some of the moves observed in the case of VAM also operate in the case of DTC genetic tests. More work is needed to establish the ubiquity of these moves in the latter case.

1. Questioning Motives of Technical Concerns (QMTC)

As noted in Chapter 5, the rhetorical ingredient required in QMTC is a nonexpert group that raises technical concerns in arguments against tool use and that has a broadly recognized political reason for doing so. Recall that in the case of VAM, this was most often teachers or teachers' unions. Because AUIT involves, by definition, high-stakes decisions or plans, it is hard to imagine a case entirely divorced from political interests. That is, both supporters and opponents of tool use are likely to have identifiable political reasons for their support or opposition. However, in different cases of AUIT, there may not be a particularly visible *nonexpert* group that opposes tool use, so QMTC—as I described the move in Chapter 5—may not be deployed.

What I observed in the case of DTC genetic testing was a modified version of QMTC. Rhetors still limit the decision-making heft of technical concerns by suggesting they are raised to further a political agenda. But in this modified version, instead of drawing attention to the politics of *nonexperts* (who learn of and raise the technical concerns), rhetors emphasize the politics of *experts*. The likely effect on the reader or listener also shifts in this modified use of QMTC. Now, not only is her attention drawn to the political dimensions of the debate (as in the

96

original version of QMTC), she's been invited directly to question the objectivity of the science: might the integrity of the research that prompted expert caution have been compromised by the experts' vested interests?

The following excerpt, taken from an opinion piece in *The New York Times*, illustrates this modified version of QMTC. The "issue" the author refers to is whether companies should be able to sell DTC tests, in light of technical concerns and possible misinterpretation (by nonexperts) of risk estimates.

> Why do experts differ from consumers on this issue? You could argue that the experts are better informed, but you could also argue that some of them are swayed by their own self-interest. Traditionally, people have had to go through a doctor to get a test, which could mean paying a fee to the physician as well as to a licensed genetic counselor. Buying tests directly from a company like Navigenics or 23andMe can cut out hundreds of dollars in fees to the middlemen. (Tierney, 2011, para 14)

I conjecture that we are likely to observe a "QMTC-esque" move in other cases of AUIT, by which I mean that when technical concerns are raised (either by experts or nonexperts), rhetors will limit their decision-making heft by drawing attention to a political agenda. What may vary, from case to case, is whether rhetors impugn the motives of experts or of nonexperts who learn of and raise the concerns.

2. Manufacturing Resolution to Technical Concerns (MRTC)

What's required to deploy MRTC is what I referred to in Chapter 5 as a *misaligned finding*: a research result that a rhetor can juxtapose with a technical concern in order to suggest that the concern can be or has been resolved. A misaligned finding needs to be similar enough (e.g., substantively, methodologically, etc.) to the technical concern with which it's juxtaposed that a reader or listener doesn't perceive the mismatch. It seems reasonable to expect this sort of "juxtaposable finding" in any sufficiently rich research context, so I conjecture that the necessary rhetorical ingredient for MRTC is likely to be available in other cases of AUIT.

However, it's not clear to me whether we should expect to see the move deployed in other cases. At its most basic, the message of MRTC is something like, "even though technical concerns are raised, we don't really have to worry about them because there are ways to resolve them." In the case of VAM, rhetors saw persuasive potential in this story, but this may not happen in other cases of AUIT. I didn't observe the move deployed in the case of genetic testing, but this may be because I my understanding of the technical subtleties of this science is limited and I wasn't able to perceive the misalignment.

3. Eclipsing Technical Concerns (ETC)

To deploy ETC, the required rhetorical ingredient is a moral dimension in the controversy about tool use, one that resonates strongly enough with the reader or listener that technical concerns fade in comparison. The specifics of this rhetorical ingredient (i.e., what I referred to as "the variant" of ETC in the previous chapter) will depend on the social, political, or ethical complexities of the case. A variant of ETC that is compelling in one case may not work in another.

For example, one of the variants of ETC deployed in the case of VAM involves the suggestion that opponents of VAM-based teacher evaluation want to avoid reality (in this case, the reality of bad teachers). This pushes moral buttons because of the nature of our education system: teachers are public employees, and their professional competence shapes our children's futures. Trying to hide from responsibilities is, in this context, an ethical affront (compensation for a demonstrably poor performance) and a moral one (jeopardizing the potential success of a future generation). But the idea of avoiding the reality that the tool reveals doesn't have the same persuasive power in the inventional landscape of DTC genetic testing. For sure, individuals may choose not to test themselves because they don't want to face the results. But because this is a

personal decision, it doesn't elicit zealous accusations of "hiding from the truth." I didn't see

this variant of ETC deployed in our second case of AUIT.

On the other hand, the idea of "withholding information" *does* appear to push moral buttons

in the case of DTC genetic testing, as the following examples illustrate. Consider this excerpt

from the closing paragraph of a research brief whose authors work for a company that sells DTC

test kits:

> We believe that DTC genetic tests play a key translational role for the science of genetics, democratizing and disseminating privileged knowledge to the public. No matter how clichéd it sounds, knowledge is power. While some medical experts may complain about patients armed with results from DTC genetic tests or information about disease symptoms from the internet, we believe that a knowledgeable public is an empowered public. (Helgason & Stefánsson, 2010, p. 67)

The phrases "democratizing and disseminating privileged knowledge" and "an empowered

public," appeal to the same topoi we saw in the case of VAM: right-to-information and political-

censorship.

The third variant of ETC observed in the case of VAM, "fear of change," also appears to

have persuasive heft in the case of genetic testing. The co-founder of a genetic testing company

was asked in an interview with *Time Magazine* to describe the biggest misconception about her

industry. Her response follows the same construction that we saw in the case of VAM. She

draws an analogy between the information tool and a past technological breakthrough that

initially elicited anxiety but ultimately proved innocuous.

> I think one thing is that people are worried this information could be harmful, and we're really not convinced of that. Like with all new things that come out, we don't always know exactly what the implications are going to be. When the car first came out, for example, it didn't have air bags or seat belts or headrests and all the protections that were built into cars ultimately to make them safer to drive. (Lynch, 2008, para. 3)

So, should we expect to see ETC in other cases of Audacious Use? Yes, I think so, since the

only rhetorical ingredient required to deploy the move is some sort of moral complexity in the

technical decision-making context. Furthermore, the above examples suggest that there are case-specific variants and there may also be variants that are deployed across cases.

4. Ignoring Technical Concerns (ITC)

All that's required to deploy ITC is a collection of *nontechnical* arguments against tool use. A rhetor raises and addresses these, creating what appears to be a balanced synthesis of the issue, while ignoring technical concerns. This move seems also to be deployed in the case of DTC genetic testing, as is illustrated in an *NBC Washington* article titled, "Direct-to-Consumer Genetic Testing Debate: The Pros and Cons." The author makes no mention of technical concerns about information quality. Instead, she mentions the following objections to tool use:

- Some people would rather not know if they're at increased risk for a certain disease (Cleary, 2011, para. 1).

- "… average consumers aren't always equipped with the proper knowledge and understanding of test results" (para. 4).

- "The probabilities and odds that a DTC test might report to a consumer could lead to confusion or anxiety" (para. 5).

- "The concern is that consumers will adopt a kind of fatalistic attitude when they receive test results that put them at an increased risk for developing a disease" (para. 7).

I expect that in all cases of AUIT, because of the high-stakes nature of the decision or plans involved, there will be nontechnical as well as technical concerns about tool use. For this reason I conjecture that we are likely to observe ITC in other cases.

*   *   *

One way to understand the discoursal differences between the cases of VAM and genetic testing is to consider how "expert stake in the use of the tool" differs. Those with

methodological expertise relative to VAM do not have an obvious political or financial interest in whether their tool is used in high-stakes teacher evaluation. Many experts in predictive genetic testing, however, have a clear, personal interest in the use of their tool: they are health-care providers, for whom the nature of their work and their job security could change dramatically with widespread access to DTC genetic tests. This creates an inventional possibility for rhetors that doesn't exist in the case of VAM: questioning the impartiality of experts and, by extension, the validity or rigor of their technical concerns (i.e., a modified version of QMTC). Perhaps because this move is not available to rhetors in the case of VAM (where it would be hard to throw shade on, say, the American Statistical Association), other moves (like the original version of QMTC or MRTC) have more persuasive oomph.

The moral of the story here is that the messages that are compelling in a case of AUIT will depend in part on the specifics of the case: who the various stakeholders are, what their stake is, how tool use has evolved, whether legislation or regulation tends to support or oppose tool use, etc. The way these elements come together in the case of VAM lent persuasive potential to QMTC and MRTC. The way they came together in the case of genetic testing made a modified version of QMTC (one that foregrounded the possibly suspect motives of experts) particularly compelling, but rhetors didn't seem to make use of MRTC.

Despite this variation between cases of AUIT, we should expect to see some commonalities: we are likely to see ITC, ETC, and a version of QMTC, directed either at nonexperts (who learn of and raise expert concerns) or at experts. We may or may not see MRTC. And—at the risk of stating the obvious—we should not be surprised to see other moves not identified in this study.

D. Rhetorical Take-Aways

By identifying four rhetorical moves that rationalize AUIT, this study has implicitly suggested some ways the phenomenon might be avoided or resolved. I'll first point out strategies a nonscientist reader or listener might use to recognize the moves in action and thus limit their persuasive potential. Then I'll suggest some inventional possibilities for scientists who find that the moves may have circumvented their technical concerns.[18]

1. Take-Aways for Nonscientists

- QMTC works by steering a reader or listener away from the substance of a technical concern and towards the broader political context of tool use. A reader or listener can limit the persuasive potential of this move by dissociating the two strands of the debate that are being conflated. She might first consider whether she identifies, politically, with the group raising the concern or with what she understands to be their motivation for raising it. Having established this, she might set that dimension of the debate aside momentarily and turn to the substance of the technical concern and its implications for tool use. Or she might keep the following question in mind when she encounters arguments about the politics of tool use: *does the affiliation or motivation of the group raising this concern change the implications of the concern?*

- MRTC derives its rhetorical oomph from misleading juxtapositions. Detecting this move, therefore, depends on being able to recognize when a pair of scientific findings is mismatched. As the examples in Chapter 5 suggest, this mismatch can be extremely subtle, making it difficult to detect for the nonexpert reader or listener. She can try,

however, by watching for arguments in which a technical concern is raised and followed by a phrase like "however," or "on the other hand," or "but others have found." She should ask, *does the second scientific finding align with the technical concern… that is, do they pertain to the same issue?* If the answer is no, then she should reevaluate the author or speaker's (perhaps implicit) assertion that the technical concern can be or has been resolved.

- ETC rationalizes Audacious Use by sidestepping technical concerns and impugning opposition to tool use on moral grounds. A reader or listener can recognize this move by asking, *would this argument still hold water if the information provided by the tool were of low quality?* If the answer is no, the argument may become less compelling.

- Although ITC is the most conceptually straightforward move I identified, its power to shape audience perception should not be underestimated. It's a subtle move, drawing persuasive potential from what's *excluded* from an argument, and this may make it difficult to detect and deflect. One possible strategy is for a reader or listener to step back from any particular argument and locate—in the overall technical decision-making context—the key issues that are outside her expertise. To do this, she might ask herself, *in this debate, are there questions whose answers (1) I can't determine myself and (2) would shape my opinions about the use of the tool?* By cultivating an awareness of the kinds of questions that call for expertise, she may be more likely to notice when these questions are left out of an argument. What makes ITC so powerful, I think, is the "shape" of AUIT. It's bizarre. There's a highly technical object (i.e., the tool) snagged in a tangle of deeply controversial *nontechnical* questions. Furthermore, there's a common

---

[18] I remind the reader that this study did not analyze any expert discourse, including expert discourse in response to AUIT, so my suggestions should be viewed as thoughtful conjectures,

103

and tempting oversimplification of what the tool can do (in the case of VAM, "find the good and bad teachers"). The combined result, I believe, is a tendency to overlook the relevance of technical expertise. To the extent that nonscientists can recognize where that expertise is relevant, it may be possible to guard against the effect of ITC.

2. Take-Aways for Scientists

- When a scientist observes QMTC[19] at work (i.e., technical concerns have been associated with political interests or ulterior motives), a straightforward response to this move may be merely to clarify: while the technical concerns have been cited by a particular nonexpert group, they reflect the views of the scientific community. She may also point out that the technical properties of the tool are invariant to politicking. Capacious confidence intervals are capacious; it doesn't matter who brings them up. Furthermore, the imprecision of VAM scores concerns *anyone* interested in using them to evaluate teachers, irrespective of political affiliation or agenda.

- As noted in the previous section, misaligned findings give MRTC its persuasive power. This move might be countered simply by pointing out the misalignment and reiterating the technical concern. For example, a scientist might counter with "while X (the misaligned finding) is true, it only applies only in special cases, so it doesn't change the widespread concern about information quality" or "both of these findings are important and widely accepted; notice that they pertain to different issues and that information quality remains a widespread concern."

---

rather than evidence-based conclusions.
    [19] I'll discuss the version of QMTC I observed in the case of VAM, not the modified, expert-directed version that seems to operate in the case of genetic testing.

- When an argument for tool use is framed in moral terms (i.e., ETC has been deployed), it makes little sense for scientists to engage the rhetor on these terms, for the argument hinges on nontechnical issues where scientific expertise is beside the point. It may also be inadvisable simply to reiterate technical concerns, which will likely seem out of place in the argument. Instead, an effective strategy might be to ask the rhetor to imagine if and how her argument would change if the information returned by the tool were of low quality. By doing this, scientists compel the rhetor *herself* to connect the moral and technical dimensions of the issue, effectively "un-eclipsing" the technical concerns.

- When a discussion of the pros and cons of tool use excludes technical concerns (i.e., ITC), an expert might respond by first establishing that she'll not address the many important issues the rhetor has already raised, but rather she'll raise an issue that's been neglected. Then, she might point out that the issue she'll raise is one she's uniquely qualified to weigh in on: the methodological capacity of the tool. After this preemptive brush clearing, she'll be better situated to introduce her technical concern.

- My final suggestion doesn't relate to a particular rhetorical move; it's about expert discourse *in general* in cases of AUIT. As I pointed out in Chapter 1, a great deal of the controversy in cases of AUIT arises from contrasting values or beliefs, and as members of the public, scientists hold various of these values and beliefs. That is, they have reasons to support or oppose tool use that are unrelated to the tool's technical capacity. But if their goal is for nonexperts to appreciate the technical limitations of the tool, then I would advise them to die on the sword of methodological capacity and avoid the nontechnical dimensions of the debate. By limiting their public-facing discourse to technical issues, they foreground the unique perspective they bring to the conversations (knowledge of the inner workings of the tool), preserve their credibility as experts, and deflect accusations

of partisanship or politicking. If experts *do* choose to weigh in on questions of values and beliefs, then I would advise them to make this explicit. For example, "it's my opinion, as a parent not a statistician, that a test-based definition of teacher quality is too narrow." By distinguishing epideictic commentary from forensic commentary, experts may be able to express the former without jeopardizing the decision-making heft of the latter. This is a risky move, however, as it's well established (e.g., Fahnestock, 1986) that qualifying statements and nuance tend to disappear in popular reporting of scientific issues.

How effective are these take-aways likely to be in avoiding or resolving AUIT? There are many interrelated factors that might contribute to the avoidance or resolution of AUIT. So isolating the efficacy of the above suggestions (from the myriad other forces at work in the technical decision-making context) is methodologically less-than-tractable. A better question might be, "how likely is it that these strategies will have the intended audience effect?" I return to this issue briefly in the section titled "Future Work" below.

Regardless, it's important to remember that inoculating the public against persuasive, pro-AUIT arguments and helping scientists foreground their technical concerns may be only a first step towards resolving AUIT in cases like VAM. Resolution may require something more in these scenarios, where the policymakers who endorse Audacious Use may be aware of technical concerns about the tool, but support its use nonetheless.

E. A Note on Manufactured Scientific Controversy and VAM

In the review of my corpus, I found five experts[20] who tended to be quoted or to have their research cited in arguments for VAM-based teacher evaluation. I expected that these experts

---

[20] Dr. Erik Hanushek, Dr. William Sanders, Dr. Raj Chetty, Dr. John Friedman, and Dr. Jonah Rockoff. The last three—who conducted an influential study about VAM and the persistence of teacher effects—are nearly always referenced as a group.

would show up in rhetorical efforts to create a sort of Manufactured Scientific Controversy[21] (MSC) around technical concerns about VAM, but my expectation was wrong. Distortion of the status of the disciplinary debate turned out not to be a primary persuasive tactic in the case of VAM. I found a few instances where rhetors suggested disciplinary controversy[22], but this did not turn out to be "the primary message about the science-y part of the debate" in the case of VAM as it is in the case of anthropogenic global warming or other cases of MSC. Even in "expert-augmented" MRTC, where experts are juxtaposed across an adversative, the purpose of the tactic is not to portray "warring experts" but rather "resolved technical concerns." For this reason, I've chosen not to describe the case of VAM as a case of MSC.

F. Future Work

The rhetorical take-aways above pertain most directly to participants in instances of Audacious Use, but this study also has implications for scholars who study technical decision making and its language. Below, I articulate five questions this work has raised and suggest how future research might address them. The first three pertain Audacious Use, broadly speaking; the last two deal specifically with rhetorical dimensions of the phenomenon.

1. How Does AUIT Start?

By definition, tool use is not "audacious" until it has elicited expert caution. But might there be a way to recognize when a case of AUIT is brewing? If so, this might suggest ways to avoid it before it becomes established. Predictors of AUIT might be discovered by analyzing sources of public-facing expert discourse (e.g., congressional hearings, letters-to-the-editor, general-

---

[21] In this well-documented phenomenon (see, for example, Ceccarelli, 2011), rhetors create a sense of widespread expert controversy that doesn't actually exist.

audience books or magazine articles, etc.); perhaps expert caution rumbles here before it

coalesces into something more formal and united. Another tack would be to study historical

precursors to AUIT, two of which are hinted at in our established cases. Significant

technological and methodological advances in the mid 2000's made it feasible and cheap to

produce the genetic risk profiles whose use eventually became audacious (Black & Clark, 2016),

suggesting that developments within science may be precursors to AUIT. The use of VAM in the

evaluation of individual teachers came into vogue when decades of accountability-based

education policy (Asen, 2012) became focused on teacher quality (Polikoff & Porter, 2014):

perhaps societal exigence can invite (or even rush) the novel use of a tool if that use appears to

solve a pressing problem.

2. How Does AUIT Stop?

The dissonance that defines AUIT (between scientists and tool users) might be resolved in a

few ways. The tool may change: its methodological capacity may develop to support the

proposed use. The use may change: different, lower-stakes decisions may be made with the tool.

Or the use may stop altogether, for a variety of reasons.

For example, litigation or government regulation may discourage the use, which is what

began to happen in 2013 in the case of DTC genetic testing: the FDA forced one company,

23&Me, to stop marketing nearly all of its "genetic risk profiles" (Gutierrez, 2013).

Alternatively, those negatively affected by the use may amplify scientists' caution beyond the

realm of peer-reviewed literature or policy statements and into mainstream discourse, which may

turn public opinion against the tool use. This may partially explain why the case of VAM is

evolving as it is at the time of this writing. The most recent reauthorization of the Elementary

---

[22] The reader may have noted one such example in Chapter 5, where Hanushek is referred to

and Secondary Education Act no longer presses for VAM-based teacher evaluation (Every Student Succeeds Act, 2015), and it remains to be seen how this will ripple into state policies.

Using the tools of historical, policy, or media analysis, we can study how AUIT resolution has been pursued, facilitated, complicated, thwarted, or successfully brokered in past cases… which might provide insight into how to expedite resolution in the future.

3. What are Other Cases of AUIT?

The type-naming work of Chapter 2 involved a broad survey of technical decision-making contexts, and two of these stood out as likely instances of AUIT. A careful analysis of these cases (of the sort conducted in Chapter 3) was outside the scope of the dissertation. Both cases appear to involve the following elements of AUIT: an information tool taken out of a research context, high-stakes decisions or plans, and expert concern about the methodological suitability of the tool to the proposed use. The future work that's needed in both cases is to establish: whether there is widespread agreement in the scientific community about the concerns expressed and whether the concerns deal specifically with issues of precision, consistency, and accuracy.

The first potential case is the current use of Dynamic Stochastic Generalized Equilibrium (DSGE) models and similar macroeconomic models to make economic policy decisions and plans. Two years after the economic crash of 2008, these widely used models were the focus of a congressional hearing whose purpose was to understand the appropriate use and limitations of the models, to question their suitability for prediction and policy setting, and to figure out "how… we get out of the mess we are in" (Building a Science, 2010, p. 7). Nobel Laureate economist Robert Solow of MIT testified in this hearing, expressing strong caution about this use of DSGE models. The heart of his argument was that the models are based on an "essentially implausible

as "a supporter of value-added measures" (Brown, 2015, para. 9).

assumption" (Building a Science, 2010, p. 13). Others have expressed even stronger caution about this use of the models, deeming them "too dangerous to be used for forecasting purposes and even more dangerous for policy making" (Garcia, 2011, p. 169).

The second case I'll suggest is the use (during the 2000's) of predictive quantitative mathematical models of earth processes in environmental and natural resource management. One example, in particular, is the use of a set of modeling techniques called the Total System Performance Assessment (TSPA) to make million-year predictions about the dynamics of nuclear waste storage in Yucca Mountain. This time-span, some experts suggest, reflects a "massively misplaced confidence in predictive mathematical models" (Pilkey-Jarvis & Pilkey, 2008, p. 474), given uncertainties that arise from "the data gaps, assumptions, weightings, and extrapolations on which mathematical models are inevitably based" (p. 470).

4. What Additional Rhetorical Moves are Deployed in the Case of VAM?

Not surprisingly, there are interesting rhetorical moves in my corpus that didn't make their way into this dissertation. In order to give adequate attention to the four moves I presented, I set aside two moves for future research. The first I referred to as "Rose-Colored Stakes," in which a rhetor emphasizes the positive or low-stakes consequences of tool use and ignores negative or high-stakes consequences. The second move is "Creating Urgency," in which a rhetor frames her argument in the context of a pressing problem, emphasizes the need to do *something* to solve the problem, and then suggests that the Audacious Use of the tool is the solution. While I've cultivated a general sense of how these moves work (and gathered plenty of significant examples), additional research is needed. The structure and mechanics of the moves need to be thoroughly analyzed—including their effect on the reader or listener—and these results need to be written up in a coherent and compelling way.

5. Are the Rhetorical Moves Identified in the Case of VAM Deployed in Other Cases of AUIT?

A thorough rhetorical analysis of the case of DTC genetic testing is the first and most obvious step to investigate whether we should expect to see the rhetorical moves identified in this dissertation deployed in other cases of AUIT. Additionally, if the two cases mentioned above turn out to be instances of AUIT, analyzing their language will solidify our understanding of the rhetoric of Audacious Use even further.

6. Do AUIT-Rationalizing and AUIT-Countering Rhetorical Moves Work the Way We Think They Do?

As in all rhetorical analyses, a key step in this study was to imagine the cognitive or emotional effect of a persuasive tactic on a reader or listener. This conjectural work showed up in two places in this study: when I explained how my AUIT-rationalizing rhetorical moves work (in Chapter 5) and when I suggested strategies scientists might use to counter these moves (previously in this chapter). Verifying this conjectural work (i.e., directly investigating how a reader or listener experiences a rhetorical move) might entail cognitive interviews, focus groups, surveys, or a content analysis of comments posted online in response to articles in which identified rhetorical moves are deployed. Notice that such work is not intended to provide direct evidence that AUIT-rationalizing moves *cause* AUIT or that AUIT-countering moves *counter* AUIT. A causal research goal like this is too simplistic, given the many forces at work in AUIT (politics, economics, social and ethical considerations, values, beliefs, etc.). Rather, the goal of such work is to understand audience perception of particular rhetorical moves, while acknowledging that these moves are just one part of a complex public decision-making context.

G. Importance of the Dissertation

1. Importance in Academia

By naming AUIT, this dissertation has called attention to a previously overlooked type of public decision making. STS scholarship tends to focus on two main types of science-society interactions. In the first, the interaction centers around the application of what we might call a "change tool," a technique or instrument that somehow changes the state of the world (e.g., the use of genetic modification or nanotechnology in food production, the use of stem cells in medical therapies, fracking, etc.). In the second, the interaction hinges on establishing the truth value of a knowledge claim (e.g., global warming, AIDS dissent, the smoking-cancer connection, etc.). This study demonstrates that a third type of science-society interaction needs our attention: interactions where information tools are put to novel, out-of-science uses.

These interactions need our attention because they are importantly different from those involving "change tools" or knowledge claims. How? They involve *interpretation* of tool output by nonscientists. Notice that this interpretation is not part of public decision making around, say, climate change or GMOs. Very little interpretive effort is required to make sense of the statement, "the earth is getting warmer and humans are responsible." Nonscientists don't interpret GMO-tomatoes; they choose whether or not to eat them.

When a nonscientist interprets output from an information tool, she may not be aware of the overall research endeavor that should inform this interpretation (e.g., research design, data collection strategy, etc.[23]). Her understanding of tool output may be shaped more by popular, policy, or commercial accounts of what the tool can tell, and less by expert accounts. When information tools are used to study intuitive but psychometrically thorny constructs (e.g.,

---

[23] For further elaboration of this idea, see Figure 2-1 and the adjacent discussion in Chapter 2.

"intelligence" or "teacher quality"), her interpretation may be based on her intuitive understanding (e.g., "this IQ score means this person is smarter than that one" or "this VAM score means this teacher is better than that one") and may not reflect the relevant psychometric complexities. A scholarly STS lens can help us better understand these interpretive dynamics in the context of public understanding, technical decision making, and expertise. And this better understanding may provide strategies that scientists, policymakers, and members of the public can use to encourage productive and responsible use of information tools.

This dissertation serves an important function for scholars of rhetoric, as well. It points out an unexplored but critically important territory in this discipline: teacher quality and evaluation. To be sure, there are areas of scholarship adjacent to this territory. Phillips (2004), for example, has traced the evolution of discourse about high-stakes educational testing over several decades. In his analysis of presidential discourse, Asen (2012) has identified key terms used to allocate agency and responsibility in K-12 education policy. Persuasive strategies used to support the privatization of schools have been documented (Weathers, 2007), as have some of the powerful myths used to define the societal role of teachers (Steudeman, 2014) and professors (Winslow, 2015). But what this dissertation has shown is that "the discourse of teacher quality and evaluation" is a distinct rhetorical sphere, one where a great deal of work is needed as soon as possible, considering the rapid pace of teacher-focused education reforms.

2. Importance in Education Reform

Like many other sciences, the field of educational and psychological measurement holds itself to professional standards. These clarify what constitutes ethical, responsible, and technically viable application of the science's tools. When experts came together to express concern about the Audacious Use of VAM, they were cooperatively applying Standard 13.3:

"When accountability indices, indicators of effectiveness in program evaluations or policy studies, or other statistical models (such as value-added models) are used, … their technical qualities should be reported" (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 2014, p. 209-210).

Somehow, this collective attempt to apply professional standards of practice didn't curb the technically inadvisable use of VAM, and what this dissertation has done is help us understand how—rhetorically—this might have happened. This insight is important because a primary purpose of the standards is to prevent tool-misapplications that could lead to unsound decisions. In the case of VAM, these might include retaining poor teachers, dismissing good ones, or publishing a misleading or inaccurate account of a teacher's professional competence.

Are these unsound decisions likely or common? One way to answer this is to consider the degree of uncertainty involved in translating a VAM score into a qualitative classification (e.g., "developing" or "highly effective"). A report commissioned by the Department of Education shortly after the introduction of RTT found that, when models are specified with three years of data, one in four teachers is likely to be misclassified; with only one year of data, this becomes one in three teachers (Schochet & Chiang, 2010). For teachers whose scores hover at the border between categories that are associated with different professional consequences, the uncertainty inherent in VAM-based teacher evaluation poses the risk of a career-changing classification error.

This dissertation has focused on the Audacious Use of VAM in the evaluation of *individual* teachers, but its importance quickly ripples outward from any individual teacher or classroom of students. As John Ewing, former Executive Director of the American Mathematical Society, points out in his discussion of the broader implications of VAM-based teacher evaluation:

If we decide whether alternative certification is better than regular certification, whether nationally board certified teachers are better than randomly selected ones, whether small schools are better than large, or whether a new curriculum is better than an old by using a flawed measure of success, we almost surely will end up making bad decisions that affect education for decades to come. (Ewing, 2011, p. 671)

3. Importance, More Broadly

The value of this dissertation—beyond education policy and outside of academia—is extremely simple: someone who has heard of AUIT is sensitized to the possibility that it might occur. She is primed to listen for (and perhaps seek out) scientific commentary on the methodological suitability of information tools put to novel uses. Furthermore, when AUIT is on the radar of scientists, they may shift how they communicate about technical concerns, perhaps in ways that change how these concerns are perceived.

Why does any of this matter? Imagine that the use of crystal balls to predict and plan for the distant future has become mainstream. Imagine that this has sparked a complex societal conversation about who should have access to crystal balls, the implications and unintended consequences of their use, whether individuals should be allowed to predict the future of others without their consent, etc. And imagine that, in the midst of this conversation, those who make, use, and study crystal balls stand up and say, "guys, wait… crystal balls can't actually tell the future." This is not just one voice among many—it is a conversational game-changer. For if those in a position to know tell us that the tool may not be able to do what we think it can, then the other parts of the conversation become moot.

*   *   *

In closing, my hope is that this study proves useful to policymakers who seek to promote and regulate the use of information tools in responsible ways, scientists who are concerned that the

tools be used safely and effectively, and nonscientists who strive to make sense of and decisions

about the use of these tools in their lives.

Bibliography

23andMe. (2013, November 25). 23andMe statement regarding FDA warning letter [Web log comment]. Retrieved from http://blog.23andme.com/news/23andme-statement-regarding-fda-warning-letter/

Abeysinghe, S. (2014). An uncertain risk: The World Health Organization's account of H1N1. *Science in Context, 27*(3), 511–529.

Abramson, L. (News Correspondent). (2010, August 27). 'L.A. Times' teacher ratings database stirs debate [Radio program]. *Morning Edition*. Washington, DC: National Public Radio. Retrieved from http://www.npr.org/templates/story/story.php?storyId=129456212

Aldeman, C. & Chuong, C. (2014). *Teacher Evaluations in an Era of Rapid Change: From "unsatisfactory" to "needs improvement."* Washington, DC: Bellwether Education Partners. Retrieved from http://bellwethereducation.org/sites/default/files/Bellwether_TeacherEval_Final_Web.pdf

Alkin, M.C. (Ed.). (2012). *Evaluation roots* (2nd ed.). Thousand Oaks, CA: Sage.

Allum, N., Sturgis, P., Tabourazi, D. & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: A meta-analysis. Public Understanding of Science, 17(1), 35-54.

Ameer, B. & Krivoy, N. (2009). Direct-to-consumer/patient advertising of genetic testing: A position statement of the American College of Clinical Pharmacology. *Journal of Clinical Pharmacology, 49*(8), 886–888.

American College of Medical Genetics. (2008). *ACMG Statement on direct-to-consumer testing*. Retrieved from http://www.acmg.net/AM/Template.cfm?Section=Policy_Statements&Template=/CM/ContentDisplay.cfm&ContentID=2975

American College of Obstetricians and Gynecologists. (2008). Direct-to-consumer marketing of genetic testing. *Obstetrics and Gynecology, 111*(6), 1493–1494.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association. (2015). AERA Statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448 – 452.

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Alexandria, VA: Author. Retrieved from http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Anderson, G. L. (2007). Media's impact on educational policies and practices: Political spectacle and social control. *Peabody Journal of Education, 82*(1), 103-120.

Asen , R. (2012). Lyndon Baines Johnson and George W. Bush on education reform: Ascribing agency and responsibility through key policy terms. *Rhetoric & Public Affairs, 15*(2), 289-317.

Asen, R., Gurke, D., Conners, P., Solomon, R., & Gumm, E. (2013), Research evidence and school board deliberations: Lessons from three Wisconsin school districts. *Educational Policy, 27*(1), 33-63.

Audacious [Def 2]. (n. d.). In *Oxford Dictionaries*. Retrieved from http://www.oxforddictionaries.com/us/definition/american_english/audacious

Bachman, L. & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Baird, B. (2015). Perspective: Informing public policy with social and behavioral science. *Issues in Science and Technology, 31*(3). Retrieved from http://issues.org/31-3/perspective-informing-public-policy-with-social-and-behavioral-science/

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from http://www.epi.org/publications/entry/bp278

Ballou, D. & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher, 44*(2), 77–86.

Banning, M. E. (2009). When poststructural theory and contemporary politics collide: The vexed case of global warming. *Communication and Critical / Cultural Studies, 6*(3), 285 - 304.

Bauer, M. W. & Aarts, B. (2000). Corpus construction: A principle for qualitative data collection. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image, and sound: A practical handbook* (pp. 19-37). Thousand Oaks, CA: Sage Publications Inc.

Bauer, M. W. (2000a). Classical content analysis, In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image, and sound: A practical handbook* (pp. 131-151). Thousand Oaks, CA: Sage Publications Inc.

Bauer, M. W. (Ed.). (2000b). *Qualitative researching with text, image, and sound: A practical handbook.* Thousand Oaks, CA: Sage Publications Inc.

Bauer, M. W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science, 16*(1), p. 79-95.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Bell, A. (1994). Climate and opinion: Public and media discourse on the global environment. *Discourse Society, 5*(1), 33-64.

Bitzer, L. F. (1968). The Rhetorical Situation. *Philosophy and Rhetoric, 1*(1), 1 – 14.

Black, J. R. M. & Clark, S. J. (2016). Age-related macular degeneration: Genome-wide association studies to translation. *Genetics in Medicine, 18*(4), 283 – 289.

Boorsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110,* 203-219.

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information, 14*(6), 19-47.

Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher, 44*(2), 127–131.

Briggle, A. (2008). Questioning expertise. [Review of the books *The honest broker: Making sense of science in policy and politics*, by Pielke, R. A., Jr. and *The philosophy of expertise*, edited by E. Selinger and R. P. Crease]. *Social Studies of Science, 38*(3), 461–475.

Briggs, D. & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the* Los Angeles Times. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/due-diligence

Brock, B. L., Scott, R. L., & Chesebro, J. W. (1990). An introduction to rhetorical criticism. In B. L. Brock, R. L. Scott, & J. W. Chesebro (Eds.), *Methods of rhetorical criticism* (pp. 9-22). Detroit, MI: Wayne State University Press.

Brock, B. L., Scott, R. L., & Chesebro, J. W. (Eds.). (1990). *Methods of rhetorical criticism* (3rd ed.)*.* Detroit, MI: Wayne State University Press.

Bromme, R. & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist, 49*(2), 59-69.

Brookhart, S. M. (2009). The many meanings of multiple measures. *Educational Leadership, 67*(3), 6-12.

Brown, E. & Balingit, M. (2015, March 16). Virginia pushed into debate of teacher privacy vs. transparency for parents. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/parent-suing-state-officials-to-make-teacher-evaluation-data-public/2015/03/15/9b441a58-c98f-11e4-b2a1-bed1aaea2816_story.html?utm_term=.ed8f6cba927c

Brown, E. (2015, February 28). Contentious teacher-related policies moving from legislatures to the courts. *Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/contentious-teacher-related-policies-moving-from-legislatures-to-the-courts/2015/02/28/6ec1f31e-b83d-11e4-aa05-1ce812b3fdd2_story.html?utm_term=.1089f34dc258

Bruner, J. S. & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality, 18*, 206-223.

Building a science of economics for the real world: Hearings before the Subcommittee on Investigations and Oversight, of the House Committee on Science and Technology, 111th Cong. (2010) (testimony of Robert Solow). Retrieved from https://www.gpo.gov/fdsys/pkg/CHRG-111hhrg57604/pdf/CHRG-111hhrg57604.pdf

Byers, M. & Johnson, M. (Eds.). (2009). *The CIS effect: Television, crime, and governance*. Lanham, MD: Lexington Books.

Carusi, A. (2012). Making the visual visible in philosophy of science. *Spontaneous Generations: A Journal for the History and Philosophy of Science, 6*(1), 106-114.

Ceccarelli, L. (2011). Controversy over uncertainty: Argumentation scholarship and public debate about science. In F. H. van Eemeren, B. Garssen, D. Godden, & G. Mitchell (Eds.), *Proceedings of the seventh conference of the international society for the study of argumentation* (pp. 254-260). Retrieved from http://rozenbergquarterly.com/issa-proceedings-2010-controversy-over-uncertainty-argumentation-scholarship-and-public-debate-about-science/

Ceccarelli, L. (2011). Manufactured scientific controversy: Science, rhetoric, and public debate. *Rhetoric & Public Affairs, 14*(2), 195-228.

Center for Education Reform. (2010). *Making the grade? A report card on performance pay programs across the US*. Washington, DC: Author. Retrieved from https://www.edreform.com/wp-content/uploads/2011/10/MakingTheGrade-PolicyAlert.pdf

Cleary, L. (2011, January 17). Direct-to-consumer genetic testing debate: The pros and cons. *NBC Washington*. Retrieved from http://www.nbcwashington.com/news/health/Direct-to-Consumer-Genetic-Testing-Debate-112986579.html

Coles, D. & Frewer, L. J. (2013). Nanotechnology applied to European food production: A review of ethical and regulatory issues. *Trends in Food Science & Technology, 34*(1), 32 – 43.

Collins, H. (2004). Interactional expertise as a third kind of knowledge. *Phenomenology and the Cognitive Sciences, 3*, 125–143.

Collins, H. M. & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science, 32*(2), 235–296.

Condit, C. M., Lynch, J., & Winderman, E. (2012). Recent rhetorical studies in public understanding of science: Multiple purposes and strengths. *Public Understanding of Science, 21(4)*, 386-400.

Connecticut Department of Public Health. (2009). *Direct-to-Consumer Personal Genomics Services: Information to Consider*. Hartford, CT: Author.

Coopmans, C. & Button, G. (2014). Eyeballing expertise. *Social Studies of Science, 44*(5), 758-785.

Corcoran, S & Goldhaber, D. (2013). Value added and its uses: Where you stand depends on where you sit. *Education Finance and Policy, 8*(3), 418-434.

Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: Annenberg Institute for School Reform, Brown University. Retrieved from http://www.annenberginstitute.org/publications/can-teachers-be-evaluated-their-students%E2%80%99-test-scores-should-they-be-use-value-added-me

Cormick, C. (2009). Why do we need to know what the public thinks about nanotechnology? *Nanoethics, 3*(2), 167 – 173.

Correnti, R. & Martínez, J. F. (2012). Conceptual, methodological, and policy issues in the study of teaching: Implications for improving instructional practice at scale. *Educational Assessment, 17*(2-3), 51-61.

Creswell, J. W. (2007). *Qualitative enquiry and research design: Choosing among five approaches*. London, UK: SAGE Publications.

Crick, N. & Gabriel, J. (2010). The conduit between lifeworld and system: Habermas and the rhetoric of public scientific controversies. *Rhetoric Society Quarterly, 40*(3), 201 – 223.

Critchley, C., Nicol, D., Otlowski, M., & Chalmers, D. (2015). Public reaction to direct-to-consumer online genetic tests: Comparing attitudes, trust and intentions across commercial and conventional providers. *Public Understanding of Science, 24*(6), 731 – 750.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich, Inc.

Cronbach, L. J. (1980). Selection theory for a political world. *Public Personnel Management, 9*, 37-50.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher, 44*(2), 132–137.

Darling-Hammond L., Amrein-Beardsley A., Haertel E. H., & Rothstein J. (2011). *Getting teacher evaluation right: A background paper for policy makers.* Washington, DC: American Educational Research Association and National Academy of Education. Retrieved from

http://www.aera.net/Portals/38/docs/News_Media/AERABriefings/Hill%20Brief%20-%20Teacher%20Eval%202011/GettingTeacherEvaluationRightBackgroundPaper(1).pdf

Dickinson, G. (2002). Joe's rhetoric: Finding authenticity at Starbucks. Rhetoric Society Quarterly, 32(4), 5–27.

Dillon, S. (2010, August 31). Formula to grade teachers' skill gains acceptance, and critics. *The New York Times*. Retrieved from http://www.nytimes.com/2010/09/01/education/01teacher.html

Direct-to-consumer genetic testing and the consequences to the public: Hearings before the Subcommittee on Oversight and Investigations, of the House Committee on Energy and Commerce, 111th Cong. (2010) (testimony of Jeffrey Shuren). Retrieved from http://www.fda.gov/newsevents/testimony/ucm219925.htm

District of Columbia Public Schools. (2016). *General education teachers with individual value-added student achievement data*. Retrieved from http://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/1GeneralEducationTeacherswithIndividualValue-AddedStudentAchievementData.pdf

Dobbs, D. (2013, November 27). The F.D.A. vs. personal genetic testing. *The New Yorker.* Retrieved from http://www.newyorker.com/tech/elements/the-f-d-a-vs-personal-genetic-testing

Drabek, M. L. (2009). Interactive classification and practice in the social sciences: Expanding Ian Hacking's treatment of interactive kinds. *Poroi, 6*(2), 62-80.

Dreyfus, H. (2006). How far is distance learning from education? In E. Selinger & R. P. Crease (Eds.), *The philosophy of expertise* (pp. 196-212).

Duncan, A. (2012, February 7). Fighting the wrong education battles: Remarks of Secretary Duncan at the Askwith Forum, Harvard Graduate School of Education. Retrieved from http://www.ed.gov/news/speeches/fighting-wrong-education-battles

Duncan, T. V. (2011). Applications of nanotechnology in food packaging and food safety: Barrier materials, antimicrobials and sensors. *Journal of Colloid and Interface Science, 363*(1), 1–24.

Dunwoody, S. (1992). The challenge for scholars of popularized science: Explaining ourselves. *Public Understanding of Science, 1*(1), 11-14.

Ebbesen, M., Andersen, S., & Besenbacher, F. (2006). Ethics in nanotechnology: Starting from scratch? *Bulletin of Science, Technology & Society, 26*(6), 451 – 462.

Ericsson, K. A. (2012, October 28). The danger of delegating education to journalists: Why the APS Observer needs peer review when summarizing new scientific developments [Web log comment]. Retrieved from https://radicalscholarship.wordpress.com/2014/11/03/guest-post-the-danger-of-delegating-education-to-journalists-k-anders-ericsson/

European Society of Human Genetics. (2010). Statement of the ESHG on direct-to-consumer genetic testing for health-related purposes. *European Journal of Human Genetics, 18*, 1271–1273.

Evans, W. & Hornig-Priest, S. (1995). Science content and social context. *Public Understanding of Science, 4*(4), 327-340.

Every Student Succeeds Act, 114th Cong. (2015). Retrieved from https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf

Ewing, J. (2014). Mathematical intimidation: Driven by the data. *Notices of the American Mathematics Society, 61*(5), 667-673.

Fahnestock, J. & Secor, M. (1988). The stases in scientific and literary argument. *Written Communication, 5*, 427-443.

Fahnestock, J. (1986). Accommodating science: The rhetorical life of scientific facts. *Written Communication, 3*(3), 275-296.

Fahnestock, J. (2011). *Rhetorical style: The uses of language in persuasion*. New York, NY: Oxford University Press, Inc.

Farrell, T. B. & Goodnight, G. T. (1981). Accidental rhetoric: The root metaphors of Three Mile Island. *Communication Monographs, 48*(4), 271-300.

Felch, J. & Song, J. (2010, August 16). U.S. schools chief endorses release of teacher data. *The Los Angeles Times*. Retrieved from http://articles.latimes.com/2010/aug/17/local/la-me-0817-teachers-react-20100817-

Felch, J., Song, J., & Smith, D. (2010, August 14). Who's teaching L.A.'s kids? *The Los Angeles Times*. Retrieved from http://articles.latimes.com/2010/aug/14/local/la-me-teachers-value-20100815

Ferdman, R. A. (2015, July 6). Why we're so scared of GMOs, according to someone who has studied them since the start. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2015/07/06/why-people-are-so-scared-of-gmos-according-to-someone-who-has-studied-the-fear-since-the-start/

Florida Department of Education. (2014). *Annual legislative report on teacher evaluation*. Retrieved from http://www.fldoe.org/core/fileparse.php/7503/urlt/1314AnnualLegisReportTeacherEval.pdf

Florida Department of Education. (n.d.). *Overview of Florida's teacher evaluation system*. Retrieved from http://www.fldoe.org/core/fileparse.php/7503/urlt/0102688-overviewfloridasteacherevaluationsystem.pdf

Frankfurt, H. G. (2005). *On bullshit*. Princeton, NJ: Princeton University Press.

Freudenburg, W. R., Gramling, R., & Davidson, D. J. (2008). Scientific Certainty Argumentation Methods (SCAMs): Science and the politics of doubt. *Sociological Inquiry, 78*(1), 2-38.

Funtowicz S. & Strand, R. (2007). Models of Science and Policy. In Traavik, T. & and Lim, L.C. (Eds.), B*iosafety first: Holistic approaches to risk and uncertainty in genetic engineering and genetically modified organisms* (pp. 263 – 278). Trondheim, Norway: Tapir Academic Publishers.

Funtowicz, S & Ravetz, J. (n.d.). *Post-normal science: Environmental policy under conditions of complexity*. Retrieved from http://www.nusap.net/sections.php?op=viewarticle&artid=13

Funtowicz, S. & Ravetz, J. (2003). Post-normal science. *Online encyclopedia of ecological economics.* Retrieved from: http://isecoeco.org/pdf/pstnormsc.pdf

Funtowicz, S. O. & Ravetz, J. R. (1993). Science for the post-normal age. *Futures, 25*(7), 739-755.

Gabriel, R. & Lester, J. N. (2013). The romance quest of education reform: A discourse analysis of The LA Times' reports on value-added measurement teacher effectiveness. *Teachers College Record, 115*(12), 1-32.

Garcia, N. E. (2011). DSGE macroeconomic models: A critique. *Economie Appliquée, 64*(1), 149 – 171.

Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gerring, J. (2012). *Social science methodology: A unified framework* (2nd ed.). New York, NY: Cambridge University Press.

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. Chicago, IL: University of Chicago Press.

Glazerman, S., Loeb, S. Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: The Brookings Institution.

Goldhaber, D. & Theobald R. (2012). *Do different value-added models tell us the same things?* Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from http://carnegieknowledgenetwork.org/briefs/value-added/different-growth-models

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher, 44*(2), 87–95.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44*(2), 96–104.

Goldstein, R. A. (2011). Imaging the frame: Media representations of teachers, their unions, NCLB, and education reform. *Educational Policy, 25*(4), 543-576.

Gopnik, A. (2015, November 30). Spooked: What do we learn about science from a controversy in physics? *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2015/11/30/spooked-books-adam-gopnik

Gordon, R. Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: The Brookings Institution.

Gregory, J. & Miller, S. (1998). *Science in Public: Communication, Culture, and Credibility*. New York, NY: Plenum Press.

Gross, A. G. (1994). The roles of rhetoric in the public understanding of science. *Public Understanding of Science, 3*(1), 3-23.

Gross, A. G. (2006). *Starring the text: The place of rhetoric in science studies*. Carbondale, IL: Southern Illinois University Press.

Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Introduction to the special issue: Policy and research challenges of moving toward best practices in using student test scores to evaluate teacher performance. *Journal of Research on Educational Effectiveness, 8*(1), 1 – 7.

Gupta, N., Fischer, A. R. H., & Frewer, L. J. (2015). Ethics, risk and benefits associated with different applications of nanotechnology: A comparison of expert and consumer perceptions of drivers of societal acceptance. *Nanoethics, 9*(2), 93–108.

Gusfield, J. (1981). *The culture of public problems: Drinking-driving and the symbolic order*. Chicago, IL: University of Chicago Press.

Gutierrez, A. (2013). *Warning letter to 23andMe, Inc*. Retrieved from www.fda.gov/iceci/enforcementactions/warningletters/2013/ucm376296.htm

Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Princeton, NJ: Education Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/PICANG14.pdf

Hamilton, A. (2008, Oct. 29). Invention of the year: The retail DNA test. *Time Magazine*. Retrieved from http://content.time.com/time/specials/packages/article/0,28804,1852747_1854493,00.html

Hammersley, M. (2014). *The limits of social science: Causal explanation and value relevance*. Thousand Oaks, CA: SAGE Publications Ltd.

Hanushek, E. (2010, October 27). Parents deserve teacher data. *The New York Daily News*. Retrieved from http://www.nydailynews.com/opinion/uft-wrong-fight-joel-klein-attempt-release-teacher-data-leading-education-researcher-article-1.190233

Harding, S. (1998). *Is science multicultural? Postcolonialisms, feminisms and epistemologies.* Bloomington, IN: Indiana University Press.

Hardwig, J. (2006). Epistemic dependence. In E. Selinger & R. P. Crease (Eds.), *The philosophy of expertise* (pp. 328-341).

Harris, D. N. & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher, 44*(2), 71–76.

Harris, D. N. (2011). *Value-added measures in education*. Cambridge, MA: Harvard Education Press.

Hartas, D. (Ed.). (2010). *Educational research and inquiry: Qualitative and quantitative approaches.* London: Continuum International Publishing Group.

Hawhee, D. (2009). *Moving bodies: Kenneth Burke at the edges of language*. Columbia, SC: University of South Carolina Press.

Heider, K. G. (1988). The Rashomon effect: When ethnographers disagree. *American Anthropologist, 90*(1), 73-81.

Heinrick, J. (2006). Everyone's an expert: The CSI effect's negative impact on juries. *The Triple Helix, 3*(1), 59-61. Retrieved from http://www.cspo.org/documents/csieffectheinrick.pdf

Helgason, A., & Stefánsson, K. (2010). The past, present, and future of direct-to-consumer genetic tests. *Dialogues in Clinical Neuroscience*, *12*(1), 61–68.

Henig, J. R. (2008). *Spin cycle: How research is used in policy debates, the case of charter schools.* New York, NY: Russell Sage Foundation.

Hermansen, P. M. (2014). "There was no one coming with enough power to save us": Waiting for "Superman" and the rhetoric of the new education documentary. *Rhetoric & Public Affairs, 17*(3), 511 – 540.

Hiebert, J. & Morris, A. K. (2012). Teaching, rather than teachers, as a path toward improving classroom instruction. *Journal of Teacher Education, 63*(2), 92-102.

Ho, T. (2015). *Exploring the decision making process in statistical data analysis: A qualitative study of quantitative researchers* (Unpublished doctoral dissertation). University of California Los Angeles, Los Angeles, CA.

Holloway-Libell, J. & Amrein-Beardsley, A. (2015). "Truths" devoid of empirical proof: Underlying assumptions surrounding value-added models (VAMs) in teacher evaluation

[Commentary]. *Teachers College Record*. Retrieved from
https://www.tcrecord.org/Content.asp?ContentID=18008

Howitt, S. M. & Wilson, A. N. (2014). Revisiting "Is the scientific paper a fraud?" *EMBO Reports, 15*(5), 481–484.

Hu, W. (2012, March 15). Most New York voters approve of city's making teacher ratings public. *The New York Times*. Retrieved from
https://query.nytimes.com/gst/fullpage.html?res=9403E4D8103BF936A25750C0A9649D8B63

Hudson, K., Javitt, G., Burke, W., & Byers, P. (2007). ASHG statement on direct-to-consumer genetic testing in the United States. *American Journal of Human Genetics, 81*(3), 635 – 637.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance.* Alexandria, VA: Center for Public Education.

Hume, D. (2004). *An enquiry concerning human understanding, (posthumous ed.).* Mineola, NY: Dover Publications Inc.

Hunter, D. J., Khoury, M. J., & Drazen, J. M. (2009). Letting the genome out of the bottle: Will we get our wish? *New England Journal of Medicine, 358*(2), 105 – 107.

IdeaStream. (2013, June 16). *Grading the teachers: Why we're publishing teachers' value-added ratings*. Retrieved from http://wcpn.ideastream.org/stateimpact/2013/06/16/grading-the-teachers-why-were-publishing-teachers-value-added-ratings

Imai, K., Kricka, L. J., & Fortina, P. (2011). Concordance study of 3 direct-to-consumer genetic-testing services. *Clinical Chemistry, 57*(3), 518 – 521.

Institute of Education Sciences. (2012). Learning from recent advances in measuring teacher effectiveness. Washington, DC: Author. Retrieved from
http://ies.ed.gov/director/pdf/measuring_teacher_effectiveness.pdf

Institute of Medicine. (2009). *Nanotechnology in food products: Workshop summary.* Washington, DC: The National Academies Press. Retrieved from
http://www.nap.edu/catalog/12633/nanotechnology-in-food-products-workshop-summary

International Society of Nurses in Genetics (2009). *Position statement: Direct-to-consumer marketing of genetic tests*. Retrieved from
http://www.isong.org/pdfs2013/PS_Marketing_Genetic_Tests.pdf

Jasanoff, S. (2008). Speaking honestly to power. [Review of the book *The honest broker: Making sense of science in policy and politics*, by Pielke, R. A., Jr.]. *American Scientist, 96*(3), 240-243.

Jasanoff, S. (2014). A mirror for science. *Public Understanding of Science, 23*(1), 21-26.

Jasanoff, S. S. (1987). Contested boundaries in policy-relevant science. *Social Studies of Science, 17*(2), 195-230.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105–116.

Jordan, J. W. (2004) The rhetorical limits of the "plastic body." *Quarterly Journal of Speech, 90*(3). 327–358.

Jordan, J.W. (2009) Reshaping the "pillow angel": Plastic bodies and the rhetoric of normal surgical solutions. *Quarterly Journal of Speech, 95*(1), 20–42.

Kasanmoentalib, S. (1996). Science and values in risk assessment: The case of deliberate release of genetically engineered organisms. *Journal of Agricultural and Environmental Ethics, 9*(1), 42-60.

Keller, E. F. (1985). *Reflections on gender and science*. New Haven, CT: Yale University Press.

Keränen, L. (2010). *Scientific characters: Rhetoric, politics, and trust in breast cancer research*. Tuscaloosa, AL: University of Alabama Press.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford: Oxford University Press.

Klein, J., Rhee, M., Gorman, P. C., Huberman, R., Johnson, C. R., Alonso, A. A., … Sheffield, L. (2010, October 10). How to fix our schools: A manifesto by Joel Klein, Michelle Rhee and other education leaders. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/10/07/AR2010100705078.html

Koerber, A., Arnett, E. J., & Cumbie, T. (2008) Distortion and the politics of pain relief: A Habermasian analysis of medicine in the media. *Journal of Business and Technical Communication, 22*(3). 364-391.

Kraft, P. & Hunter, D. J. (2009). Genetic risk prediction: Are we there yet? *The New England Journal of Medicine, 360*(17), 1701 – 1703.

Krathwohl, D. R. (2009). *Methods of educational and social science research: The logic of methods* (3rd ed.). Long Grove, IL: Waveland Press, Inc.

Kruse, C. (2012). The Bayesian approach to forensic evidence: Evaluating, communicating, and distributing responsibility. *Social Studies of Science, 43*(5), 657-680.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kyvik, S. (2005). Popular science publishing and contributions to public discourse among university faculty. *Science Communication, 26*(3), 288-311.

Labinger, J. A. (1995). Science as culture: A view from the Petri dish. *Social Studies of Science, 25*(2), 285-306.

Lachenbruch, P. (2009). Communicating statistics and developing professionals. *Journal of the American Statistical Association, 104*(485), 1–4.

Ladner, M., & Lips, D. (2012). *Report card on American education: Ranking state K-12 performance, progress, and reform*. Washington, DC: American Legislative Exchange Council. Retrieved from https://www.alec.org/publication/2012-education-report-card/

Latour, B. & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Beverly Hills, CA: Sage Publications.

Latour, B. (1987) *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Latour, B. (1998). From the world of science to the world of research? *Science, 280*(5361) 208-209.

Layton, L. (2012, January 17). States should make better use of academic data, study says. *Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/states-should-make-better-use-of-academic-data-study-says/2012/01/17/gIQAb6cs6P_story.html?utm_term=.194e9b1e8e43

Leach, J. (2000). Rhetorical analysis. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image, and sound: A practical handbook* (pp. 207-226). Thousand Oaks, CA: Sage Publications Inc.

Lee, L. (2011). What did the teachers think? Teachers' responses to the use of value-added modeling as a tool for evaluating teacher effectiveness. *Journal of Urban Learning, Teaching, and Research, 7*, 97-103.

Leff, M. (1980). Interpretation and the art of the rhetorical critic. *Western Journal of Communication, 44*(4), 337-349.

Lerner, B. H. (2003). *The breast cancer wars*. New York: Oxford University Press.

Levin, N. (2014). Multivariate statistics and the enactment of metabolic complexity. *Social Studies of Science, 44*(4), 555-578.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106 – 131.

Longino, H. (1990). *Science as social knowledge. Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*(5), 304-305.

Lovett, I. (2010, November 9). Teacher's death exposes tensions in Los Angeles. *The New York Times*. Retrieved from http://www.nytimes.com/2010/11/10/education/10teacher.html

Lowrey, A. (2012, January 6). Big study links good teachers to lasting gain. *The New York Times*. Retrieved from http://www.nytimes.com/2012/01/06/education/big-study-links-good-teachers-to-lasting-gain.html

Lynch, S. N. (2008, July 22). Q&A: What can genetic tests tell you? *Time Magazine.* Retrieved from http://content.time.com/time/health/article/0,8599,1825679,00.html

MacFarlane, A. & Ewing, R. (Eds.). (2006). *Uncertainty underground: Yucca Mountain and the nation's high-level nuclear waste*. Cambridge, MA: MIT Press.

Mair, M., Greiffenhagen, C., & Sharrock, W. (2013). *Social studies of social science: A working bibliography.* Southampton, United Kingdom: National Center for Research Methods. Retrieved from http://eprints.ncrm.ac.uk/3219/

Majdik, Z. P. (2009). Judging direct-to-consumer genetics: Negotiating expertise and agency in public biotechnological practice. *Rhetoric & Public Affairs, 12*(4), 571-605.

Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine, 363*(2), 166 – 176.

Marrani, D. (2013). Nanotechnologies and novel foods in European law. *NanoEthics, 7*(3), 177 – 188.

Martin, B. & Richards, E. (1995). Scientific knowledge, controversy, and public decision-making. In S. Jasonoff, G. E. Markle, J. C. Petersen & T. Pinch (Eds.), *Handbook of Science and Technology Studies* (pp. 506 - 526). Newbury Park, CA: SAGE.

Maslin, M. & Austin, P. (2012). Uncertainty: Climate models at their limit? *Nature, 486*, 183-184.

Maynard, D. W. & Schaeffer, N. C. (2000). Toward a sociology of social scientific knowledge: Survey research and ethnomethodology's asymmetric alternates. *Social Studies of Science, 30*(3), 323-370.

Mayo, D. G. (1991). Sociological versus metascientific views of risk assessment. In D. G. Mayo & R. D. Hollander (Eds.), *Acceptable evidence: Science and values in risk assessment* (pp. 249-279). New York, NY: Oxford University Press.

McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica: RAND Corporation.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R. & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy 4*(4), 572-606. Retrieved from http://www.mitpressjournals.org/doi/pdf/10.1162/edfp.2009.4.4.572

McClure, K. (2009). Resurrecting the narrative paradigm: Identification and the case of young earth creationism. *Rhetoric Society Quarterly, 39*(2), 189-211.

McFadden, B. R. & Lusk, J. L. (2015). Cognitive biases in the assimilation of scientific information on global warming and genetically modified food. *Food Policy, 54*, 35–43.

McIntush, H. G. (2000). Defining education: The rhetorical enactment of ideology in A Nation at Risk. *Rhetoric & Public Affairs, 3*(3), 419-443.

McQuarrie, E. F. (2008). A visit to the rhetorician's workbench: Developing a toolkit for differentiating advertising style. In E. F. McQuarrie & B. J. Phillips (Eds.), *Go figure! New directions in advertising rhetoric* (pp. 257 - 276). Armonk, NY: M. E. Sharpe, Inc.

Mead, S. (2012). *Recent state action on teacher effectiveness*. Washington, DC: Bellwether Education Partners. Retrieved from http://bellwethereducation.org/sites/default/files/RSA-Teacher-Effectiveness.pdf

Medawar, P. B. (1964). Is the scientific paper a fraud? In D. Edge (Ed.), *Experiment: a series of scientific case histories first broadcast in the BBC Third Programme* (pp. 7-12). London: British Broadcasting Corporation.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation* (3rd ed.). San Francisco, CA: John Wiley & Sons, Inc.

Merton, R. K. (1973). The normative structure of science. In R. K. Merton (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 267-278). Chicago, IL: University of Chicago Press.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*(9), 9-20.

Messick, S. (1988). The once and future issue of validity: Assessing the meaning and consequence of measurement. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.

Messick, S. (1994). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning.* Princeton, NJ: Educational Testing Service.

Miller, C., Guston, D., Barben, D., Wetmore, J., Selin, C., & Fisher, E. (2007). *Nanotechnology and society: Ideas for education and public engagement* (Report No. R07-0001). Tempe, AZ: The Center for Nanotechnology in Society. Retrieved from https://cns.asu.edu/sites/default/files/library_files/lib_millercagustondb_0.pdf

Molnar, A. (2001, April). *The media and educational research: What we know vs. what the public hears.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, Washington. Retrieved from http://epsl.asu.edu/epru/documents/cerai-01-14.htm

Morgan, D. L. (2008). Emergent design. In L. M. Given (Ed.), The SAGE encyclopedia of qualitative research methods (pp. 246 - 248). Thousand Oaks, CA: Sage Publications Inc.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison, WI: University of Wisconsin Press.

Myers, G. (2003). Discourse studies of scientific popularization: Questioning the boundaries. *Discourse Studies, 5*(2), 265-279.

Nano Research Facility. (n.d.). *Societal and ethical implications of nanotechnology*. St. Louis, MO: Washington University in St. Louis. Retrieved from http://www.nano.wustl.edu/doc/Forms/NRF%20Nano%20Ethics.pdf

National Academy of Science & National Research Council. (1983). *Risk assessment in the federal government: Managing the process*. Washington, DC: National Academy Press.

National Center for Education Statistics (NCES). (2015). *Selected statistics from the public elementary and secondary education universe: School year 2013–14*. Retrieved from http://nces.ed.gov/pubs2015/2015151/tables/table_02.asp

National Council on Teacher Quality (NCTQ). (2010). *2009 State teacher policy yearbook: Florida.* Washington, DC: Author. Retrieved from http://www.nctq.org/dmsView/2009_State_Teacher_Policy_Yearbook_Florida_NCTQ_Report

National Council on Teacher Quality (NCTQ). (2013). *State of the states 2013 – Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Washington, DC: Author. Retrieved from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

National Council on Teacher Quality (NCTQ). (2015). *State of the states 2015 – Evaluating Teaching, Leading and Learning*. Washington, DC: Author. Retrieved from http://www.nctq.org/dmsStage/StateofStates2015

National Institutes of Health. (2016). Help me understand genetics: Genetic testing. Retrieved from https://ghr.nlm.nih.gov/primer/testing.pdf

National Research Council. (2009). *Letter report to the U.S. Department of Education on the Race to the Top fund*. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog/12780/letter-report-to-the-us-department-of-education-on-the-race-to-the-top-fund

National Research Council. (2010). *Getting value out of value-added*. Washington, DC: National Academies Press.

National Society of Genetic Counselors (2007). *Position statement: Direct to consumer genetic testing*. Retrieved from

https://web.archive.org/web/20120915004550/http://www.nsgc.org/Advocacy/PositionStatements/tabid/107/Default.aspx#DTC

Neethirajan, S. & Jayas, D. S. (2011). Nanotechnology for the food and bioprocessing industries. *Food and Bioprocess Technology, 4*(1), 39 – 47.

Neidhardt, F. (1993). The public as a communication system. *Public Understanding of Science 2*, 339-350.

Nichols, T. (2014, January 17). The death of expertise [Web log comment]. Retrieved from http://thefederalist.com/2014/01/17/the-death-of-expertise/

Niiniluoto, I. (1993). The aim and structure of applied research. *Erkenntnis, 38*, 1-21.

Nisbet, M. C. & Scheufele, D. A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany, 96*(10), p. 1767–1778.

O'Callaghan, T. (2010, June 14). FDA: At-home DNA tests should be regulated. *Time Magazine*. Retrieved from http://healthland.time.com/2010/06/14/fda-at-home-dna-tests-should-be-regulated/

O'Neil, M., & Kendall-Taylor, N. (2011). *Painted in a corner: How the media frame teachers' unions and education reform*. Washington, DC: FrameWorks Institute.

O'Neil, M., Kendall-Taylor, N., & Remington-Bell, T. (2012). *Cognitive media analysis on assessment: An education core story report*. Washington, DC: FrameWorks Institute.

Ogilvie, J. P. (2011, April 18). Do-it-yourself DNA testing: A risk or a right? *LA Times*. Retrieved from http://articles.latimes.com/2011/apr/18/health/la-he-dtc-gene-tests-20110418

Olson, L. C., Finnegan, C. A., & Hope, D. S. (Eds.). (2008). *Visual rhetoric: A reader in communication and American culture*. Thousand Oaks, CA: Sage.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: a necessary distinction? *Science Education, 95,* 627-638.

Otterman, S. (2010, December 26). Hurdles emerge in rising effort to rate teachers. *The New York Times*. Retrieved from http://www.nytimes.com/2010/12/27/nyregion/27teachers.html

Paroske, M. (2009). Deliberating international science policy controversies: Uncertainty and AIDS in South Africa. *Quarterly Journal of Speech, 95*(2), 148 – 170.

Pellechia, M. G. (1997). Trends in science coverage: A content analysis of three US newspapers. *Public Understanding of Science, 6*(1), 49-68.

Pew Research Center & American Association for the Advancement of Science. (2015). *Public and scientists' views on science and society*. Washington, DC: Author. Retrieved from http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Pew Research Center (2015). *Public and scientists' views on science and society.* Washington, DC: Author. Retrieved from http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Phi Delta Kappa International. (2015). The 47th Annual PDK/Gallup poll of the public's attitudes toward the public schools. Retrieved from http://pdkpoll2015.pdkintl.org/wp-content/uploads/2015/10/pdkpoll47_2015.pdf

Phi Delta Kappa International. (2016). The 48th Annual PDK/Gallup poll of the public's attitudes toward the public schools. Retrieved from http://pdkpoll2015.pdkintl.org/wp-content/uploads/2016/08/pdkpoll48_2016.pdf

Phillips, A. M. (2012, April 9). Release of city teachers' rankings prompts lawmakers to weigh limiting access. *The New York Times.* Retrieved from http://www.nytimes.com/2012/04/10/education/new-york-legislature-to-weigh-limiting-access-to-teacher-rankings.html

Phillips, D. C. (2006). Muddying the waters: The many purposes of educational inquiry. In C. F. Conrad & R. C. Serlin (Eds.), *The SAGE handbook for research in education: Encouraging ideas and enriching inquiry* (pp. 7-21). Thousand Oaks, CA: Sage Publications Inc.

Phillips, K. R. (2004). *Testing controversy: A rhetoric of educational reform.* Cresskill, NJ: Hampton Press Inc.

Pianta, R. (2012). Signal in a noisy system: Can VAM serve a purpose for evaluation and improvement of teaching? In Institute of Education Sciences (Ed.), *Learning from recent advances in measuring teacher effectiveness.* Washington, DC: Editor.

Pielke, R. A., Jr. (2007). *The honest broker: Making sense of science in policy and politics.* Cambridge, MA: Cambridge University Press.

Pilkey, O. H. & Pilkey-Jarvis, L. (2007). *Useless arithmetic: Why environmental scientists can't predict the future.* New York, NY: Columbia University Press.

Pilkey-Jarvis, L. & Pilkey, O. H. (2008). Useless arithmetic: Ten points to ponder when using mathematical models in environmental decision making. *Public Administration Review, 68*(3), 469 – 479.

Polikoff, M. S. & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis, 36*(4), 399-416.

Porter, E. (2015, March 24). Grading teachers by the test. *The New York Times.* Retrieved from https://www.nytimes.com/2015/03/25/business/economy/grading-teachers-by-the-test.html

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Powell, J. L. (2015). Climate scientists virtually unanimous: Anthropogenic global warming is true. *Bulletin of Science Technology Society, 35*(5-6), 121 – 124.

Proctor, R. N. (2008). Agnotology: A missing term to describe the cultural production of ignorance (and its study). In R.N. Proctor & L. Schiebinger (Eds.), *The making and unmaking of ignorance* (pp. 1-33). Stanford, CA: Stanford University Press.

Raudenbush, S. W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher, 44*(2), 138–141.

Ravitch, D. (2014). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Knopf.

Reform Support Network (2013). Educator evaluation communications toolkit: tools and resources to support states in communicating about educator evaluation systems. Retrieved from https://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/educator-evaluation-communications-toolkit.pdf

Reisigl, M. & Wodak, R. (2009). The Discourse-Historical Approach (DHA). In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 87 – 121). Thousand Oaks, CA: SAGE Publications Ltd.

Release of teacher test data in English and math is a big step forward. [Editorial]. (2012, February 25). *The New York Daily News*. Retrieved from http://www.nydailynews.com/opinion/release-teacher-test-data-english-math-big-step-article-1.1028281

Rich, M. (2012, September 11). National schools debate Is on display in Chicago. *The New York Times*. Retrieved from http://www.nytimes.com/2012/09/12/education/chicago-is-focus-of-national-debate-on-teacher-evaluation.html

Ripley, A. (2008, November 26). Rhee tackles classroom challenge. *Time Magazine*. Retrieved from http://content.time.com/time/subscriber/article/0,33009,1862444,00.html

Ripley, A. (2010, September 23). Waiting for "Superman": A call to action for our schools. *Time Magazine*. Retrieved from http://content.time.com/time/specials/packages/article/0,28804,2019663_2020590_2020592,00.html

Robson, M. E., Storm, C. D., Weitzel, J., Wollins, D. S., & Offit, K. (2010). American Society of Clinical Oncology policy statement update: Genetic and genomic testing for cancer susceptibility. *Journal of Clinical Oncology, 28*(5), 893–901.

Rose, M. (2010). Writing for the public. *College English, 72*(3), 284 – 292.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science, 5*(3), 127-134.

Rotherham, A. J. & Mitchel, A. L. (2014). *Genuine progress, greater challenges: A decade of teacher effectiveness reforms*. Sudbury, MA: Bellwether Education Partners. Retrieved from http://bellwethereducation.org/publication/genuine-progress-greater-challenges-decade-teacher-effectiveness-reforms

Rotherham, A. J. (2011, October 20). Blame game: Let's talk honestly about bad teachers. *Time Magazine*. Retrieved from http://ideas.time.com/2011/10/20/blame-game-lets-talk-honestly-about-bad-teachers/

Rothstein, J. (2011). *Review of "Learning about teaching: Initial findings from the Measures of Effective Teaching project."* Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/thinktank/review-learning-about-teaching

Russell, J. L. (2011). From child's garden to academic press: The role of shifting institutional logics in redefining kindergarten education. *American Educational Research Journal, 48*(2), 236-267.

Saltelli, A. & Funtowicz, S. (2014). When all models are wrong. *Issues in Science and Technology, 30*(2). Retrieved from http://issues.org/30-2/andrea/

Schochet, P. Z. & Hanley, S. C. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf

Schurman, R., & Munro, W. A. (2010). *Fighting for the future of food: Activists versus agribusiness in the struggle over biotechnology*. Minneapolis, MN: University of Minnesota Press.

Scott, L. M. (1994). Images in advertising: The need for a theory of visual rhetoric. *Journal of Consumer Research, 21*(2), 252-273.

Scott, L. M. (2008). Expanding rhetoric. In E. F. McQuarrie & B. J. Phillips (Eds.), *Go figure! New directions in advertising rhetoric* (pp. 297 - 314). Armonk, NY: M. E. Sharpe, Inc.

Selinger, E. & Crease, R. P. (2006a). Expertise and practical knowledge: Introduction. In E. Selinger & R. P. Crease (Eds.), *The philosophy of expertise* (pp. 193-195).

Selinger, E. & Crease, R. P. (2006b). Trusting experts: Introduction. In E. Selinger & R. P. Crease (Eds.), *The philosophy of expertise* (pp. 11-13).

Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New Directions for Program Evaluation, 60*, 13-57.

Shapin, S. & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton University Press.

Shapin, S. (1979). The politics of observation: Cerebral anatomy and social interests in the Edinburgh phrenology disputes. *Sociological Review, 27* [monograph], 139-178.

Shulman, L. S. (1997). Disciplines of inquiry in education: A new overview. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 3-29). Washington, DC: American Educational Research Association.

Siegrist, M., Cousin, M., Kastenholz, H., & Wiek, A. (2007). Public acceptance of nanotechnology foods and food packaging: The influence of affect and trust. *Appetite, 49*(2), 459 – 466.

Skirton, H., Goldsmith, L., Jackson, L., & O'Connor, A. (2012). Direct to consumer genetic testing: A systematic review of position statements, policies and recommendations. *Clinical Genetics, 82*(3), 210 – 218.

Song, J. (2011, March 23). U.S. education secretary calls for overhaul of No Child Left Behind. *The Los Angeles Times.* Retrieved from http://articles.latimes.com/2011/mar/23/local/la-me-0323-duncan-20110323

Song, J. & Felch, J. (2011, April 12). L.A. Unified releases school ratings using 'value-added' method. *The Los Angeles Times*. Retrieved from http://www.latimes.com/local/la-me-0413-value-add-20110414-story.html

Steudeman, M. J. (2014). "The guardian genius of democracy": The myth of the heroic teacher in Lyndon B. Johnson's education policy rhetoric, 1964–1966. *Rhetoric & Public Affairs, 17*(3), 477-510.

Stewart, B. E. (2006). *Value-added modeling: The challenge of measuring educational outcomes*. New York, NY: Carnegie Corporation.

Students First. (2013). *Elevating the teaching profession: Increasing teacher quality*. Sacramento, CA: Author. Retrieved from https://www.studentsfirst.org/page/-/publications/TeacherBriefUpdate.pdf

Tavory, I. & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*. Chicago, IL: University of Chicago Press.

Taylor, M., Kent, M. L., & Saffer, A. J. (2013). Using inoculation theory in public relations messages: Preparing the audience for arguments to be made later. In C. Liberman (Ed.), *Casing persuasive communication* (pp. 127 – 144). Dubuque, IA: Kendall Hunt Publishing.

te Kulve, H., Konrad, K., Palavicino, C. A., & Walhout, B. (2013). Context matters: Promises and concerns regarding nanotechnologies for water and food applications. *NanoEthics, 7*(1), 17 – 27.

Tierney, J. (2011, January 17). Heavy doses of DNA data, with few side effects. *The New York Times*. Retrieved from http://www.nytimes.com/2011/01/18/science/18tier.html

Turner, S. (2001). What is the problem with experts? *Social Studies of Science, 31*(1), 123 – 149.

Uhlmann, W. R. & Sharp, R. R. (2012). Genetic Testing Integration Panels (GTIPs): A novel approach for considering integration of direct-to-consumer and other new genetic tests into patient care. *Journal of Genetic Counseling, 21*(3), 374–381.

United States Department of Education. (2009). *Race to the top program*. Washington, DC: Author. Retrieved from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

United States Department of Education. (2009, April 30). *Partners in truth-telling: Secretary Arne Duncan's remarks to the Education Writers Association*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/partners-truth-telling

United States Department of Education. (2009, July 2). *Partners in reform: Remarks of Arne Duncan to the National Education Association*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/partners-reform

United States Department of Education. (2009, June 14). *States will lead the way toward reform: Secretary Arne Duncan's remarks at the 2009 Governors Education Symposium*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/states-will-lead-way-toward-reform

United States Department of Education. (2010, August 25). *Secretary Arne Duncan's remarks at the Statehouse Convention Center in Little Rock, Arkansas*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-statehouse-convention-center-little-rock-arkansas

United States Department of Education. (2011, March 22). *The road less traveled: Remarks of U.S. Secretary of Education Arne Duncan at the United Way of Greater Los Angeles Education Summit*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/road-less-traveled

United States Department of Education. (2012). *ESEA flexibility*. Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

United States Department of Education. (2012, February 7). *Fighting the wrong education battles: Remarks of Secretary Duncan at the Askwith Forum, Harvard Graduate School of Education*. Washington, DC: Author. Retrieved from https://www.ed.gov/news/speeches/fighting-wrong-education-battles

Vatz, R. E. (1973). The myth of the rhetorical situation. *Philosophy & Rhetoric, 6*(3), 154 – 161.

Vaughan, D. (2014). Theorizing: Analogy, cases, and comparative social organization. In R. A. Swedberg (Ed.), *Theorizing in social science* (pp. 61-84). Stanford, CA: Stanford University Press.

Walsh, L. (2010). Before Climategate: Visual strategies to integrate ethos across the "is/ought" divide in the IPCC's Climate Change 2007. *Issues in the Rhetoric of Science and Technology, 6*(2), 33-61.

Warriner, K., Reddy, S. M., Namvar, A., & Neethirajan, S. (2014). Developments in nanoparticles for use in biosensors to assess food safety and quality. *Trends in Food Science & Technology, 40*(2), 183-199.

Watanabe, T. (2010, November 14). D.C. schools may hold lesson for L.A. *The Los Angeles Times*. Retrieved from http://articles.latimes.com/2010/nov/14/news/la-teachers-dc-schools-lesson

Weathers, J. (2007). Privatizing schools: The struggle over how we define democracy and the role of public institutions. *Working Papers in Educational Linguistics, 2*(2), 67 – 93.

Weber, M. (2002). *The Protestant ethic and the spirit of capitalism*. (P. Baehr and G. C. Wells, Trans.). New York, NY: Penguin Books.

Weinberg, A. M. (1972). Science and trans-science. *Minerva, 10*(2), 209 – 222.

Williams, M. (2000). *Science and social science: An introduction*. London, UK: Routledge.

Willingham, D. T. (2012). *When can you trust the experts? How to tell good science from bad in education*. San Francisco, CA: Jossey-Bass.

Wilson, T. D., DePaulo, B. M., Mook, D. G., & Klaaren, K. J. (1993). Scientists' evaluations of research: The biasing effects of the importance of the topic. *Psychological Science, 4*(5), 322-325.

Winerip, M. (2011, November 27). Principals protest role of testing in evaluations. *The New York Times*. Retrieved from http://www.nytimes.com/2011/11/28/education/principals-protest-increased-use-of-test-scores-to-evaluate-educators.html

Winslow, L. (2015). The undeserving professor: Neoliberalism and the reinvention of higher education. *Rhetoric & Public Affairs, 18*(2), 201-245.

Wodak, R. & Meyer, M. (2009). Critical discourse analysis: History, agenda, theory and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 1 – 33). Thousand Oaks, CA: SAGE Publications Ltd.

Wynne, B. (1989). Sheep farming after Chernobyl: A case study in communicating scientific information. *Environment, 31*(2), 10-15, 33-39.

Wynne, B. (1992). Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science, 1*(3), p. 281-304.

Yanchar, S. C. & Williams, D. D. (2006). Reconsidering the compatibility thesis and eclecticism: Five proposed guidelines for method use. *Educational Researcher, 35*(9), 3-12.

Yettick, H. (2009). *The research that reaches the public: Who produces the educational research mentioned in the News Media?* Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from http://epicpolicy.org/publication/research-that-reaches

Yettick, H. (2015). One small droplet: News media coverage of peer-reviewed and university-based education research and academic expertise. *Educational Researcher, 44*(3), 173 – 184.

Young, A. M. (2014). *Prophets, gurus, & pundits: Rhetorical styles and public engagement.* Carbondale, IL: Southern Illinois University Press.

Zarefsky, D. (2008) Knowledge claims in rhetorical criticism. *Journal of Communication, 58*, 629-40.

Zompettie, J. P. (Ed.). (2013). *Essential readings on rhetoric*. New York, NY: International Debate Education Association.