Writing assessment's "debilitating inheritance": Behaviorism's dismissal of experience

Maja Joiwind Wilson

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation
Wilson, Maja Joiwind, "Writing assessment's "debilitating inheritance": Behaviorism's dismissal of experience" (2013). Doctoral Dissertations. 708.
https://scholars.unh.edu/dissertation/708

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
WRITING ASSESSMENT'S "DEBILITATING INHERITANCE": BEHAVIORISM'S DISMISSAL OF EXPERIENCE

BY

MAJA JOIWIND WILSON
BAS, Eastern Michigan University, 1997

DISSERTATION

Submitted to the University of New Hampshire in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in English

May, 2013
This dissertation has been examined and approved.

Thomas Newkirk
Dissertation Director, Thomas Newkirk, Professor of English

Crisi Beemer, Assistant Professor of English

Ann Diller, Professor of Education

Chris Gallagher, Professor of English, Northeastern University

Paula Salvio, Professor of Education

4/8/2013
Date
TABLE OF CONTENTS

DEDICATION.........................................................................................................................vii

ACKNOWLEDGMENTS.....................................................................................................viii

ABSTRACT.............................................................................................................................ix

CHAPTER PAGE

INTRODUCTION.....................................................................................................................1

I. CLUES TO THE DISMISSAL OF EXPERIENCE .......................................................13
   The MEAP dismisses Carl’s experiences and so he dismisses it ......................17
   A campaign against “experience, folk wisdom, and ideology” .............21
   Experience: explorations and definitions .....................................................22
   Standardization’s logic of replication .........................................................28
   Following the logic of replication to its own absurd conclusion .............32

II. JOHN B. WATSON’S BATTLE AGAINST MIND AND CONSCIOUSNESS:
    PARALYZING TO HUMAN EXPERIENCE .......................................................36
    Is a tool ever just a tool? ..............................................................................36
    Making sense of a muddle of terms ............................................................38
    Low educational overpasses .......................................................................41
    Watson frightens Little Albert, a “lowly piece of unformed protoplasm” ....45
    Titchener’s psychology: talk about mind ....................................................51
    Watson talks about consciousness: as *unscientific as soul* ......................56
    Paralyzing to human effort ..........................................................................63
III. “A DEBILITATING INHERITANCE: WATSON’S DISMISSAL OF MIND INFECTS EDUCATION”............................ 69

Watson’s model of behavior and the industrial model of production: \textit{wait, what's the difference}? ................................................................. 71

Watson’s model of behavior and Taylor’s model of efficiency: \textit{wait, what's the difference}? ........................................................................ 74

Watson’s model of behavior and Thorndike’s model of learning, teaching, and educational measurement: \textit{wait, what's the difference}? ........................................ 79

IV. EDUCATION’S NEW MASTER......................................................... 88

The standardization committee: democracy and efficiency at odds.................... 90

The pupil’s new master: the capitalist.............................................................. 92

Idiots, imbeciles, and morons: the psychological sorting machine.................... 96

“So much for observations”: four letters in which Lieutentant Furgeson objects to his impersonal rating................................................................. 99

“Great will be our good fortune”: in which wartime lessons in “human engineering” are applied to peacetime civil institutions............................ 101

Psychological testing: gears in the educational sorting machine....................... 103

Satisfying stakeholders with Thorndike’s rating scales.................................... 109

V. CARL IN THE BEHAVIORIST MATRIX: MEASURE, REWARD, AND PUNISH................................................................. 113

VI. INTROSPECTION AND EXPERIENCE GO UNDERCOVER AND SHOW UP TO THE COMPOSITION AND RHETORIC PARTY IN NEW PEDAGOGICAL CLOTHES................................................................. 125

“Reports of my death have been greatly exaggerated”: introspection lives........ 127

The writing process movement: “experience squared” .................................... 132

From introspection to Freud: layers of self..................................................... 138

VII. TESTING TOLLBOOTHS AND TEMPORARY OVERFLOW LOTS: COMPOSITION’S ROLE IN THE BEHAVIORIST INFRASTRUCTURE........... 142
DEDICATION

To the memory of my mother, Nancy Carol Rozell Wilson (1952 – 2012).
ACKNOWLEDGEMENTS

I am grateful for the love and support of my family, especially during this past year: my sons, Isaiah and Ben; my father, Ken Wilson; my brother Jesse; and my sisters Amy, Judy, and Grace.

I am also grateful for my committee -- Tom Newkirk, Chris Gallagher, Ann Diller, Paula Salvio, and Cristy Beemer. Years of conversations about assessment and education with Tom and Chris -- even before they were on my committee -- were invaluable.

Many others read or listened to drafts or portions of drafts and gave helpful responses along the way: my father, Phyllis Brazee, Nancy Patterson, Sierra Holmes, Penny Kittle, Patrick McFarlane, Mark Childs, Paul Thomas, Linda Rief, Alfie Kohn, Barry Lane, Jim Webber, Bob Broad and Dan Sharkovitz.

I am indebted to those Tom Newkirk refers to as “intellectual allies,” whose work provides a very real sense that I am not in this alone. There are too many to name but several deserve mention: Alfie Kohn, whose focus on the experience of the child resonates deeply; Peter Elbow, whose commitment to the experience of the writer never wavers; Rich Haswell, whose ear for poetry and contrarian nature are greatly comforting; Tom Newkirk, whose belief in narrative helped me avoid becoming the academic writer I didn’t want to be; and Bob Broad, with whom I share the love of a good metaphor and a deep distaste for holistic scoring calibration sessions.

I am grateful to the Yale University Archives for permission to quote from the collection of Robert Mearns Yerkes.
In this project, I examine the legacy of behaviorism’s dismissal of experience on contemporary writing assessment theory and practice within the field of composition studies. I use an archival study of John B. Watson’s letters to Robert Mearns Yerkes to establish behaviorism’s systematic denial of experience and its related constructs: mind, consciousness, thought, emotions, purpose, and meaning. I trace this denial through the efficiency movement’s effects on education and educational measurement in the early 20th century and the establishment of the behaviorist infrastructure of assessment -- an infrastructure that contributed, paradoxically, to the early focus in composition studies on experience. I analyze contemporary writing assessment’s principles and practices for remnants of behaviorism’s dismissal of experience. I conclude by proposing a new principle of principles for writing assessment based on the concept of experience.
Dear Readers,

This all began when I was ten years old. Doing my Sunday School homework, a four-color-clickey pen in one hand, a thin plastic ruler in the other. Zoom into the text on the ruler, and you’ll see the cause of angry tears on my face: *Two green underlines = the words of Jesus. A red asterisk = the words of the Pharisees. A single blue underline = a parable. A black circle = the words of the devil.*

I like to read the bible. My parents would be shocked, but the story of Jezebel in 2nd Kings fills my imagination, with the angry dogs and the ripping of flesh. But every time I have to do my homework and put this pen to these onion-skin thin pages of sticky silk, the words of God crumple up and rip underneath it.

My dad always picks up a bible and says affectionately, “Onion-skin thin pages of sticky silk.” It is a line he wrote as a first year college student when he was still an English major, before he became a nurse and then an evangelical preacher. I am proud of this line my father wrote before he was a grown man, because I am proud that he is a writer. He disappears into his office next to my room for hours, the loud clacking of his typewriter punctuating my play. One of his books has already been translated into Portuguese. But now, when I think of my father’s words, I cry harder. Because of these onion-skin thin pages, I will never finish my homework.
Later that week, my father will ask me how I like the new Sunday School program, this ruler and the clickey pen that the church has purchased from a Christian supply catalog. I will be on the verge of tears again. I will want to tell him how I hate the hours spent alone in my bedroom, how little the words and how thin the pages and how cramped my hand. But I will sense a danger in this answer: he will tell me it is a virtue to persevere.

Instead, I will tell him a true untruth.

After listening, my father will look at me thoughtfully, pull on his beard, and say, “I see what you mean.” The next week, my Sunday School teacher will tell us we have a new Sunday School program. I will throw away my ruler, but I will keep the pen.

What I will have told my father is this: All this marking stands between me and the words of the Lord, and I lose my sense of the story as I peer at the words through the ruler.

I have been saying the same thing, in different ways, all my life.

* * *

Years later, as a teacher in my own classroom, I will again be driven to angry tears by the commercial programs that stand between readers and their books, by the formulaic instruction that stands between writers and their meaning, by the rubrics that stand between teachers and their readings. I will know that my professional arguments against these programs are a reaction, in some way, to the systems of truth and
interpretation I struggled with as a child in my father’s church. What I won’t realize until just now is how my father’s response to me – *I see what you mean* – will give me the courage to keep speaking my mind even when it has proven professionally dangerous to do so.

**A METHOD TO MY MADNESS: TRICKS FOR MAKING ASSESSMENT**

**OBJECTS SPEAK**

I have been writing about writing assessment since 2004. I was a high school English teacher, and the rubric had fairly recently made its debut on the classroom scene. Almost immediately, it hit Platinum, was labeled a Best Practice, and before anyone could say, “Exceeds Expectations,” it had topped the charts as a Mandate. I was not a fan.

Whoever filled in the blanks, the rubric rarely reflected my experience of reading particular texts in ways that seemed satisfying to me or useful to my students. Quite apart from the question of how it worked to generate a number or grade, I resented how it mediated my readings and responses – and how it seemed to set up my students’ relationship with their intentions and their readers. No matter how I played with the rubric, it seemed incapable of doing what I wanted it to do -- communicating my experience of reading a text, filtered through my ever-changing sense of what a student writer needed to hear when.

I first posted my criticisms of rubrics informally on a professional listserv. An editor at Heinemann read them, asked me to write a book on the subject, and suddenly, I was in the middle of a profession-wide debate about writing assessment. During the course of this debate, played out in journals, conferences, online forums, and my own
district, I was often informed that the rubric wasn’t the problem. The rubric was just an object -- a tool, really -- that depended on the skill and intentions of its human users. By implication, the problem was the use to which it was put, or the person (mis)using it -- me.

I didn’t think I had a problem, except in the sense that it seemed a problem to expect a slab of marble to be as responsive to my fingers as clay. But people I respect continued to assert that, when created and used the right way, the rubric isn’t a bad thing (See Anson et al; Spandel; Thompson). This was assessment’s version of, “Guns don’t kill people, people do.” In ways I couldn’t quite articulate, this seemed both obviously true and obviously not true -- about guns and rubrics.

This dissertation is obviously not about guns, and not even obviously about rubrics. But, it is my attempt to explain how a seemingly neutral assessment object or tool, like a rubric, might carry intents and effects in ways that contradict the intentions of its users. It is also my inquiry into why, for all its insights and innovations, the field of writing assessment has not helped me to understand the questions and concerns I have brought to it -- particularly, questions and concerns about the experience of the readers and writers involved in assessment.

Although I was prepared to go where my questions took me, I am still surprised by the ground I have covered. Starting in an eighth grade alternative education classroom in Michigan in 2004 where a student named Carl angrily dismissed an exam question that he felt had dismissed him, I followed my inquiry through portions of three centuries and across settings I’d never before encountered: underneath one of Robert Moses’ shockingly low Long Island highway overpasses; through John Watson’s infant
laboratory; alongside one of Edward Thorndike’s vigorously struggling cats; and into a biology department meeting of the High School Teachers Association of New York where, in a single presentation sub-point, Mrs. Pinageay redefined the pupil as product, laborer, and raw material.

These digressions have helped me to understand why even contemporary writing assessment theory and its tools has not been able to account for the concept of experience in ways that satisfy me. Still, I feel some obligation to explain these digressions, since they add up to a rather eccentric dissertation. I am interested in stepping outside the obvious boundaries of the field in order to understand it differently. I do not aim, at least yet, to create and promote a new assessment technology or practice. Instead, I want this dissertation to allow different understandings about writing assessment.

Ultimately, I am very interested in the practical work of change. I have spent my fifteen-year career working to persuade, to build allies, and to change assessment practices, procedures, and policies in the schools I have worked in. Still, I have chosen to ignore the writing assessment establishment’s advice regarding change — advice that is often more like a warning than a suggestion. This advice asserts that speaking the language of the dominant educational measurement paradigm is the only way to be taken seriously, and thus, the only way to make any real difference (see Huot; Huot, et al; Adler-Kassner and O’Neill). While I respect the intentions of those who engage in this way, I find this stance both limited and limiting. I have seen this strategy used to curtail and censure our understandings more often than I have seen it change the system we claim to want to change. Most importantly, I am convinced that the perspective and
language of educational measurement gives me little insight or understanding into the
nature of the problem as I have experienced it.

In ways that will become clearer later, I have come to believe that the problems I
see in writing assessment are not technical problems, but social problems. I have found
Actor-Network Theory (ANT) helpful in explaining what I have attempted to do in this
dissertation. In Reassembling the Social, Bruno Latour begins his explanation of ANT by
counteracting a common understanding of "the social." In Latour's view, "the social" is
not a substance or factor that can be used to explain other things that are not social.
Latour would not say, for instance, "This economic theory breaks down in practice
because of 'social factors.'" Instead, "the social" is the process of people, ideas, and
objects -- what Latour calls human and non-human "actors" -- coming into association,
always moving and being moved by each other.

The work of the sociologist, then, is not to identify the substance of the social or
to identify ever more social factors, but to trace the associations that assemble actors --
people, things, and ideas -- together into networks. In a sense, this is what I have tried to
do in this dissertation: to begin tracing associations between human and non-human
actors through space and time as they have come together -- always moving and being
moved by each other -- to form the assemblage we call the field of writing assessment.

In order to trace these actors and their associations, Latour believes that the
sociologist must listen to what the actors are saying. He questions a hasty jump from the
actor's words to the sociologist's own explanations of them, explanations that often serve
to conclude rather than to extend the enquiry. He implores the enquirer to listen more
respectfully -- to "treasure" the actor's "odd way of speaking":

6
When a famous soprano says, ‘It is my voice who tells me when to stop and when to begin’, how quickly should the sociologist jump to the conclusion that the singer offers here a ‘typical case’ of ‘false consciousness’, artists being always too ready to take what is of their own making as the fetish that makes them do things? Is it not abundantly clear that this singer should not be listened to but instead ‘freed from her own delusion’ by the courageous exposition of her lies. Down with Muses and other undocumented aliens! And yet, the soprano did say that she shared her life with her voice that made her do certain things. Are we able to treasure this odd way of speaking or not? It was very precise, very revealing, very telling, and also very moving. Is not being moved, or rather, put into motion by the inform-ants exactly what we should mean by an enquiry (*Reassembling 48*)?

To “treasure [actors’] odd way of speaking,” the sociologist must listen without too quickly superimposing the limited concepts of her field. This can be a daunting enough task for those trained in traditional sociology’s reductions -- when the actors speak willingly. But what does it mean to listen to the actors when they don’t say anything -- when they remain silent?

Forcing objects to speak, Latour warns, is tricky but critical: objects become inaudible and invisible as soon as they are established in a network, making them difficult to trace and listen to. His discussion of the critical but quiet role of objects in networks has been useful to me in thinking about how and why the rubric interacted with my readings and my students’ intentions in ways that I didn’t like. An object’s silent and invisible function in a network helps me explain why so many people have protested that the rubric can’t be the problem simply because it is “just” a thing, “just” an object, “just” a tool. As Latour puts it, “...carrying their effects while becoming silent is what [objects] are so good at...” (79).

Latour proposes that “…specific tricks1 have to be invented to make [objects] talk, to offer descriptions of themselves, to produce scripts of what they are making others -- humans and non-humans -- do” (79). I find that I have used at least two of
Latour's tricks for making assessment objects speak. One trick involves looking at technological objects in the moment of their breakdowns. After the wonder of their creation -- our marvel at how we have brought objects together in ways that carry out our wishes -- these technologies fade into the background. We not only forget how they were made, but *that* they were made: we forget that our own assumptions, beliefs, intentions, and controversies are embedded in their materials and structures. And we forget how they are making us.

These new objects become part of the landscape of our lives -- when they work properly. But when these technologies fall apart or stop carrying out our wishes, we remember once again that and how they were assembled, as the flurry of activity and interaction that long ago brought them into association with other actors can again come into focus. In this way, otherwise invisible and silent objects become again visible and audible in the moments of their breakdowns. Latour uses the example of the Columbia shuttle disaster. One moment, the shuttle was a silent object -- "autonomous, automatic, completely devoid of human agents" ("Networks" 797). The next moment, as pieces of it rained down on Texas, observers could see,

swarms of entities that...appear in retrospect necessary for its sustenance...the Columbia shuttle was an object ready to fly in the sky, and then suddenly...you realize that it needed NASA and its complex organizational body to fly safely in the sky -- here is the hall where the disjointed parts have been assembled for the task force to inquire into what went wrong. ("Networks" 797)

In the shuttle's unmaking, we saw the myriad of actors -- materials, people, needs, wishes, ideas, knowledge, beliefs, assumptions, compromises, and negotiations -- that had been brought together in its making.
Making an assessment object speak in the moment of its breakdown is my opening move in this dissertation. I begin Chapter One with the story of a teacher named Amy, and an eighth grade alternative education student named Carl, brought together and then thrown apart by the 2004 Michigan Educational Assessment Program (MEAP) writing exam. This move is a slight variation on Latour’s: in my story, it is not the assessment tool itself that breaks down -- it is actually doing what it was designed to do. Instead, as the tool comes into association with Amy and Carl, three breakdowns take place. Carl breaks down, Amy’s teaching breaks down, and their relationship with each other breaks down. In these three breakdowns, the object begins to give voice to the beliefs, assumptions, and human purposes that brought it together as an assessment tool.

I find the 2004 MEAP writing exam as eloquent in this moment as if it had literally shattered in front of Amy and Carl like the Columbia exploding above a horrified crowd. To understand how and why this large-scale assessment tool affected these rather dramatic breakdowns, I use a second trick recommended by Latour for making objects speak: objects can be illuminated in the archives, as we see them again in the moment of their making. In Chapters One and Two, I describe going to the archives of Robert Yerkes, the father of standardized testing. In Yerkes’ collection, I find moments in which the concerns, tools, and procedures of large scale standardized testing were first emerging in relationship with ideas -- behaviorism, industrial efficiency, and “human engineering,” -- and people -- John B. Watson, Edward Titchner, William MacDougall, Edward Thorndike, and John Dewey.

In Chapter Three, I find another small moment of breakdown in Yerkes’ papers: letters of protest written to Yerkes during World War I from Lieutentant George O.
In these letters, Lt. Furgeson claims that Yerkes’ rating system has failed, impeding his work and forcing him to resign. As I trace associations between these early actors forward again to Carl and Amy in Chapters Four and Five, and, eventually, to the field of composition studies and writing assessment in Chapter Six, several new actors come into focus, including the New York City biology teachers as they translated the work of efficiency engineers into education.

In Chapter Seven, I build on the history uncovered in the previous six chapters to describe the concept of an infrastructure: how it facilitates our movements, how it suggests and constrains the technologies we use to navigate and maintain it, how it affects the values we want to enact in our lives, and how it structures our experiences. I argue that the behaviorist infrastructure of assessment actually gave rise to composition studies, and that anomalous subfield within it, writing assessment. In chapter Eight, I tell the story of two of writing assessment’s early resistances to the very field of educational measurement from which it emerged: holistic scoring and portfolio assessment.

In Chapter Nine, I tell the story of my first experience with Brian Huot’s five preliminary principles for writing assessment -- how I used these them to resist a centralized mandated assessment scheme, and how I consequently ran up against what I consider to be their limitations: in their focus on “context’ and the “local,” Huot’s principles of writing assessment work well as resistance strategies, but never fully account for the concept of experience. I will examine Bob Broad’s Dynamic Criterion Mapping in a similar light: while it evades or puts off quantification, it stops short of making experience a central construct, leaving it vulnerable to being drawn back into the service of the behaviorist infrastructure. In Chapter Ten, I will lay out some of the
possibilities for assessment that might come from making “experience” one of writing assessment’s key conceptual constructs – a “principle of principles.”

The work of tracing associations between actors in this network and compelling them to speak does, as Latour predicts, proceed slowly. But I hope that the reader will find these actors, their scripts, and the associations between them as fascinating as I have. And I hope that the understandings that emerge from this work will help us have new conversations about assessment -- conversations that will allow us, in the words of William James (paraphrasing Dewey and other pragmatists) “to get into more satisfactory relation with other parts of our experience” (100).

While Latour’s theoretical framework helps explain the shape of my work, it does not explain (nor did it provide) my motivation to do it. On the day I decided to write my dissertation on assessment and not on a study I was planning to conduct of new teachers, Richard Haswell sent me his then unpublished review of A Guide to College Writing Assessment by Peggy O’Neill, Cindy Moore, and Brian Huot. While praising the authors’ impeccable scholarship and admirable refusal to write people out of the assessment picture, Haswell critiques the book as a rhetorical act. First, he describes Scott Robert Olson’s explanation of how certain Hollywood movies play well outside the culture that produced them. By opening a space for all audiences to project their own beliefs and rituals onto the culturally specific beliefs and rituals represented in the movies, Olson found these cultural cross-dressers share five transcendent mythotypes: awe, wonder, purpose, joy, and participation. Haswell then looks for these mythotypes in A Guide, asking how it will play outside the scholarly community that produced it. Where in A Guide, he wonders, is the sense of awe, wonder, purpose, joy, and
participation that “hold at bay the equally universal fear that the world is vast, material, meaningless, and in ultimate control”? Haswell concludes his review of *A Guide* with this assessment:

“This is the book -- with its full and cutting-edge coverage of the *needs-to* -- that I will recommend when I learn that a colleague is ready to take up assessment. But for the book that creates the *ready-to*, that opens up the culture of writing assessment to other cultures within our field, that creates openings for the human yearning to challenge the gods of testing and to do good deeds -- in short the book on the adventure of writing assessment -- that book is yet to be written (117).

I do not claim to have written such a book. But Haswell’s words have helped me to hold at bay my own fear that the world is vast, material, meaningless, and in ultimate control. What follows is my attempt to give voice to the sense of purpose -- and, yes, joy -- I have experienced while talking back to the gods of testing. I don’t expect that everyone will throw out her ruler. But I’m hoping at least a few people will say, “I see what you mean.”

Maja Wilson
CHAPTER I

CLUES TO THE DISMISSAL OF EXPERIENCE

“i’m not going to write about friends...”
--Carl, eighth grader

I found my first professional home in the high school English department where I taught literature and writing to college prep students for five years and to alternative education students for two years. Although I hadn’t read enough professional literature when I joined that department to know it, it had been shaped in profound ways by composition studies. My department head had studied with Stephen and Judy Tchudi, creating a series of writing courses that paid homage to the writing process rather than the five-paragraph essay. A colleague kept a lending library of professional books -- many of them written by scholars I would encounter in my graduate program. Another colleague had spent formative summers at Breadloaf with Kenneth Marcorie and James Britton. “Those were the days,” he’d hold forth, whenever I asked him about it, “hitting the volleyball back and forth with Ken and Jimmy.” A beloved fixture in a conservative school community, he swore in class and wrote inventive poetry with and for his students. In this school, I was free to experiment with my writing and my teaching to my heart’s content. These were my people!

Even when I found something to resist in the second half of my secondary teaching career -- the increasingly intrusive assessment mandates of NCLB -- I experienced a sense of belonging while reading the work of the writing assessment scholars within composition studies to whom I had turned for support. In these scholars, I recognized the spirit of resistance. Here was Edward White, resisting the ultimate
reduction of the multiple-choice grammar test with a celebration of the holistic integrity of the text! Here were Peter Elbow and Pat Belanof, confounding the scientistic hubris of the one-shot standardized writing exam with multiple samples of student writing gathered in naturalistic settings! Here was Brian Huot, wielding the concept of “local assessment” to fend off imposed national assessment brands! Here was Bob Broad, exploding the traditional rubric’s monopoly on textual factors by uncovering no less than 89 values within a single group of readers! And, here was Patricia Lynne, rejecting the psychometrician’s terms in one fell swoop. Vive le Resistance! These were my people!

However, in ways I will explain later, I misread much of this writing assessment literature in order to support my developing positions. These misreadings weren’t completely intentional. I was trying to understand and join a complex academic conversation on my own. In addition to parenting two young sons, I taught the community’s neediest teenagers five days a week: twenty minutes of professional reading was all I could sneak in a night before losing consciousness. All would be well as I’d read and re-read the author’s opening critique, gleeful witness to the skewering of a common enemy. Edward White didn’t like assessments that fractured the text or the act of writing? Me, neither! Bob Broad thought that traditional rubrics were too reductive? Me, too! Patricia Lynne thought reliability a liability? Right on, sister!

The trouble would begin as I’d read what followed the critique, usually the author’s proposal for change. Sometimes, I just didn’t have time to read this second part carefully. More often, I didn’t fully understand it. Without a community of readers by which to check my understandings, I felt free to take what I wanted, ignore what I couldn’t comprehend, and even misunderstand in a way that served my own purposes.
After teaching for ten years in Michigan’s public schools, I joined the doctoral program in composition studies at the University of New Hampshire. In my new role as graduate student, I re-read the work on assessment I’d encountered as a high school teacher, and more. With the luxury of time and the benefit of mentors, I now understood that these proposals for change often didn’t follow the momentum I’d projected from the critiques that had preceded them.

I have resisted this realization mightily. I have struggled to look clearly at my misreadings because they underscore differences with writing assessment scholars from whom I have taken so much comfort. Those who know my contrarian nature may be surprised to learn that I didn’t want to write this critique of the field of writing assessment. But, the more I understand, the more distressed I find myself by how the field is emerging in relationship with its material and theoretical contexts.

It will take ten chapters to develop my critique, but I’ll tip my hand now. I have come to understand the story of writing assessment within composition studies as a series of resistances to various aspects of behaviorism expressed through educational measurement theory. I ally myself with the spirit of these resistances, and I am grateful for them. However, I believe that these resistances have rarely confronted the behaviorist infrastructure of decision-making and social control in which educational measurement emerged. Without fully understanding the nature of this infrastructure -- the values embedded in it and then perpetuated through the technologies that support it -- our resistances don’t go far enough.

Without challenging or changing the infrastructure itself, we allow our work to be defined for us. Any truly innovative technologies that we would design are ignored.
because they do not work within the existing infrastructure, or changed so that they do.
We accept these limits, often internalizing and becoming advocates for them, because we
want to be relevant. Hostage to and often complicit in the very behaviorist system that
goes against our best instincts, we perpetrate the indignities foisted on us back onto
ourselves, silencing our own more radical proposals for change (or the understandings
that might lead to them) with the charge that they would be “impractical.” In the process,
we stunt our theoretical imaginations.

Others have told the story of writing assessment differently. For example, in
“Historicizing Writing Assessment,” Kathleen Blake Yancey focuses on how technical
concepts from educational measurement theory have shaped the field. In Yancey’s
telling, writing assessment’s history can be characterized by three waves of shifting
emphases between the concepts of reliability and validity, each concept surfacing on the
crest of a wave before being drawn back into the undertow, returning to shore in a
different layer of emphasis. I have found much to admire and use in Yancey’s story. But
the forces at work in my story are -- pointedly -- more personal and social than the
movement of waves.

I believe that the story of writing assessment exists within a larger story: the
story of a society that has abdicated its obligation to grapple with the very questions that
a democracy allows (and demands) it keep open. These are questions about power and
relationships, exclusion and acceptance, rejection and belonging. Who belongs to which
groups? Who decides the boundaries of these groups? Which groups receive what
resources? Who has the power to make such judgments, and on what basis? How do we
justify them? How do we decide how we decide? What are the consequences of these
decisions -- and the methods we have chosen for making them -- in our individual and collective experiences? I believe that we have displaced the moral weight of these questions onto the technical apparatus of assessment and the behaviorist infrastructure that called it into being, obscuring its role in shaping our view of what assessment can do and be.

Our participation in this behaviorist infrastructure has blinded us to the potential for the concept of experience to transform our theory and practice of writing assessment for teaching and learning. This blindness is no surprise; it is a consequence of the infrastructure’s originating values. Experience, in fact, is the very concept that behaviorism purposefully denies, allowing if not encouraging industrial models to run roughshod over human experience in the pursuit of profit. Composition scholars have heroically attempted to loosen behaviorism’s hold on assessment by putting several concepts at the center of writing assessment theory and practice: validity and context, to name two. But without attention to the concept of experience -- which cannot survive the behaviorist infrastructure that currently sponsors most of the work that we attempt to do in writing assessment -- our well-intentioned resistances risk collapsing in on themselves, doing damage, in the end, to the very thing that they might allow us to protect, understand, and develop.

THE MEAP DISMISSES CARL’S EXPERIENCES AND SO HE DISMISSES IT

To understand what I mean by the concept of experience and writing assessment’s blindness to it, you’ll have to follow me back to 2004, to a warehouse on the side of a highway in Michigan. It was here in this warehouse (otherwise known as an alternative
middle and high school) that an eighth grader named Carl and a teacher named Amy helped me think more deeply about assessment and experience. I take a risk by inviting readers who are primarily concerned with college writers to follow me back into an eighth grade classroom -- especially when they find we will linger in the world of K-12 long past a single illustrative anecdote. Surely, this diversion should benefit college teachers for all the usual reasons: we all teach the same students and the same pressures for accountability that shaped my experience of K-12 are quickly exerting themselves on postsecondary institutions.

But I also believe that the alternative education students I worked with are the proverbial frogs in the polluted river: they suffer more quickly and visibly from the educational toxins that should worry us all. Compositionists and writing assessment theorists will not have met my alternative education students. Few of them ever made it to a university -- Carl himself didn’t live past tenth grade. To clearly see the effects of the toxins that I’m claiming are even now polluting writing assessment theory, teachers of post-secondary students and writing assessment theorists are going to have to spend some time mucking around with me in the K-12 river.

The two years I spent at this alternative school on the side of a highway in Michigan were the most challenging years of my career. Add adolescence to difficult lives and complicated school histories, pour it all into a badly lit building with the thinnest of hastily constructed internal walls and no counseling or special education services, and you’ve got a recipe for challenging work. I’d already taught for three years in an adult basic education program and for two years at a traditional high school. While I’d always known that relationships in the classroom were important, I found that I
couldn’t survive at the alternative school (let alone teach anything) unless I simultaneously got to know students and redesigned the curriculum so that students could create relationships between their lives and what happened in class. More than any teacher I’d met, my colleague Amy had always known the power of relationships in her teaching.

Unfortunately, the relationships Amy and I worked so hard to establish with students and the curriculum no longer mattered during MEAP week. For one week in October, students across the state put aside projects, papers, lectures, and labs for the Michigan Educational Assessment Program (MEAP). Students were commanded to sit still, be quiet, and fill in the multiple-choice bubbles or write essays in response to canned writing prompts such as “Justice” or “Community Service.” And teachers were reduced to reading scripts from the MEAP manual.

With a history of failure on the MEAP, I was frankly surprised that our students bothered to show up during MEAP week at all. But the stakes were high. A school’s failure to meet Adequate Yearly Progress (AYP) several years in a row (as measured, in large part, by the MEAP) resulted in sanctions under the No Child Left Behind (NCLB) Act of 2001, including dismissal of teachers and administrators, and takeover by the state or a private management company.

During lunchtime in the middle of MEAP week, Amy handed me a crumpled up piece of paper. Earlier that morning, she had given her eighth grade students the writing MEAP. Amy was required to read verbatim from a series of scripts that directed students when to turn test pages or sharpen their pencils or go to the bathroom. The MEAP
manual forbade her from discussing the prompts beyond repeating the information in the script, lest she invalidate the tests.

When she’d broken the seal on the MEAP writing exam, Amy scanned ahead to the theme of the prompt: “Friends.” She winced, knowing that her students had difficult relationships; one had been suspended from her class earlier that year for pulling a knife on his best friend, whom he’d accused of stealing his “girl.” But there was nothing she could do about the prompt or her students’ reaction to it. Her job, according to the MEAP, was to read the directions and the prompt out loud, clearly and slowly. Nothing more, nothing less.

When she introduced the prompt, she saw Carl’s face redden before he blurted out, “I’m not doing this!” After Amy finished reading the directions, she sat down next to Carl to talk, violating at least one MEAP code and probably several international statutes in the process. Carl asserted that he had nothing to write about because he didn’t have any friends. Amy encouraged him to write about this; any essay was better than no essay. Carl wrote furiously on a scrap paper for about thirty seconds, balled it up, threw it on the ground, and stormed out of the building. Amy had saved the scrap to show me at lunchtime. I smoothed out the wrinkles and read,

‘i’m not going to write about friends because i don’t have any friends
and i had a friend once and he tried to kill me the end.’

Amy was furious -- at the test, not at Carl. “I would know better than to ask my students to write about friends,” she fumed, “at least not without a group of therapists present.” If she ever wanted to invite her students to write about their friends, she said, she would do so only after careful discussion, and only after she knew there was enough trust between
herself and her students to handle the fallout. But none of this -- not her experience as a
teacher, and not her knowledge of her students’ experiences -- mattered during MEAP
week.

A CAMPAIGN AGAINST “EXPERIENCE, FOLK WISDOM, AND IDEOLOGY”

I’d opposed large scale standardized testing for years, but always on the grounds
that the tests intimidated students, misrepresented their abilities, narrowed the
curriculum, and encouraged drill and kill teaching (see Kohn, Ohanian). But Amy’s story
about Carl brought another troubling aspect of the testing situation into focus for me.
Even though the writing prompt solicited students’ experience of friendship, the larger
structure of the test seemed designed to dismiss the experiences of everyone involved.
No one who knew or cared about Carl’s experiences would have asked him to write about
friends in such a setting. And no one who valued Amy’s teaching experience or her
knowledge of Carl’s experiences would have required that she distance herself from what
she knew for the sake of objectivity for one moment, let alone for an entire week.

But the federal Department of Education (DOE) didn’t just want Amy to distance
herself from her experience for the duration of MEAP week. The DOE wanted to change
educators’ everyday relationship with experience. Three years before, the Institution of
Education Sciences (IES), founded and funded by the DOE to carry out the mandates of
NCLB, had laid out its “ambitious agenda” to replace educators’ reliance on “personal
experience, folk wisdom, and ideology”:

A number of recent reports have described current education practice as not resting
on a solid research base. Instead, policy and practice decisions are often guided by
personal experience, folk wisdom, and ideology. Grounding education policy and
practice in the United States on evidence will require transformation of both the
research and practice fields. Practitioners will have to turn routinely to education research when making important decisions... To achieve this ambitious agenda, there is a need for a cadre of well-trained scientists capable of conducting high quality research that is relevant to practitioners and policy makers. (Institution of Education Sciences)

My first instinct was to laugh. I imagined white-coated researchers hovering behind Amy in her eighth grade classroom, clipboards and stopwatches in hand. Who doesn't want a scientist or two over her shoulder? When Carl crumples up his paper and storms out of the building, Amy cocks an eyebrow and calls out, “So, what’re we gonna do, guys?” Members of our well-trained cadre exchange blank looks before a collective shrug in Amy’s direction. “We’ve never seen that in the laboratory.”

EXPERIENCE: EXPLORATIONS AND DEFINITIONS

Still, my laughter belied grief and anger. That experience would fall victim to this “ambitious agenda” felt deeply, profoundly wrong to me. To be fair, I was probably thinking of experience more broadly than IES was using it in this particular passage. The IES was careful to vilify teachers’ “personal” experience, as opposed, perhaps, to their “professional” experience. But, when it came to teaching, I didn’t see a bright line dividing the two. In “The Overcoming of Experience,” Margaret Buchmann claims that while the counterintuitive knowledge in some fields -- such as medicine or physics -- require professionals to deny their personal experience, the knowledge in other fields -- such as teaching -- are drawn directly from personal experience. She quotes philosopher Renford Bambrough, who refers to both writing and teaching as human birthrights:

Thinking, teaching and learning are human birthrights: By birthright we are all not only thinkers but also singers and dancers, poets and painters, teachers and storytellers. This means that the professional singer or painter, poet or teacher, dancer or story-teller, is a professional in a different way from the solicitor or doctor,
Unlike the doctor or statistician, who must be trained to think, act, and see in ways that do not derive from common personal experience, the teacher, storyteller, and writer can call upon experience when doing, “in a more high powered way” (192), what we all already do. But, as Buchmann points out, to become professionals in professions that cannot draw on common personal experience, highly ritualized “breaks from experience” are built into training -- periods of time such as the doctor’s residency when everything an apprentice might have known is turned upside down, when even the normal experience of sleep is disrupted to signal to the trainee that she is entering a profession that requires a “new way of experiencing, sensing, and communicating” (184).ii

By insisting on a “transformation” (Institute…) of the field of education by demanding that educators break from personal experience and turn instead to scientific educational research, the IES implied that education was no different than fields such as medicine: teachers, like doctors, should put personal experience behind them in order to take on the sometimes counterintuitive insights generated by science. In fact, in the first working group conference of the IES in 2002, Susan Neuman, the Assistant Secretary for Elementary and Secondary Education, proclaimed that, “…when we teach students we really are engaging in a kind of brain surgery.” The prevalence of medical images in educational discourse renders this statement unremarkable. Even progressive educators such as Harvey Daniels have frequently invoke the medical metaphor, using the term “best practice” to align education with the legal and medical professions -- and thus, to
extend to teachers the cultural respect reserved for doctors and lawyers (See Daniels, 
Hyde, and Zemelman).

But the effect of the IES and the federal DOE's invocation of the medical model was to vilify teachers, so naïve as to rely on "folk wisdom." So while Daniels' aim in invoking the medical model is to lend educators respect by association, thereby increasing their authority and autonomy, the effect (if not the direct aim) of the IES and the DOE's invocation of the medical model was to undermine educators' authority. And once you undermine the collective authority of a large number of individuals (teachers who rely on dangerous experience, ideology and folk-wisdom!) then you set up the need for control by a small group. ("Cadre," is, by definition, a small, specially trained group). The research produced by this new cadre of researchers would control the decisions of educators. This cadre, in turn, would itself be controlled by the group that had called for its formation and training in the first place: the federal DOE. Amy and I, it seemed to me, were pawns in a chess game that had little to do with Carl or any of the students in our classrooms. The winner of the game would control public education.

But beyond the question of who controlled public education, I worried that the IES' vilification of personal experience worked to shut down the important reflective process that Bambrough and Buchmann implied was key to teachers' development and professionalism. If teaching is a profession in which teachers do, "in a more high powered way," what each of us already does by virtue of being human, then personal experience and reflection are key tools by which teachers become professionals. Consider the way that Amy drew on her personal experience in the classroom. To solve one of the problems frequently encountered in her profession -- middle school readers
who don’t want to read -- Amy eschewed the traditional advice to increase the punishment for noncompliance and the reward for compliance. Instead, she got to know her students’ interests, drew on her wide reading experiences, and helped every student find a book she or he connected with. Drawn in by the enjoyment of reading the right book and the connection with a teacher who took the time to get to know them, these resistant readers began to read more, seek out more books, and read with more confidence and skill.

What informed Amy’s decision about how to approach resistant middle school readers? She didn’t rely on educational research -- though it might be possible to find research that supported her decision. She didn’t rely on the curriculum at the school where she taught, a curriculum that would attempt to solve the problem of resistant readers with more frequent and heavily weighted reading quizzes. Instead, Amy’s approach was informed by her beliefs about the role of relationships in education, beliefs with strong experiential roots in her own formative experiences and relationships. She had experienced this relational approach practiced by her own father, who had modeled it in school as a guidance counselor and at home as a parent. Amy had then cultivated this approach through years of reflection, study, and practice.

There seemed little separation, then, between Amy’s personal experience and her professional knowledge and skills. She had built an intentional connection between them. If Amy, Buchmann, and Bambrough were right about the relationship between teaching and personal experience, then the IES’ call for teachers to break from their personal experience meant that teachers’ development might be stunted. Requiring Amy to break from her personal experience in her role as a teacher would not only be
unnecessary, but also damaging: to Amy and to her students. Indeed, asking Amy to break from her experience by replacing it with the MEAP testing situation and writing prompt had already been deeply damaging to Carl.

When I use the term *experience* in this dissertation -- when I accuse a theory or practice of dismissing experience, or when I call for more attention to it -- I purposefully include a range of meanings. I mean what we commonly mean by “professional” experience: the number of years a professional has practiced, and the breadth and depth of professional problems she has encountered and worked through. I also mean something close to John Dewey’s inclusive definition of the term: an organism’s transaction with its environment. This definition includes the human organism’s bodily sensations as it bumps up into the world around it, but also the thoughts, beliefs, and emotions by which a person *makes sense* of those sensations.

These meanings -- the *sense* a person makes of her experiences in the moment and through later reflection -- transform the experience itself. This definition sets up what will later prove to be a crucial difference: while John B. Watson, the Father of Behaviorism, would consider an organism’s sensations as it bumps into the world around it, he would not join me in considering the *sense* a person makes of these sensations based on her thoughts, emotions, wishes, and previous experiences. Experience, as I use it, includes a person’s consciousness and self-consciousness in addition to her physical sensations and reflexes. As we will discover, Watson dismissed the notion of thoughts, emotions, consciousness, and experience as valid constructs at all.

As Dewey reminded us in *Experience and Education*, the concept of experience alone does not tell teachers what to do: students are always having experiences. Students
experience a lecture, and they experience a hands-on activity. The question for Dewey
was not whether or not a teaching method or activity gave a student an experience; the
question was what kind of experience the student was having, and how that experience
affected his learning in the moment and in the future. According to Dewey, educative
experiences leave the student open, receptive, and eager for future educative experiences.
Miseducative experiences, on the other hand, make him “callous” to future learning. In
Dewey’s terms, then, Carl’s experience of the MEAP writing exam was not just negative,
but also miseducative -- with the effect of making future writing less likely. The
educator’s job, then, is to set up environments and activities that provide educative
experiences and then help him reflect on (make sense of) these experiences. And to avoid
setting up miseducative experiences.

I am asking us to consider assessment as a learning experience. Are the
assessments we set up educative? Do they open up the student to further writing? Do
they open up a student’s understanding of what happens when another human being reads
and experiences -- makes sense of -- the words she has written? It probably doesn’t take
much to convince you that the MEAP was miseducative for Carl, shutting him down to
future educative writing experiences. Not only did this experience make him less likely to
want to write again, but it also provided a miseducative experience of audience.
Students’ relationships with their audiences are crucial for their development as writers.
The MEAP writing exam caused Carl to sever his relationship with his audience. Why
would he work hard to communicate something that was difficult to express in the first
place, knowing, as he must have, that his audience’s purpose in reading his text would be
to judge it rather than to understand it or to try to understand him through it?
It will likely be more difficult to convince you that contemporary writing assessment practices may be similarly miseducative: that they fail to open up a student’s understanding of what happens when another human being experiences — makes sense of — the words she has written. To convince you of this, we’ll need to muck around in K-12 a bit longer, to see how deep and wide the dismissal of experience goes. And then we’ll need to trace it back. We won’t go back to Plato or attempt to identify its “source.” But if we follow the dismissal of experience from this moment between Carl, Amy, and the MEAP back to the point from which it branches forward again into a variety of institutions, movements, and schools of thought, we’ll have a better chance of seeing how it made its way into contemporary writing assessment theory and practice.

**STANDARDIZATION’S LOGIC OF REPLICATION**

The dismissal of experience I was seeing in the tools of the accountability movement seemed to go hand in hand with standardization: the MEAP’s standardization of the prompt and testing situation dismissed Carl’s experiences and the IES’ standardization of teacher decisions dismissed Amy’s experiences. I’d long recognized standardization’s link to objectivity, but with *experience* as my lens, I saw objectivity more clearly as an attempt to transcend personal experience in a way that ended up devaluing it. If personal experience is viewed as a problem, standardization is an attempt to eradicate it: by giving everyone the same conditions, the logic goes, the experiences are no longer personal in the sense of being unique. Of course, standardization as a solution to the “problem” of personal experience doesn’t work on its own terms, because it is based on a misunderstanding of experience: standardizing the testing situation does
not result in uniform experiences for everyone who experiences it, as illustrated by Carl’s
outburst, which had been shaped, in part, by his particular past experiences of friendship.
But, while standardization doesn’t actually standardize experience, it gives the illusion
that personal experience has been transcended and that objectivity has been achieved.
The circular compound consequence of this illusion of standardization is to further deny
the role of personal experience. And when you deny the importance of experience, it
becomes easier to justify abuse — the ultimate miseducative experience.

It seemed bad enough to me that teachers and students’ experiences were
dismissed during MEAP week through the standardization of the testing situation and the
test prompts. But once introduced into the system, standardization replicated itself like a
highly contagious virus, compounding the original dismissal of experience. The MEAP
was only administered in Michigan at the time during 3rd, 5th, 8th, and 11th grades.
However, as soon as it became attached to the funding and sanctions instituted by NCLB,
districts began “benchmarking” the standards supposedly measured by the MEAP, and
testing students on these benchmarks several times yearly. Most often, these benchmark
assessments were mini-versions of the MEAP -- previously released test questions used
to test students’ progress towards future tests.

Two years earlier, I had been informed that my son’s kindergarten class would be
spending the week practicing for the TerraNOVA, a standardized test for young children
designed by McGraw-Hill. What would children be practicing, exactly? First, sitting
quietly -- a difficult (but necessary!) testing convention for kindergarteners. Secondly,
students would practice filling in bubbles — despite the oversize tests created especially
for young children, a Scantron is still baffled by incorrectly bubbled bubbles. I
wondered why kindergarteners would be taking, much less practicing, the TerraNOVA?

After all, the state didn’t require the testing of kindergarteners. Was there a new mandate I hadn’t heard about? No, there was no new mandate. The TerraNOVA, which students would spend the week practicing for, was, itself, practice for the MEAP they would encounter in subsequent years. Testing begat more testing.1

I thought of a scene from Disney’s Fantasia that had horrified me as a child. In the cartoon version of Goethe’s “Der Zauberlehrling,” Mickey, an apprentice to the Wizard, is left alone to clean the workshop, and calls on the Wizard’s magic hat to compel his broom to bring water from the well. To Mickey’s delight, the broom begins dividing and multiplying itself, carrying out its orders at an exponentially increasing pace. But Mickey’s pleasure turns to horror when the brooms won’t stop replicating -- or dumping water on him. Soon, an army of bucket-wielding brooms overwhelms him with a swirling ocean of water before the Sorcerer returns, gains control of the hat, and breaks the spell. With no Educational Sorcerer in sight to put standardization back in the hat, I figured it was only a matter of time before interim benchmarks (benchmarkbenchmarks?) would be used to measure students’ progress towards the benchmarks that would measure their progress towards the MEAP. With less and less time left for teaching, assessment itself would be redefined as teaching.

Ironically, the proponents of these standardized benchmark assessments in my district often used the language of “local assessments” to attempt to get teachers to buy into them. They drew a distinction between the state-level MEAP, scored by unknown readers, and our own “local” common writing assessments, designed and scored by us. This distinction mirrored the mantra of “local assessment” I was encountering in Brian
Huot's (Re)Articulating Writing Assessment for Teaching and Learning, a mantra I'll return to in Chapter Nine. But the fact that local educators were creating, administering, and scoring these benchmark assessments was not a sign that power was becoming decentralized, that teachers' professional judgment was being honored, or that assessment was becoming a local matter. These benchmarks were based on standards that local educators had had no hand in creating. And they were administered and scored with procedures and tools created primarily to guard against local teacher observation and judgment.

Besides, these supposedly local assessments showed little to no local variation. ELA assessments had begun to look virtually identical from district to district across the state: quarterly MEAP-style reading comprehension questions; monthly DIBELS probes; and county-wide timed writing prompts scored using some variation of the state rubric. In other words, you might be shopping at your local east coast grocery store, but all the produce was still coming from the same mega-farm in California. The fact that the highest-scoring writing samples might be published in a booklet distributed to next year's students seemed a patronizing nod and wink to the value that writing process proponents placed on giving students opportunities to write for audiences beyond their own teachers.

I had never heard anyone bother to explain how -- or even to argue that -- all this standardization and testing would actually improve teaching or learning. Although I'd missed it, this argument had been made when NCLB was originally under consideration. Diane Ravitch, Assistant Secretary of Education during George H. W. Bush's administration, helped to design and implement a voluntary state standards and testing program that paved the way for NCLB. An early supporter of NCLB, Ravitch has since
questioned and spoken out against the very standards and test-based system of accountability system that she helped to create. She tells the story of the “Texas Miracle” that convinced legislators and the public that a standards and testing program would improve teaching and learning and close the achievement gap:

Soon after the election of George W. Bush as president, we learned that he was the architect of this miracle in Texas. The miracle occurred because of this strategy: the state tested every child every year in grades 3-8; disaggregated their scores by race, ethnicity, and other characteristics; published the scores; and then honored the schools where scores went up and shamed the schools where they did not. Mirabile dictu, it worked! Or so a credulous press told us. Test scores went up, graduation rates went up, and the achievement gap began to close.

A few scholars warned that the miracle was an illusion. Walt Haney of Boston College and Stephen Klein of the Rand Corporation published critical reviews of the claims in Texas, but Congress ignored them...the Texas miracle was good enough to persuade Congress to pass sweeping legislation that affected every public school classroom in the nation, imposing federally mandated testing...But what we now know is that there never was a Texas miracle. At best, it was wishful thinking. At worst, it was a lie. On the National Assessment of Educational Progress of reading, 8th grade students in Texas had exactly the same score in 2009 as they had in 1998. No progress, period. No miracle.

No verification of standardization and testing’s effectiveness was necessary, it seemed, because standardization and testing had its own logic: the logic of replication.

FOLLOWING THE LOGIC OF REPLICATION TO ITS OWN ABSURD

CONCLUSION

Later, I’d see this logic of replication -- which needs no justification other than itself, and has no other goal than self-reproduction -- play out to its own absurd conclusion when an award winning middle school in New Hampshire failed to make Adequate Yearly Progress (AYP) two years in a row, and was thus designated a School in Need of Improvement (SINI). Teachers and administrators at the middle school knew full well why they had “failed.” In 2008, the principal had brought results of the previous
year’s state test to a staff meeting. Linda Rief, an eighth grade ELA teacher present at that meeting, recounted the principal’s announcement, “There’s good news and bad news. The good news is that we knocked the state test out of the park. The bad news is that we did so well that we’re going to fail to make AYP this year” (2009).

The principal was referring to the AYP ceiling effect. NCLB’s continuous improvement model doesn’t demand that schools must score well on the state tests, but that they must always score better than the previous year -- not just as a school, but subgroup by federally designated sub-group. This expectation is particularly difficult for small schools and high-performing schools; extreme individual scores drastically skew the average in small classes (and in accordingly small sub-groups), and high-performing schools have nowhere to go but down. This New Hampshire middle school was not only high performing, but also small -- a double whammy.

The principal’s remarks proved prescient. That year and the following year, students at the middle school scored high but not higher enough, and the school was designated by the state as a School in Need of Improvement (SINI) in 2009. But the principal’s understanding of the cause of that designation didn’t matter; the New Hampshire DOE used Indistar’s® web-based program, Steps To Success, to determine what had gone wrong. Steps to Success did not employ teams of education experts to investigate, consider, and weigh the cause of each school’s failure on a case by case basis, a system that might have led to the following recommendation to the teachers at this NH middle school: “Keep doing what you’re doing!” Instead, Steps to Success was a computer program, with a standardized menu of inputs and outputs. After processing the middle school’s data, the program spit out five “root causes” of the school’s failure. In
keeping with standardization’s logic of replication, three of the five “root causes” involved a lack of standardization: the school’s lack of a standardized reading curriculum, lack of standardized reading assessments, and lack of teacher training in standardized test administration (Barry 5).

Ceiling effect or no ceiling effect, the middle school was required to form a SINI team that would address these “root causes” of their failure. And so, in 2009, the SINI team set to work standardizing the school’s reading curriculum, implementing standardized reading assessments, and training its teachers in standardized test administration. The solution to failure in a standardized system? More standardization.

One of the teachers subjected to this training in standardized test administration was Linda Rief, who was told during the fall 2010 inservice that the “new school motto” was “Fill in the box.” According to the SINI team members leading the inservice, students who write more score better on the state writing test, and so teachers should admonish students to “fill in the box.” In addition to teaching at this school for over thirty years, Linda Rief is the founding editor of NCTE’s *Voices from the Middle*, recipient of the NCTE’s outstanding middle school educator award in 2000, author of *Seeking Diversity* (1992), a student and friend of Donald Murray and Donald Graves, and a leader of the *Teachers as Writers* movement. Linda described her outrage about the training:

> I kept raising my hand and asking, ‘Wait a minute. Let me get this straight. You want me to tell my students that it doesn’t matter what the write, they just need to fill in the box.’ They kept saying, ‘Yes, research shows that longer essays get higher scores!’ I almost walked out of the room and quit right then and there. (2011)

The irony of ordering Linda Rief (or any writing teacher!) to represent the act of writing to her students by telling them to “fill in the box” was apparently lost to the members of
the SINI team, obligated as they were to accept and then address the mysteriously identified “root causes” of their school’s failure.

In 2004, I had been frustrated that neither Carl nor Amy’s experiences seemed to matter during MEAP week at the alternative school in Michigan. Five years later, no one’s knowledge or experience mattered at this middle school in New Hampshire. Parents’ experience of the school’s excellence mattered less than the AYP formula that declared it a failure. Teachers and administrators’ knowledge that their failure was really just NCLB’s ceiling effect mattered less than the “root causes” identified by Steps to Success. And Linda’s considerable expertise and experience in teaching writing (and in teaching teachers to teach writing) was dismissed with the admonition: Just tell students to “Fill the box.”

When I went to the University of New Hampshire to pursue my doctorate in composition studies in 2008, four years after Carl crumpled up his “friendship” essay, I wanted to use the time to try to understand the dismissal of experience that was besieging K-12 education -- most often in the form of assessment. In the moment of its breakdown, as Latour would predict, the MEAP writing exam spoke eloquently of its disregard for Carl and Amy’s experience. I wanted to see if I could compel this assessment apparatus to speak again, and so, on Latour’s good advice, I went to archives, to see large scale standardized testing in the moment of its making.
JOHN B. WATSON’S BATTLE AGAINST MIND AND CONSCIOUSNESS: PARALYZING TO HUMAN EFFORT

If all men believed the teaching of the mechanical psychology...no man would raise a finger in the effort to prevent war, to achieve peace or to realize any other idea.
—William MacDougall, 1929

IS A TOOL EVER JUST A TOOL?

If I had been a strict ANT scholar, I surely would have gone to the archives with a completely open mind, ready to hear whatever the actors had to say. But I was looking for something specific. I was going into the archives with a working hypothesis: that tools continue to enact the purposes embedded in their very design, no matter the intent of their contemporary users. If the MEAP writing assessment tool had had the effect of dismissing Amy and Carl’s experiences in 2004, than I suspected I’d find a similar dismissal embedded into the very design of standardized testing -- as an intent.

I’d begun thinking about the origins, intents, and effects of tools when the word, tool, usually preceded by the word, just, had been repeatedly invoked to deflect my criticisms of rubrics. Rubrics are just assessment tools. They aren’t good or bad. It’s all about how they’re used. The implication was that the person using the tool, not the tool itself, should be blamed for a bad result. While I knew it would be silly to argue that the skill of the person using a tool doesn’t have some effect on the outcome, I couldn’t accept that tools were neutral. Although I wouldn’t use these terms until I’d read the work of Althusser, Berlin, and others in my doctoral program, I suspected that our conception of the work that needs to be done is, itself, ideologically constructed, and so any tool created to carry out this work might have those purposes embedded in its very structure. In turn,
that ideologically constructed tool might continue to enact its original ideological purposes quite apart from the intentions of those wielding it.

To illustrate how the shape of particular technologies continue to enact their original purposes long after their ideological moments (or creators) have passed, Langdon Winner describes the highway infrastructure designed by Robert Moses on Long Island: the parkways he designed passed beneath surprisingly low overpasses, some as low as nine feet from the curb. Those fortunate enough to afford cars or taxis would be free to use the parkways Moses had designed. But those who relied on public transportation -- most often the poor and minorities -- would be denied access.

Moses, whose racial and socio-economic prejudices were documented by his biographer Robert Caro, purposely designed those low overpasses to keep public transportation (busses with clearances of over nine feet) off his parkways. Moses’ ideological agenda was consistent throughout his half-decade of public service to New York City. He famously vetoed an extension of the Long Island Railroad to Jones Beach, ensuring that the “public” beach he designed would remain inaccessible to those needing public transportation. The intent of his public infrastructure matched its effects: to uphold socio-economic and racial disparity.

Those overpasses might appear to be “just another part of the landscape” (Winner 534), neutral but for the intentions of contemporary users. But Winner points out that decades after Moses’ career was over and the political alliances he forged to carry out his work have fallen apart, Moses’ public designs continue to serve his ideological agenda: restricting access to those who have the least. As planner Lee Koppleman put it to Caro,
That old son of a gun made sure that buses would never be able to use his goddam parkways.” (qtd. in Winner 534).

**MAKING SENSE OF A MUDDLE OF TERMS**

To pursue my hunch that assessment tools have ideological purposes embedded in their very structures -- purposes expressed through their effects -- I have drawn on Langdon Winner's example of Robert Moses’ highway overpasses. I’ve referred to these overpasses as part of an infrastructure, a term I’ll explore in more detail in Chapter Seven. But Winner himself never uses this term to describe these overpasses, preferring “technologies” and “artifacts.” The word “infrastructure” appears only once in Winner’s *The Whale and the Reactor*, in a much different context: “the technical infrastructure of freedom” (72). My choice to refer to Moses’ parkways as an infrastructure -- despite the fact that Winner never does, and despite the fact that I’m using the intents and effects of an infrastructure to illustrate the intents and effects of a tool -- represents some major slippage of terms. And this slippage demands explanation.

I’m not the first to make a muddle of such terms. In three paragraphs on pages 35 - 36 of *The Whale and the Reactor*, Winner uses the following terms virtually interchangeably: *things*, *inventions*, *technical systems*, *technologies*, *human artifice*, *the hardware*, *technical artifacts*, *technical devices*, and *technical things*. On the next page, Winner collapses these to answer the question posed in the chapter title — “Do artifacts have politics?” -- announcing that he’ll use “technology” to refer to “all of modern practical artifice” (37).
This, for Winner, is a useful reduction: He wants to suggest that that everything from a hammer to an automobile to the design of a network of roads has “politics” in a sense we usually reserve for human beings. This also proves a useful reduction for me to explore the human intentions carried within and the effects carried out by assessment objects -- whether we’re talking about a rubric, the MEAP writing exam, or the entire structure of accountability that gives the MEAP writing exam its meaning. But it also obscures some commonplace distinctions. Since I’ll make use of these distinctions again in the second half of the dissertation, I’ll introduce them briefly now, and propose another way to reconcile rather than reduce them: Latour’s explanation of how human and nonhuman actors assemble the social.

In common usage, we distinguish between three things that Winner might lump together under “technologies”: a hammer, an automobile, and the interstate system. The hammer is a tool, a device that helps us do something to something else -- to manipulate or have some effect on materials or accomplish some task. The automobile is a technology, the application of science to materials. It might be the product and combination of a number of different tools -- pistons, wheels, gears, switches, etc. The interstate system, finally, is an infrastructure, a collection of technologies working together to organize our movement, activities, and the other technologies (the automobile, bus, or motorcycle) that we choose to navigate it. In this sense, tools are the most basic unit of objects that humans use, combined or in increasingly complex ways to form technologies and eventually infrastructures.

But clearly, the line between the three terms becomes very muddled indeed when we look closely, beginning with the question of whether or not tools must be objects at
all. For example, the pragmatists argued that ideas themselves are tools: the pragmatists thought of ideas as “instruments for coping,” forks or spoons that are true in so far as they help us to accomplish various tasks (Menand 361). Furthermore, the line between a tool, a technology, and an infrastructure gets muddy upon close inspection. The materials used in the modern hammer are, themselves, the application of science to materials, requiring complex networks of people, materials, and ideas to build them. Why is the hammer a tool and not a technology? And (to shift momentarily from the interstate highway infrastructure to the international airway infrastructure), the international airway infrastructure could itself be considered just one of many technologies making up the larger infrastructure of globalization.

While the distinctions between a tool, a technology, and an infrastructure are intuitively satisfying, then, they are also inherently slippery. Once again, Latour’s gives me my idea fork, the tool by which I pick through this muddle of terms: the social as network. Latour reminds us that “the social” is really just the momentary networks created between human and nonhuman actors as they assemble into association with each other. These associations are necessarily provisional, shifting, temporary, fleeting. But they can be stabilized when certain patterns of interaction and assembly are repeated. Technologies (in Winner’s inclusive use of the term) or nonhuman actors (in Latour’s use of the term) help to solidify certain patterns of association, stabilizing networks (the social) by bringing the actors around them together in specific ways, over and over again.

Automobiles and highways, then, solidify particular associations of humans and nonhumans by creating repeated kinds of interactions. We keep producing automobiles — repeating particular associations between engineers, mechanics, factories, and highways
in the process -- because the highways that allow and demand them remain in place long past the original interactions between people and technologies that built them. Every time a particular kind of interaction is repeated, that part of the network is stabilized -- and these stabilizations are the basis of power, in Latour’s view. So, while there are interesting distinctions between tools and infrastructures that deserve preservation, they are all part of the same dynamic of assemblage, in which the objects brought together by people continue to bring people and objects together in particular ways long past the moments of their making.

**LOW EDUCATIONAL OVERPASSES**

What if the infrastructure of standardized testing were like Moses’ overpasses, a tool designed to restrict access long after its prejudiced architects were dead and gone -- even when contemporary users of this infrastructure were now touting *equal access*, vowing to leave no children behind? I’d seen this dynamic play out before: I’d noticed that the stated intent of the testing mechanism at the heart of NCLB (to promote equity) was at odds with its actual effects (to increase inequality). Well-funded districts retained their freedom and funding with predictably high scores (at least until the ceiling effect kicked in), while under-funded districts staggered under the weight of sanctions imposed on them for failing to do more with less.

Even when standardized tests were implemented to encourage supposedly progressive pedagogy, the effect was still to widen the gap between districts secure enough to invest in the slow growth supported by progressive pedagogy and those vulnerable enough to insist on short-term, superficial returns on scarce resources,
otherwise known as test-prep pedagogy. Massachusetts instituted a high stakes state-wide science test with the stated purpose of holding schools accountable for the state standards, which explicitly encouraged inquiry-based teaching in 1998. In a study of six Massachusetts school districts conducted as the tests went into effect, the test widened the gap between the inquiry-based pedagogy that one higher SES school was already using and the drill-and-kill teaching used at the three lower SES schools. Most telling is what happened in South Allenville, a middle SES school. In South Allenville, the principal and curriculum director were strongly in favor of an inquiry based approach and teachers devoted significant time to student collaboration, labs, and student investigations. But they gave up their focus on inquiry favor of more traditional methods at first whiff of the statewide test. According to researchers Falk and Drayton,

Inquiry-based science requires time for students to develop their own questions; time for data to be collected, discussed and analyzed; time for reflection; and time for students to share strategies and learn from each other (Drayton & Falk, 2001). The methods adopted for increasing MCAS scores in South Allenville diminished the time, the freedom, and the leisure for scientific exploration. (377)

Unlike the pre-test pedagogy at South Alleville, which had been relatively student-focused, the pre-test pedagogy at the three low SES schools were overwhelmingly “teacher centered” and continued to be so after the test. However, Falk and Drayton found two “inspiring” teachers at these schools who engaged students in inquiry-driven explorations. They explain the effects of the test on these two teachers.

Perhaps not surprisingly the MCAS exam had the strongest effect on these two inspiring teachers who had created inquiry-based classrooms despite the lack of labs and resources. These teachers had the most to lose. They worried that covering all the material tested on the MCAS would result in needing to change their pedagogical approach with their students. One of these teachers commented that there was no longer any time for him “to engage his students’ athletic or artistic mind” as the MCAS assesses only one type of intelligence. The other teacher commented that she would need to drop her “daily problem,” which was
motivating for her students, in order to increase material covered. (373-4)

The stated intent behind the implementation of the test -- to hold schools accountable for the inquiry-based pedagogy of the state science curriculum frameworks -- didn't matter; its effects were to restrict the access of lower SES students to pedagogy still accessible to many of the higher SES students. The high stakes test widened a pedagogical gap that some claimed, with all good intentions, that it was supposed to close.

Before I encountered the history of standardized testing in general and writing assessment in particular (Lemann; Gould; Hanson; Williamson; Broad; Yancey; White; Huot; O'Neill; Elliot; et al), I posed a question to a friend that sounded, even to me, like a conspiracy theory: what if the creation of standardized testing and writing assessment had originated with the desire to perpetrate inequality? What if its original intents matched its modern-day effects? It felt surreal shortly after to learn how this was true: I found that organized writing assessment had begun -- as a concern and set of practices -- to help universities develop gate keeping procedures they hadn't needed before the explosion of applicants in the early 1900's. Organized systems of writing assessment had begun, in part, to keep certain students out of higher education. While a great many people have worked very hard to ensure that these assessment gates let through individuals from traditionally under-represented groups, efforts to rid standardized testing of gender, racial, or socio-economic bias has only increased the number of groups or the percentage of individuals within a given group who go to college: these efforts have no bearing on the total number of individuals turned away.

In a more overtly disturbing sense, large scale standardized testing, which helped shape writing assessment practice and theory, was inextricably entwined with the
The eugenics movement in the United States: the first large scale standardized test in the United States, the Army Alpha Test (AAT), had been created by eugenicists and used to promote their causes. Robert M. Yerkes, who had helped create and administer the AAT during both World Wars, was an avowed eugenicist. Yerkes’ assistant, Carl Brigham, had analyzed the results of the AAT after World War I. In 1923, Brigham published this analysis (with a foreword by Yerkes) in *A Study of American Intelligence*, pointing out that southern and eastern Europeans had scored lowest on the test and arguing for “selective breeding” to preserve the integrity of the “Nordic race.”

Brigham’s book fueled growing anti-immigrant sentiment in the United States, and was used by Harry Laughlin, appointed by a House committee as an “expert eugenics agent,” to propose and pass the Immigration Restriction Act of 1924. This Act imposed quotas on all immigrants, enforcing the most restrictive quotas…wait for it…on those who had scored lowest on the AAT: eastern and southern Europeans. Brigham renounced his position in the 1930s, but not before he had helped to transform the AAT into the Scholastic Aptitude Test (SAT) and worked as a test maker for ETS’ parent organization, the College Board.

Although it had been news to me when I’d originally stumbled upon it in 2004, this troubling story was no secret. Pulitzer prize-winning journalist Nicholas Lemann had described the SAT’s links to the eugenics movement in *The Big Test*, published in 1999. But, in 2010, I was looking for evidence of a different troubling story: the dismissal of experience that I’d seen as Amy administered the MEAP and a New Hampshire middle school was declared a failure and then rehabilitated via standardization.
Given my interest in education, Brigham’s direct connection to ETS made him an obvious starting place for my search for the dismissal of experience in standardized testing. But several factors drew me to Yerkes instead. First, Brigham had been Yerkes’ assistant in developing the AAT. But, more importantly, when comparing biographies of the two men, I noticed Yerkes’ close connection to John B. Watson, the father of behaviorism. Yerkes and Watson had worked together on primate research, and had kept up an almost life-long correspondence, including the period before and after Yerkes worked on the AAT. I’d heard about Watson’s infamous infant experiments. They sounded cruel, and I didn’t think he could have undertaken them if he took infants’ experiences seriously. If Yerkes was going to discuss his dismissal of experience with anyone, I figured it just might be Watson.

JOHN B. WATSON FRIGHTENS LITTLE ALBERT, A “LOWLY PIECE OF UNFORMED PROTOPLASM”

Watson was famous for his research on primates and the theory of behaviorism, but also for his popular child-rearing book, *The Psychological Care of Infants and Children*, in which he argued that, “mother love is a dangerous instrument” (87). The book, written in 1924 “with the assistance” of his second wife, Rosalie (she was not given a proper byline), was based on Watson’s infant experiments. Watson was interested, among other things, in knowing if he could condition fear in infants. With Rosalie’s help, he systematically conditioned his young test subject, eight-month-old “Little Albert,” (who was naturally unafraid of any animal) to be afraid of a fuzzy bunny, and, by association, a fur muff and a furry-faced Santa Claus (23-30). Watson proudly
presented this research in *The Psychological Care of Infants and Children* as proof that parents (and, specifically, mothers) are to blame for children's fears, laziness, and neurosis. Furthermore, in his estimation, no parent knew how to be a good parent, and his work in behaviorism was the answer.

How could Watson live with himself as he systematically instilled fear in Little Albert? Was this simply the case of a researcher's natural enthusiasm in the days before International Review Boards? Or was there something particularly callous about Watson and this experiment? I knew that Watson and Yerkes had worked together at John Hopkins, sharing an interest in primate research and corresponding for decades. Some even credited their long relationship with Watson's full development and articulation of the theory of behaviorism. Had Watson spoken of his infant experiments to Yerkes? Had Yerkes spoken of the AAT to Watson? I felt that Yerkes, imposing standardized testing on hundreds of thousands of soldiers and then generations of schoolchildren, was somehow akin to Watson, instilling fear in a baby -- at least in the sense that I suspected each man of blindness to his test subjects' experiences.

As I paged through letter after letter, I was scanning for any mention of these infant experiments or the AAT. At first, I didn't think I was finding what I was looking for. However, I was gaining a more nuanced picture of Watson and his work. I came to admire his straightforward and dramatic voice, and I even gained a slightly different view of his work with infants, which I'd first seen as preposterous. I still think them preposterous, but I learned to respect at least one part of the larger argument he was using them to make. A thread of conversation between 1922 and 1932 demonstrated the best of
Watson’s voice, and also clued me in to one of the larger arguments that his “baby work” served.

In January of 1922, Watson wrote to Yerkes expressing his affection for this work in characteristically melodramatic fashion: “I have lost interest in university work. If I could get the baby work I would be willing to starve to death.” Ten years later, when Watson worked for the J. Walter Thompson advertising company, he would return to this wish, then dismiss it by explaining to Yerkes why “…there is too much water over the dam for me ever to go back into observational work.” He and his family had become accustomed to a lifestyle not supported by university work. But more to the point:

…if I ever went back I would want a real infants laboratory and I would want a hospital for about one hundred pure-blooded negroes at birth and a hundred pure-blooded Chinese and a hundred pure-blooded Anglo-Saxon Americans (if any such exist any more.)…Then I would say to old Father Time that any time he was ready, I was. I think I still have the guts to do this but it will require so much money that I am sure it will not come in my lifetime. (22 Jan 1932)

Watson didn’t discuss in this letter why he wanted such a racially specific groups of infants. But his reasons were easy to infer from his larger body of work, which I began to read. These reasons point to a major ideological difference between Watson and Yerkes, and this difference helped me see a more sympathetic context for Watson’s otherwise odious infant experiments. Both Yerkes and Watson were part of a generation of scientists dedicated to measuring and cataloging differences between the races and genders that supported the superiority of their own (Gould). While Yerkes was entirely sympathetic to eugenics, Watson was not, and his ideal infants laboratory was part of his plan to discredit the scientific foundation of the eugenicist view -- the theory of inherited intelligence.
Eugenics -- the idea that the progress and purification of the human species depends upon the extinction of inferior races -- rested on the theory of inherited intelligence, talents and instincts. Brigham and Yerkes claimed to have found support for this theory from the results of the AAT. But Watson’s study of infants convinced him that there were very few true instincts at all, and that all healthy infants had them -- breathing, babbling, wriggling, sucking, flailing, grasping, and startling at loud sounds and falling -- in more or less equal measure. In *Ways of Behaviorism*, Watson outlines his view of the infant:

> The behaviorist finds that the human being at birth is a very lowly piece of unformed protoplasm, ready to be shaped by any family in whose care it is first placed. This piece of protoplasm breathes, makes babbling, burbling, cooing sounds with its vocal mechanisms, slaps its arms and legs about, moves its arms and toes, cries, excretes through the skin and other organs the waste matter from its food. In short it squirms (responds) when environment (inside or out) attacks it (stimulates it). *(Ways 28)*

This view of the infant as a more or less blank slate allowed Watson to apply Pavlov’s stimulus-response model (a model Pavlov had created to explain animal behavior) to human learning. If, as Watson believed, the human infant is a wriggling mass of aimless movements, organized by conditioning into socially approved behavior, then conditioning, not inheritance, must be the reason for any differences between the races.

Watson’s fantasy multicultural infant laboratory was likely part of his plan to prove his claim -- taken up fictionally by B.F. Skinner a generation later in *Walden Two* -- that he could take any healthy infant and condition him into whatever kind of man he wanted:

> Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I’ll guarantee to take any one at random and train him to become any type of specialist I might select -- doctor, lawyer, artist, merchant-
chief, and, yes, even beggarman and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors (Behaviorism 82).

In fact, Watson labeled the foundational belief of the eugenicist -- the belief that intelligence or abilities are passed on within families as a matter of genetics -- “more dangerous than bolshevism” (Ways 26). He chalked up this dangerous belief to a “clever but unverbalized device (the Freudians would call it unconscious) for living forever (Ways 97). He explained,

It is hard for most of us to believe when we are dead that we are dead all over, like Rover. Some of us may boast of the fact that we are irreligious and even claim that we do not believe in a life after death -- but we soothe the personal hurt of this by believing and teaching that we as individuals will carry on in our children just as our parents have carried on in us...We hate to give up. We hate to die...this...is the reason for our unbounded belief in the inheritance of ‘mental traits,’ ‘dispositions,’ ‘capabilities,’ ‘tendencies,’ and ‘special abilities. (Ways 26-27)

You’ll notice Watson’s nod to Freud here -- eugenics represents an unconscious fear of death! -- at the same time that he was careful to redefine the unconscious as an “unverbalized device.” I wouldn’t understand this redefinition until later when I’d come to understand Watson’s view of consciousness itself.

Watson and Yerkes also discussed Yerkes’ work in testing, focusing mostly on Watson’s objections to Yerkes’ attempts to give multiple-choice tests to animals. But their conversation about testing occasionally wandered round to human subjects. In March of 1919, Watson made a request of Yerkes:

The Gilman School here has become interested in giving their youngsters some kind of a test and they asked me to ask you if you could free three hundred of the alpha blanks so that they can run through the school. I told them I didn’t know whether this would be allowed until the war was over, but that I would write you about it. My youngster goes there and of course I have a great interest in the school and I am interested in encouraging them to try modern methods. If you can free three hundred blanks, what would be the cost to them?
Yerkes obliged.

* * * *

I had hoped to find -- and, indeed, had found -- evidence that Yerkes and Watson had exchanged information about Yerkes’ testing program and Watson’s infant experiments. I’d built a larger context for Watson’s experiments. But I hadn’t found the dismissal of experience I’d gone to the archives looking for. I don’t know what I’d imagined I’d find: a secret, blood-stained pact in which Yerkes and Watson pledged to dismiss experience, starting with the torture of infants and generations of schoolchildren?

I felt a bit silly, actually, until a part of Watson and Yerkes’ conversation that had originally seemed secondary to my interests came more clearly into view upon repeated and more open-minded readings.

In looking for evidence that Watson’s callousness to Little Albert’s experiences had somehow influenced Yerkes as he developed the Army Alpha Test, I was focusing too narrowly around any mention of testing or “baby work” in the letters. It was as if I’d brought google search into the archives with me, typing, “testing, baby, dismissal of experience” into the search box. Of course, googling key words isn’t really listening to the actors, so it is no surprise that I couldn’t hear much. I mentally dropped out of graduate school for two days, comforting myself with chocolate ice cream while composing imaginary apologies to everyone who had rearranged their lives so that I could give up my high school teaching career to play scholar.

On the morning of the third day, determined to salvage some kind of seminar paper from the hopeless mess, I picked up the letters again. Once I resigned myself to not finding what I was looking for, I began to hear it in a thread of the conversation that...
had first seemed tangential to my interests. On April 7, 1913, Watson tries to understand a slight disagreement he’s having with Yerkes regarding Edward Titchener’s version of psychology:

I think our main difference lies in this: you [Yerkes] are willing to let psychology go its own gait, whereas I have probably an earlier and deeper interest in psychology than you have; consequently I am not willing to turn psychology over to Titchener and his school. The wise way would probably be to do as you suggest, -- call behavior physiology or biology, and leave psychology to the introspectionists. But I have sincere an attachment for psychology to do this way. I believe that it can be made a desirable field for work...I am not willing to turn psychology over to Titchener...

As I built a context around this letter -- Who was Titchener? What was an introspectionist? -- I realized that Watson was building a version of psychology opposed to the very notion of experience, labeling anyone interested in studying experience and its constellation of concepts (sensations, emotions, consciousness, mind) as “religious” or “mediaeval.” Moreover, Watson and Yerkes would swiftly, successfully, and personally apply their “mechanistic dogma” (MacDougall 69) to multiple realms of American life: the war effort, parenting, advertising, and education. Understanding the origins of Watson’s theory -- what it was a reaction to and what it was used for -- helped me to understand Amy and Carl’s experience of the MEAP and what I see as the frustrating inability of contemporary writing assessment theory and practice to adequately account for the experience of the reader and the writer.

**TITCHENER’S PSYCHOLOGY: TALK ABOUT MIND**

To understand how Watson’s “mechanistic dogma” discounted the very concept of experience, you have to understand the key concepts and methodology of Edward Titchener’s psychology. Titchner, the professional dragon that Watson would spend his
career slaying, began his 1889 *A Primer of Psychology* with the definition of psychology that would become the focus of Watson’s crusade:

> The Meaning of ‘Psychology.’ -- The word ‘psychology’ comes from the two Greek words psyche, ‘mind,’ and logos, ‘word.’ Psychology therefore means, by derivation, ‘words’ or ‘talk about mind’. (1)

Titchener defined mind not as an object inside the body that either holds or does things to thoughts and feelings, but as ‘the sum’ (5) of thoughts and feelings. In Titchener’s view, “we must not say that mind ‘has’ thoughts and feelings; but that mind is thoughts and feelings” (6). If mind is thoughts and feelings, not an object, then the way to study mind, the only method of the psychologist, was to look inward and talk about his experience -- introspection.

Introspection had also been used as a psychological method by Titchener’s teacher, the German philosopher Willhelm Wundt, who had had created one of the world’s first psychology laboratories in 1879. Because Titchener was interested in distinguishing psychology from philosophy and from the work of his teacher, he went to great lengths to make psychology more like an objective natural science in two ways. First, he wanted to use introspection to do for mind what the method of any other natural science did for its subject matter: to describe and classify its constituent parts and discover the laws that govern their interactions. Second, he sought to make introspection itself a rigorous, objective method that required several years of training.

Titchener believed that psychology must first be able to describe and classify its own territory -- mind. He began by dividing “total mind” into three parts: “the child mind, the adult mind, and the senile mind” (*Primer* 8). These three minds, he claimed, are made up of consciousnesses, or “‘mind now’; the mind of the present moment”
He divided these consciousnesses further into “concrete processes: ideas feelings, wishes, resolutions, etc.” (Primer 9-10). Each idea or feeling was a “specific item” of our consciousness, and each specific item of consciousness was made up of “elementary processes” -- affections and sensations that combine in unique ways to form our experience. The problem of psychology, as Titchener saw it, was to accurately describe these elementary processes through introspection and then to discover the rules that govern these elementary processes as they merge into concrete processes and then into consciousness and experience itself.

Introspective psychology depended on an individual’s description of his private experience, an admittedly subjective basis for a field that Titchener claimed should be more objective and scientific. Titchener’s approach to this problem began with implementing rigorous training for each Introspector -- Titchener called them “Observers” -- consisting of a series of standardized introspection exercises: “Only by looking inward can we gain knowledge of mental processes; only by looking inward under standard conditions can we make our knowledge scientific” (32). For example, Observers in-training were instructed to describe what they experienced when listening to certain tones or when exposed to various lights. Titchener invented several instruments for standardizing these exercises himself, including a “sound cage,” a mesh of wires surrounding the head connected to a telephone receiver designed to give each Observer practice in pinpointing the exact location of an auditory stimuli. In “Class Experiments and Demonstration Apparatus,” Titchener proposed a standard set of instruments for all psychology classrooms:

...whenever possible, we should call on the class to do psychology for themselves. The demonstration apparatus which I have in mind are, then,
apparatus which shall subserve this latter purpose: apparatus that shall standardise
the conditions for such introspections as the lecture-room and the lecture-hour
allow. (440)

Titchener considered Observers themselves to be highly trained scientific instruments,
and he bemoaned psychology’s great disadvantage in its ability to share results and
instruments across space and time. In the external sciences, scientists could easily ship
specimens and the conclusions of their experiments to other interested scientists. But to
facilitate the sharing of results and specimens in psychology, the inner science, Observers
themselves would have to be shipped at great expense and inconvenience (Lectures 278).

Still, the knowledge gained by Titchener’s Observers was not scientific, objective,
or standard enough for John Watson. Plenty of philosophers and psychologists had
critiqued Titchener’s methods and aims, including the philosopher John Dewey, under
whom Watson had studied at the University of Chicago at the turn of the century.
Dewey, too, had criticized introspection, but on far different grounds than Watson would.
In his 1891 textbook, Psychology, Dewey outlined a transactional objection to
introspection:

> When introspective analysis begins, the anger ceases. It is well understood that
> external observation is not a passive process... We shall see hereafter that there is
> no such thing as pure observation in the sense of a fact being known without
> assimilation and interpretation through ideas, already in the mind. This is as true
> of the observation of the facts of consciousness as of perceiving physical facts.
> (8-9)

Dewey took no issue with introspection as a psychological method, but simply pointed
out that observation is never objective. Watson, however, claimed to find Dewey’s ideas
altogether incomprehensible, proclaiming in 1936 that, “I never knew what he was
talking about then, and unfortunately for me, I still don’t know” (Watson, qtd. in Cheney
and Pierce 14). We can imagine Watson pausing and winking at his audience --
unfortunately for me. After living with almost a century’s accumulation of behaviorist influence in everything from advertising to educational policy, we can, of course, wink back -- unfortunately for us.

Dewey’s colleague, the psychologist William James, also took issue with Titchener’s ideas. He didn’t discredit introspection as an appropriate method for accumulating psychological knowledge, but he disagreed with Titchener’s assumption that the goal of introspection was to discover and describe elementary mental processes. In “On Some Omissions of Introspective Psychology,” James objects to “mental atomism”:

The traditional psychology talks like one who should say a river consists of nothing but pailsful, spoonsful, quartpotsful, barrelsful and other moulded forms of water. Even were the pails and pots all actually standing in the stream, still between them the free water would continue to flow. It is just this free water of consciousness that psychologists resolutely overlook. Every definite image in the mind is steeped and dyed in the free water that flows around it. With it goes the sense of its relations, near and remote, the dying echo of whence it came to us, the dawning sense of whither it is to lead. (16–17)

But Watson wasn’t interested in Dewey’s inherently subjective observations or James’ “free water of consciousness.” While James’ objection to Titchener’s mental atomism led to a conception of experience that influenced Husserl and other phenomenologists (Schuetz 442), Watson’s objections would extend to the very concept of consciousness -- along with purpose, value, and meaning -- itself (Costall 643).

*Experience* wasn’t a key term for Watson -- he more often attacked consciousness, emotion, introspection, or even thoughts. This explains, in part, why my search in the archives for any mention of experience led nowhere. However, as I read more of both Watson and Titchner’s work, I realized that Watson was reacting against the whole of Titchener’s psychology. And, in Titchener’s psychology, thoughts, emotions,
mind, consciousness, and experience were connected: thoughts and feelings were specific items of consciousness, which constituted various states of mind, which formed our experience. If Watson spoke of experience at all, he was careful to indicate that he wasn't talking about consciousness or emotion: experience to Watson was simply an organism's instinctive reactions to its environment.

**WATSON TALKS ABOUT CONSCIOUSNESS: AS UNSCIENTIFIC AS SOUL**

In a private disagreement with Yerkes about the role of introspection in 1915, Yerkes suggests that, "...there should be encouragement given those who are willing to make use of it [introspectionism]" (October 30, 1915). Watson counters two days later with a seemingly mild, almost conciliatory reply: "I am mulling over your statements about introspection. There doesn't seem to be as much disagreement as I had supposed" (November 1, 1915). But, as his next comment reveals, this small area of disagreement that remains is actually the crux of the matter.

I think the disagreement is rather on the subject of consciousness. I don't believe consciousness is any more a scientific concept than soul; and you can't agree with this or else you would not be willing to admit that introspection was a possible method. This is pretty radical I know, but it expresses my present views.

In other words, Watson realizes that he doesn't just disagree with the *method* of introspection, but also with the very *construct* on which the method is based -- consciousness itself. He thought it best to leave the soul and its secular counterpoint, consciousness, to religion (or philosophy, as long as psychology disentangled itself from philosophy).

In fact, "religion," along with "mediaeval tradition" and "philosophy," headed the list of insults that Watson was most likely to employ in putting down a rival theory (or
theorist). Consider how he uses these terms in the opening four sentences of Psychology from the Standpoint of the Behaviorist, published nine years after Watson’s small disagreement with Yerkes:

Mediaeval Tradition Has Kept Psychology From Becoming a Science.—Psychology, up to very recent times, has been held so rigidly under the dominance both of traditional religion and of philosophy—the two great bulwarks of mediaevalism—that it has never been able to free itself and become a natural science. Chemistry and physics have freed themselves. Zoology and physiology are now in the process of becoming emancipated. (1)

Strong language: held rigidly under; dominance; free itself; becoming emancipated. In fact, it is interesting to look at how Watson characterizes psychology’s concern with mind and consciousness in the first six pages of this book:

The Old Psychology of Mind and Consciousness
...deistic idol already fashioned and worshipped (vii)...crude dualism
...theological mysticism ...mediaeval tradition ...religion ... philosophy
...mediaevalism ...soul ...so-called states of consciousness ...phenomena of spiritualism ...not objectively verifiable ...no community of data ...mental curiosities ...introspection ...serious bar to progress ...failed to become a science ...deplorably failed ...it would not bury its past ...hang onto tradition ...will not bury their ‘medicine men’ ...subjective subject matter ... (vii - 3)

For all of his emphasis on objectivity, Watson presented his own ideas in an emotionally charged narrative. In Watson’s story, psychological medicine men such as Titchener had been sacrificing science and truth on the altar of mediaeval philosophers. Armed with a community of objectively verifiable data (his stimulus-response experiments), Watson was prepared to bury these psychological medicine men and emancipate psychology once and for all. Freed from the hocus pocus of mind and consciousness, he would help Man -- including the military, parents, advertisers, and teachers -- finally get control of his actions. Or, rather, the actions of others.
For those who cut our teeth on Freudian psychology, it might be tempting to note here that Watson chafed under the rearing of a strict fundamentalist mother who expected him to become a southern Baptist minister (Buckley 5). We might wonder if Watson’s demonization of mind was a “clever unverbalized device” in which Watson associated mind with mother love (“a dangerous instrument”) and thus religion (a “serious bar to progress”), disposing of both in his own attempt to “become emancipated.” But that would be to put Watson on Freud’s couch, a place he would never voluntarily lay his own head (not to be confused with his mind!).

In any case, Watson was convinced that if philosophy wanted to take up the issue of mind or consciousness (“a concept as unscientific as the soul”), then psychology must separate wholly from philosophy and study behavior alone. By 1912-13 Watson was busy trying to do just that at John Hopkins University. In May of 1912, he told Yerkes the results of his local institutional efforts to redefine his field: “President Remsen has resigned and I am exceedingly sorry….Today he gave me permission to separate from philosophy and psychology so that hereafter psychology will be independent.” But Watson wasn’t content with this local institutional separation. In the 1913 article often used to mark the beginning of behaviorism, he sets an ultimatum (which he disingenuously refers to as a “compromise”) for the entire field regarding the organization of philosophy and psychology:

It seems reasonably clear that some kind of compromise must be effected; either psychology must change its viewpoint so as to take in acts of behavior, whether or not they have bearings on the problem of ‘consciousness’; or else behavior must stand alone as a wholly and separate and independent science. (“Psychology as the Behaviorist Views It” 158)
As Watson would demonstrate in coming years, he wasn’t at all interested in compromise. In fact, a full decade before Watson’s Manifesto, the noted British psychologist William MacDougall had begun studying the “facts of behavior” in addition to the “facts of consciousness” without either calling for or causing a rift within the field. This wasn’t enough for Watson; he wouldn’t be content until the very term “fact of consciousness” was dismissed as a dangerous oxymoron.

In March of 1916, the discussion is perhaps sensitive enough that Watson feels he must affirm his friendship with Yerkes despite it:

Friendship to me is a far more precious than agreement about psychological positions. I have felt this keenly about Titchener and I have felt it even more so with respect to you. My feeling for you has nothing whatsoever to do with the question whether you think a certain amount of introspection is justifiable. If I dropped all my friends who think that a certain amount of introspection is justifiable, I would be living in a land of the friendless.

Watson was probably not exaggerating about living in the land of the friendless if he befriended only psychologists who shared his views; he was making his mark on psychology by rejecting not only the method but also the conceptual constructs of his field.

Key terms in Watson’s version of psychology were very different indeed from key terms in the psychology he was rejecting. While Dewey and James also had serious disagreements with Titchener, they shared a lexicon of acceptable (though debatable) psychological terms: experience, consciousness, mind, thought, feeling, memory, attention, intention, purpose, meaning. Watson’s psychology not only revolved around different terms -- stimulus, response, reactions, condition, conditioning, behavior, adjustment to environment, and reflex -- but he also called for psychology to erase and replace Dewey, James’ and Titchener’s terms. As Alan Costall, an historian of
psychology points out, consciousness and introspection weren’t the only focus of Watson’s ire: his larger crusade was against “purpose, meaning and value” (643).

Watson’s rejection of mind and meaning required that he explain popularly used terms such as thought. To explain “thought” (which he consistently referred to in quotation marks) without referring to mind or meaning, he first had to redefine language itself. In Watson’s view, language was not a product of the mind. Instead, it was conditioned laryngeal behavior. Here’s what he meant. The infant, as Watson’s “baby work” had indicated, is a “very lowly piece of unformed protoplasm...” (Ways 28) This protoplasm casts about with arms, legs, and vocal chords, randomly squirming (Battle 20). Mixed with exhalation, the random squirmings of the vocal larynx and palate form sounds. These random squirmings, then, are organized into “language” through conditioning and reinforcements. Adults praise (reinforce) children only when the child’s vocal squirmings correspond with the objects with which adults associate them. Talking then, is socially conditioned laryngeal behavior.

And here comes Watson’s turn to “thought”: “Thinking is merely talking, but talking with concealed musculature” (Battle 33). Why concealed? Conditioning, again. If Mrs. Smith and Mrs. Jones both frown at you when you voice your unfavorable laryngeal responses to the visual stimulation of their fashionable but dreadfully tottering hats, their negative responses to your response conditions you to speak these responses to yourself alone. Thought, then, is simply a double dose of conditioning: socially conditioned laryngeal behavior driven underground by the further conditioning of social mores.
In redirecting psychology’s gaze from mind to behavior, Watson didn’t just change psychology’s focus, methods, and terms. He also changed the nature and goals of applied psychology. In Titchener’s psychology, an understanding of an individual’s thoughts and feelings had a crucial role to play in ethics, and Titchener went so far as to assert, “Psychology is the foundation of ethics” (296). Titchener defined ethics as general laws that must be determined from the particular “facts of life” (296). He recognized that these facts of life are different in different societies, not to mention different for different individuals, so ethics must be sensitive to these differences. One way to assure such sensitivity was to use the insights of psychology -- drawn from the experience of individuals -- as an ethical check on the laws of ethics.

We see the Titchener’s concern for individual differences in his discussion on the application of psychology to pedagogy:

The problem of pedagogy is to lay down rules or norms of education...the abstract ‘child’ of psychology does not exist for education, not ‘the child,’ but real children, Katie Jones and Tommy Smith. Psychology cannot deal with Jones-ness and Smith-ness, but only with child-ness. Science, indeed, can never be ‘applied’ offhand. (298-99)

While Titchener acknowledged that psychology is a science that generalizes, the method of his science -- the description of an individual’s thoughts or feelings -- led to his concern for the interaction of generalizations and individuals. In other words, a science that made experience its special province had a special interest in how the application of that science affected individual experience.

The reverse was true in Watson’s science: a psychology that dismissed mind, thoughts, feelings, and consciousness as central constructs showed little interest in the effects of its application on the experience of an individual. And Watson was supremely
interested in applying behaviorism to every conceivable realm. In 1913 he announced the immediate usefulness of his version of psychology,

> If psychology would follow the plan I suggest, the educator, the physician, the jurist and the business man could utilize our data in a practical way, as soon as we are able, experimentally, to obtain them. Those who have occasion to apply psychological principles practically would find no need to complain as they do at the present time ("Psychology as the Behaviorist Views It").

Watson’s views on the application of psychology to education show none of the caution and respect for individual differences we see in Titchner’s discussion of “Smith-ness” and “Jones-ness.” To Watson, learning was a change in behavior in response to a stimulus, a process that is the same for Katie Jones as it would be for John Smith as it would be for a rat. Watson was so convinced of the applicability of animal research to human learning that when Edward Thorndike, who worked at Columbia Teachers College, suggested to Watson that he “move over into experimental pedagogy,” Watson protested to Yerkes:

> In my point of view I am already doing the scientific side to experimental pedagogy. Ulrich has just completed a problem on the distribution of learning in rats which is really a remarkable piece of work. We are planning now to carry out a similar line of tests upon human beings. If this type of work is not scientific experimental education pedagogy I guess I don’t know what the latter is. (21 February 1912)

Later, B.F. Skinner would take Watson’s position on the connection between rats and humans even further, leading Aruthur Koestler to write in 1964 that, “for the anthropomorphic view of the rat, American psychology has substituted the rattomorphic view of man.”

Without an introspectionist’s grounding in the experience of the individual, Watson had no qualms about proclaiming the goal and application of psychology as the “control of behavior,” which very quickly came to mean, in practice, the control of
individuals. While early critics of behaviorism attacked Watson's lack of attention to states of mind or consciousness on moral grounds, they perhaps underestimated the potential power of behaviorism to do what Watson says it would: to control individuals' behavior.

PARALYZING TO HUMAN EFFORT

In 1929, Watson and William MacDougall published their debate about behaviorism (held publicly in Washington, DC five years earlier) in *The Battle of Behaviorism*. MacDougall, a British psychologist, begins by a tongue-in-cheek apology for what he promises will be a “ruthless” critique:

> We all know that Dr. Watson has his feelings, like the rest of us. But I am at liberty to trample on his feelings in the most ruthless manner; for Dr. Watson has assured us (and it is the very essence of his peculiar doctrine) that he does not care a cent about feelings, whether his own or those of any other person. (44)

MacDougall proceeds to trample on Watson’s feelings by making the “impudent claim” that “I, rather than Dr. Watson, am the Arch-Behaviorist” (49). MacDougall reminds Watson that he was doing behavioral observations of infants ten years before Watson would undertake his baby work. And, as early as 1905, MacDougall was publishing calls for the inclusion of behavioristic studies in psychology (51). He begins his critique by establishing their similarities: “Dr. Watson and I are, then, engaged in the same enterprise, the endeavor to reform psychology by correcting the traditional tendency to concentrate on the facts of consciousness to the neglect of the facts of behavior” (53). Finally, he puts his finger on their differences:

> The difference between us...is that I, unlike Dr. Watson, have not made myself at the same time famous and ridiculous by allowing the impetus of my reforming zeal to carry me over from one lop-sided extreme position to its opposite, from
exclusive concern with the facts of consciousness to exclusive concern with the facts of behavior. (53)

In his own rhetorical flourish, Watson had pointed out at the beginning of their debate that MacDougall was a skilled rhetorician (7). Indeed, MacDougall would effectively level all of Watson’s own favorite accusations against him, here ascribing religiosity to Watson’s professional work by referring to his “zeal.”

Next, MacDougall borrowed another favorite accusation from Watson, calling Watson’s focus on behavior to the exclusion of consciousness impractical. Watson described talk about the mind as useless metaphysical philosophizing, and touted behaviorist psychology as the practical psychology. To portray Watson’s exclusion of introspection as impractical, MacDougall described the following hypothetical experiment. Dr. Watson sticks a pin in MacDougall’s finger. Watson, interested only in the facts of behavior, notes only that MacDougall draws his hand away, dismissing MacDougall’s introspective report of the experience. When Watson repeats this experiment on “a thousand hands, hands of babies, men, and monkeys” (55) and finds the same reaction, he might generalize that “sticking a pin into an extended hand causes it to be promptly withdrawn” (56).

But what if Watson were to repeat the experiment? This time, Watson sticks the pin into MacDougall’s finger, and MacDougall steadfastly leaves his hand in place and smiles calmly. What should Watson do with this? Without MacDougall’s introspective report, he asserts, Watson “can do nothing with it” (56). MacDougall doesn’t give us this second hypothetical introspective report, but we can imagine it in all its mischievousness: I willed myself not to draw my hand away, to prove that your mechanistic psychology cannot account for what is in my mind, and what is in my mind matters.
Lest anyone think that Watson’s insistence that behavior alone matters is so trivial an issue as pins stuck in hands, MacDougall reminds his audience that the role of a jury is to determine “the intentions, the motives, and the responsibility of the accused” (70), not just his behavior. Watson’s psychology, according to MacDougall, ignores a fundamental fact of human nature:

“Man is fundamentally a purposive striving creature. He looks before and after and longs for what is not. And he does not merely long; he strives to achieve that which he longs for, to bring about what is not yet actual, what he judges and desires should be…” (72)

MacDougall’s ruthless critique is just gathering steam. According to MacDougall, Watson’s focus on behavior to the exclusion of concepts such as “‘incentive,’ ‘motive,’ ‘purpose,’ ‘intention,’ ‘goal,’ ‘desire,’ ‘valuing,’ ‘striving,’ ‘willing,’ ‘hoping,’ and ‘responsibility’” (69) is more than ridiculous, impractical, and ignorant of human nature: it leads to a psychology paralyzing to human effort:

I submit to you the proposition that any psychology which accepts this mechanistic dogma and shapes itself accordingly is useless, save for certain very limited purposes, because it is incapable of recognizing and of taking into account the most fundamental facts of human behavior...If all men believed the teaching of the mechanical psychology (and only beliefs that govern action are real beliefs) no man would raise a finger in the effort to prevent war, to achieve peace or to realize any other idea. So I say that the mechanical psychology is useless and far worse than useless; it is paralyzing to human effort. (69-72)

On one level, MacDougall was strikingly wrong: behaviorism was not useless, as it was put to use immediately. On another level he was strikingly right: the uses to which Watson and Yerkes put behaviorism -- parenting, education, military morale, and advertising -- would prove paralyzing to human effort.

Watson’s application of behaviorism to parenting confirmed MacDougall’s fears about the application of behaviorism to human affairs. After describing the American
press’ adulation of Watson after the publication of *The Psychological Care of Infants and Children* and noting Watson’s “wise policy of abstaining from all attempt to reply to criticisms” (87), MacDougall goes straight once again for that sensitive Watsonian nerve:

Watson’s ideas are a “gospel” that must be accepted not on scientific evidence, but on “blind faith.”

Dr. Watson knows that if you wish to sell your wares, you must assert very loudly, plainly, and frequently that they are the best on the market, ignore all criticism, and avoid all argument and all appeal to reason. The response of the American press to his new book shows how sound these methods are... We have to face the prospect that in a few years’ time many thousands, perhaps even millions, of young victims of this propaganda on behalf of crass materialism will be bringing up their families without other guidance than their blind faith in the Behaviorist’s formulae. Having learned that all such words as effort, striving, grab, ideal, purpose, will, are entirely meaningless, they will be seen throughout this broad continent striving to form the character of their children by ‘conditioning their reflexes’ and pathetically endeavoring to gain their affection by stimulating their ‘erogenous zones’; for according to the gospel of Dr. Watson, that is the one and only way. (96)

MacDougall’s prediction that “millions” of parents would rely on Watson’s behaviorist parenting advice was sadly accurate. But perhaps not even MacDougall would have dared predict how devastating Watson’s reliance on his own behaviorist parenting advice would be for his children.

In a series of family tragedies that Watson’s granddaughter, Mariette Hartley, would blame on Watson’s behaviorist parenting techniques, Watson’s first son with his second wife, Rosalie, committed suicide several years after Watson’s death. His second son with Rosalie suffered chronic stomachaches that began when Watson systematically broke his attachment to his mother at the age of three (“mother love is a dangerous instrument”) by sending him away to live with his grandparents and punishing his wish to express affection to his mother and his older brother. (Watson reported on these
“experiments” in breaking attachment to create “minimal fixations” in Behaviorism, referring to his son as “Subject B.”) After taking up to 20 aspirin a day as an adult to deal with these stomachaches, Watson’s son died from bleeding ulcers in his 50’s.

Watson’s daughter with his first wife, Polly, would survive multiple suicide attempts. In her 1990 memoir, Breaking the Silence, Polly’s daughter Mariette Hartley describes the damage done by Watson’s parenting ideas on her mother. In the pursuit of “minimal fixations,” Watson maintained that children should not be touched, save a light kiss on the cheek before bedtime. Hartley described once trying to grab her mother’s hand when crossing a busy street. Polly pulled away, exclaiming angrily, “People will think we’re lesbians” (103).

It is perfectly possible for a scientist who denies the value of constructs such as mind, experience, thought, consciousness, emotion and meaning in his scientific research to have affectionate, meaningful relationships with his children. But Watson consistently, proudly, and publicly applied his science to parenting, denying himself and his children -- in theory and in practice -- a relationship based on mind, experience, thought, emotion, and meaning. And he encouraged all parents to follow his behaviorist lead. The effects of Watson’s theories on his relationships illustrated MacDougall’s warning: only a psychology that acknowledges and honors the experience of individuals has any hope for the purpose of furthering human efforts.

Eerily echoing MacDougall’s postscript, Hartley refers to the effects of Watson’s theories in terms (legacy, inheritance, genetic flaw) antithetical to his brand of behaviorism:
Grandfather's theories infected my mother's life, my life, and the lives of millions. How do you break a legacy? How do you keep from passing a debilitating inheritance down, generation to generation, like a genetic flaw? (18)
John B. Watson is clearly the antagonist of my story. But in spending time with Watson in the archives and in his published work, I had found much to appreciate. Watson’s voice was dramatic, direct, and compelling. (“If I could get the baby work, I would be willing to starve to death.”) I admired the clarity of his critique of the eugenicist’s reliance on inherited mental traits. (“It is hard for most of us to believe when we are dead that we are dead all over, like Rover... We hate to die...this...is the reason for our unbounded belief in the inheritance of ‘mental traits’...”) I even laughed at the elegance of some of his reductions. (“This piece of protoplasm breathes, makes babbling, burbling, cooing sounds with its vocal mechanisms, slaps its arms and legs about...cries...In short it squirms...”)

It wasn’t just that I was amused by Watson’s phrasing. I didn’t think he was all wrong. In a sense, I accept his formulation: if you want behavior to change, you need to consider changing the conditions to which that behavior is a response. In a sense, I share Watson’ doubt that Titchener’s Observers could solve the problems confronting psychology by describing the distribution and duration of the light spots behind their eyes. So it isn’t that Titchener is the fallen hero of my story.

Amedeo Giorgi outlines two very different solutions to the problems presented by mentalism -- the idea that the “mental” is a “substance that interacts with the body or the world”:

---

CHAPTER III

“A DEBILITATING INHERITANCE”: WATSON’S DISMISSAL OF MIND INFECTS EDUCATION

Never will you get a better psychological subject than a hungry cat.
--Edward Thorndike’s dissertation, 1898
...for the phenomenologist, mentalism is a motivated error. It is trying to say something even if it is poorly expressed. Radical behaviorism tries to solve the problem posed by mentalism by translating mentalisms into behaviorism, which essentially means doing away with whatever it is that mentalism referred to. Phenomenology attempts to recover the referent of mentalism and integrate it in a more adequate expression. (208)

Giorgi captures the spirit of my judgment of Titchener: he was “trying to say something, even if it is poorly expressed.” To put it another way: Titchener may have been misguided, but he was onto something.

Had I no knowledge of what followed, I might have applied to Watson the judgment I applied to Titchener. But I do know what followed. And so, my emphasis within this judgment inverts -- Watson may have been onto something, but he was misguided -- and ends with an ominous postscript -- and his influence on American education was debilitating.

I borrow the word “debilitating” -- to make weak and infirm -- from Watson’s granddaughter (“a debilitating inheritance”). I find it particularly apt in describing the effect of Watson’s theories on education. In Education and The Cult of Efficiency, the densely documented study of the efficiency movement’s influence on education, Raymond Callahan grapples with the question of why American education in the early 1900’s was so “vulnerable” to the business model when, as he asserts, “education is not a business” and “the school is not a factory” (vii). In trying to explain the “extreme weakness and vulnerability of schoolmen” (viii) to the factory model, Callahan points to two dynamics. First, he notes the low status of educators, asserting that “the process of cultural diffusion” (vii) makes it predictable that the concepts of a high status group (business) would trickle down to the low status group (education). But, after marveling at the extent to which these business concepts came to prevail in education, Callahan
further blames the local power structure of schools that made superintendents (“schoolmen”) vulnerable to local criticism (viii-ix).

I propose a third dynamic to explain the weakness of “schoolmen” in resisting the application of industrial models to education. As one of the most prominent members of a field dedicated to the way people work, vii Watson was uniquely positioned to insist on the need to carefully consider the consequences of importing industrial or economic models to uniquely human realms such as education or family life. Such models might function to organize people’s lives and decisions more or less “efficiently.” But what effects might these industrial models have on human experience? This question seemed custom-made for the psychologist. Watson not only declined to rise to the occasion provided by such a question at a critical moment in human history, but he also set out to change psychology itself so as to be unfit to answer it. viii

WATSON’S MODEL OF BEHAVIOR AND THE INDUSTRIAL MODEL OF PRODUCTION: WAIT, WHAT’S THE DIFFERENCE?

Watson didn’t question the industrial model’s effect on human experience because it was virtually indistinguishable from his model of behavior. Both were reductionist, but more to the point here, both operated on the assumption that the consciousness or experience of the humans involved was irrelevant. And in the absence of an individual’s sense of purpose or motivation, both sought to control behavior externally through rewards and punishments.

The skilled artisan model preceding the industrial model had been based on the relationship between a human being and his or her craft, a relationship that took a lifetime
to develop. This relationship served to organize other relationships, as family members participated in various stages of production, and apprentices often lived and worked under the artisan’s roof. And the artisan was often involved in every step of production – from selecting and sometimes even growing or gathering raw materials, to fashioning or commissioning tools necessary for the craft, to selling or trading the product.

While the industrial model mesmerized the public by generating great wealth for a new American aristocracy of Rockefellers and Carnegies, it worked by dispensing with an individual’s lifelong relationship with craft. Along with technological advances, this industrial model was powered by a devastating insight: if you break a complex process of production down into a series of simpler actions, you can replace the skilled (and expensive) artisan with cheap, unskilled workers and, better yet, machines. Unlike human workers, these machines required no training, cajoling, or bonus pay on holidays. They didn’t need to sleep. They didn’t come to work with hangovers or otherwise disrupt the manufacturing process with injuries, the flu, or family emergencies. They didn’t resent the wealth and power of the owner. They never got bored. And they didn’t need to understand what they were doing or why. Their consciousness or experience was irrelevant to the production of goods.

There was still a place for human workers in this model, of course, but they were treated like the machines constantly threatening to replace them. Unlike the artisan, the lowest worker in the factory didn’t need a lifetime to develop a vision for the process that led to product he was making – or even to know what the product was. It is telling that in Charlie Chaplin’s satire of the industrial model in Modern Times, it is entirely unclear
what is actually being made at the factory where Chaplin’s character unthinkingly tightens anything resembling a nut, including the buttons on a woman’s skirt.

Even activities we’d consider cognitive -- such as the solving of complicated math problems -- was revolutionized by the factory model. A mathematician would break complex equations into simpler parts, then train and supervise teams of people with basic computational skills (often the poor or wives of professors) to create the mathematical tables required by astronomers and engineers. Person A added two numbers, Person C divided the sum of Person A’s calculation with the product of Person B’s, and so on down the mathematical assembly line, none of these human computers sure what the numbers were adding up to or how they’d be used. For workers in these human math factories that reached their height in the 1930’s, no understanding of advanced mathematics and no sense of purpose were needed (“Return...”).

Nobel Prize winning physicist Richard Feynman had supervised such a group of human computers (most of them women) at Los Alamos to do the calculations necessary to build the atom bomb (Skinner 103). It was great achievement, really -- the quick solving of complicated equations by those who can do little more than basic computations -- until you consider the experience of these housewife computers. It might be fun for about 10 minutes, a great parlor trick: “Look! You do this, I’ll do that, and betcha didn’t know we just did advanced calculus!” The work might even be tolerable for several months, especially if you labored under the hope that your calculations might add up to the end of a World War. (That is, if you don’t consider the experiences of those victims of the bomb you are busy helping to build. Notice any themes here?) But a year’s worth of such work? A decade’s? With the vision for the whole confined to the
upper levels of management, most of the human workers in the industrial model didn’t need to understand what they were doing or why -- whether they were making steel rods or mathematical solutions. The worker’s consciousness and sense of purpose was irrelevant to the production of goods.\textsuperscript{x}

\textbf{WATSON’S MODEL OF BEHAVIOR AND TAYLOR’S MODEL OF EFFICIENCY: \textit{WAIT, WHAT’S THE DIFFERENCE?}}

Behaviorism perfectly complemented the efficiency movement that shaped education in the first half of the twentieth century. The efficiency movement began by seeking to squeeze the last drop of profit out of each worker through “time-motion studies.” These studies were developed and introduced to the business world by Frederick Taylor and Lillian and Frank Gilbreth, Sr. (The Gilbreth’s son, Frank Jr, chronicled his parents’ application of their scientific efficiency studies to their family of twelve children in \textit{Cheaper by the Dozen}.) Most of us know Taylor’s story by now, reported in his 1911 bestseller, \textit{Principles of Scientific Management}, and retold by education writers such as Raymond Callahan and English Education writers and compositionists such as Robert Tremmel and Tom Newkirk. But it is a story that never fails to shock educators, as it brings aspects of their school lives into focus, so it bears repeating here.

At Bethlehem Steel, Taylor’s assistant watched 75 men loading pig-iron. Documenting the separate motions that made up the job, the assistant then calculated the time necessary for one man to perform each motion. By adding these separate times back together, Taylor’s assistant determined that the men \textit{should} be moving the pig-iron four times faster than they currently were. This was a time-motion study.
Taylor and his assistant then looked for the fastest, strongest loader -- a man Taylor referred to as “Schmidt.” Taylor told Schmidt he’d give him $.70 a day more than he was currently earning to move four times more pig-iron than the average man was currently moving; he would just have to follow the steps *exactly* as Taylor had worked them out after examining his assistant’s time-motion study. When he managed to meet this goal, Schmidt would train all the other workers to perform exactly the same movements. Those who refused or were unable to meet the goal were “persuaded or intimidated into giving it up” (Taylor 56). As it turned out, 7 of 8 men were unable to meet the goal.

This particular factory job -- loading pig-iron at Bethlehem Steel -- had been repetitive and divorced from a larger sense of purpose before Taylor’s help: *Just move this pile from here to there!* But now, with the contributions of the first efficiency expert, the worker’s every move was prescribed: *Take three steps here, pivot there, lift now.* It’s one thing to revel in finding out, for oneself, the smallest possible number of motions required in doing a task: my very organized college roommate once reported, with a sense of glee I’ll never understand, that she’d figured out how to shower with only 13 movements. It’s another thing, indeed, to have your entire day prescribed in this manner by your boss.

The steps of Taylor’s applied management work bear a remarkable resemblance to Watson’s behavioristic heuristic. Step One: Ignore the purpose, mind, and experience of the human under consideration. Step Two: Break his behavior into the smallest possible parts. (Remember Watson’s insistence that *thought* can be broken down to the random
vocal behavior of the infant?) Step Three: Measure these parts of behavior. Step Four: Insist that the way to control this (measured) behavior is to use rewards and punishments.

It was this version of efficiency -- buttressed by a theory of learning and behavior that supported it perfectly -- that transformed education. As Callahan tells the story, efficiency experts flooded the schools to conduct time-motion studies in classrooms, stopwatches and clipboards in hand, as students figured columns of numbers and memorized lists of spelling words like men loading so much pig-iron. And generations of teachers were trained in “classroom management,” taught to dole out incentives and demerits to manage the behavior and productivity of the nation’s most important natural resource, otherwise known as *children*.

Before Taylor and Watson’s work caught on, it was still possible for human experience -- including thoughts, sensations, and feelings -- to be central concerns for those who would apply the concept of “efficiency” to education. In 1900, John Dewey had addressed a concern related to the concept of efficiency: “waste.” In a lecture titled “Waste in Education” published in *Schools and Society*, Dewey acknowledged that the waste of money and things must be taken into account, but then declared his own interest in the issue: “the primary waste is that of human life the life of the children while they are at school” (77).

The primary cause of that waste, Dewey went on to argue, is the lack of a connection between the child’s experiences of the world outside of school and his experiences inside school, made all the more pronounced because of the “isolation of school from life” (89). Dewey was careful not to imply that connecting school to life meant that schools should simply teach “commercial geography and math”
Instead, Dewey was talking about a broader connection between school and life, in which educators use the child's experience as the materials of his own education, helping him encounter ideas to help the child "...get into satisfactory relation with other parts of [his] experience" (James, paraphrasing Dewey). To Dewey, efficiency in education should concern itself with experience of the child, not with the economic bottom line.

Similarly, in 1909, the prominent educator, Charles Eliot, published a book titled *Education for Efficiency* that looks nothing like the version of efficiency for education documented in Callahan's study. Eliot's book begins with his definition: "By efficiency I mean effective power for work and service during a healthy and active life" (5). Unlike Taylor's version of efficiency which ignored the worker's experience and focused on the bottom line, Eliot asserted that education for efficiency must include "training of the bodily senses and the care of the body" (6): "the training of the ear" by "reading aloud, reciting prose and poetry, and music" (7), and the "careful training" of a sense of smell, "for it is the daily source of keen gratifications, the frequent renewer of mental associations and the best natural protector against corrupted food, drink, and air" (7-8). And, in a sentiment that would be antithetical to Taylor's later efforts to outsource thinking to a special class of managers, Eliot declares that one of the "fundamental matters with which efficiency for education should deal" (6) involves thinking:

It may seem strange to say so, but it is perfectly plain to persons who have been carefully observing the rising generations that education for efficiency must especially endeavor to induce young people to think...quiet and intense thinking, and particularly that inventive thinking, which is something more than sorting or putting in order materials supplied to the mind from without. (13-14)
“thinking absolutely requires personal initiative, or a compulsion from within and not from without” (15) and so adults should not “use compulsions which depend on fear of pain” (16). Instead, all the “teacher, parent, or friend” can do is to find subjects with “a natural interest for the young thinker” (15), instill the conviction that thinking will lead to later satisfactions, and provide a “personal influence” that will “rather lead than drive; for the personal initiative in thinking is indispensible” (17).

Efficiency for education meant reciting poetry? Intense, inventive thinking? Personal initiative and influence? And cultivating the nose, that daily source of keen gratifications? There’s no whiff of business, no hint of rewards and punishments, no mention of time-motion studies, no suggestion that efficiency has anything to do with speed. In fact, Eliot ends the first section of Education for Efficiency with an injunction against the narrow versions of educational efficiency that were soon to come:

“Educational for efficiency must not be materialist, prosaic, or utilitarian; it must be idealistic, humane, and passionate, or it will not win its goal” (29). Eliot’s book and Dewey’s lecture suggest that the concept of efficiency alone was not responsible for the infiltration of education by industrial models of business. Eliot and Dewey had maintained the barrier between purely economic interests and human affairs. xi

But when John B. Watson, a nationally-renowned expert in the domain of human affairs, effectively argued that we don’t have to pay attention to human experience — feelings and thoughts — the barrier separating industrial models from realms such as parenting and education gave way rapidly. Over it rushed an enthusiasm for the business model generated by industrialism’s great material success in the nineteenth and early
twentieth centuries. Immediately, and very nearly totally, efficiency for education came to be equated with speed and short-term, measurable returns on taxpayers’ investments.

**WATSON’S MODEL OF BEHAVIOR AND THORNDIKE’S MODEL OF LEARNING, TEACHING, AND EDUCATIONAL MEASUREMENT: WAIT.**

**WHAT’S THE DIFFERENCE?**

Behaviorism didn’t just make its way into schools through the time-motion studies of the education efficiency experts. Watson’s model of behavior was so simple that it doubled (or quadrupled) as a model of learning, teaching, and educational measurement. *Model of Behavior:* something stimulates the organism to respond (behave), and it repeats or avoids that behavior when externally reinforced (rewarded or punished). *Model of Learning:* the pupil “learns” when his behavior is reinforced (rewarded or punished). *Model of Teaching:* the teacher “teaches” when he reinforces (rewards or punishes) the pupil’s behavior. *Model of Educational Measurement:* the behavior must be measured so that the appropriate reinforcement (reward or punishment) can be scientifically applied.

Note how educational measurement becomes inextricable from teaching and learning in this formulation. We hear the phrase, *assessment for teaching and learning,* regularly invoked by contemporary assessment scholars in composition studies who couldn’t be accused, by any stretch of the imagination, of being behaviorists. But it is important to acknowledge this link between educational measurement, teaching, and learning, a link that often overwhelms the link that some proponents of formative assessment would make between teaching and assessment, a point we’ll return to later.
While Watson’s behaviorist model quadrupled as learning, teaching, and educational measurement models, Watson himself wasn’t much interested in schools. Watson’s disdain for joining the world of teachers and students was clear. After receiving Edward Thorndike’s invitation to “move over into experimental pedagogy” and protesting to Yerkes that he already was doing the experimental side to pedagogy, he wrote, “I could never be intellectually happy if I have to give over my work and take up the mollycoddling of psychology for teachers.”

Thorndike, at Teacher’s College at the time, had close professional ties to Watson and Yerkes. Besides inviting Watson to join experimental pedagogy, Thorndike had developed and administered the AAT with Yerkes. Of course, the three men didn’t agree about everything. Watson objected to Yerkes’ view of inherited intelligence. Yerkes viewed some amount of introspection as allowable while Watson tolerated none. And Watson disagreed with aspects of Thorndike’s Law of Effect. But all three shared a belief in the Holy Trinity of Behaviorism: the study of animals would shed light on human behavior: a vocabulary dominated by stimulus, response, and reinforcement was sufficient to explain behavior; and terms such as consciousness and drive were, to use Watson’s phrase, “medieval hocus-pocus.” Of course, the three weren’t just believers in this Trinity: they’d penned, sanctified, and preached it.

In fact, Thorndike’s first scholarly contribution was his 1898 dissertation, in which he explained how cats, dogs, and chicks learn without reference to conscious mental experience. Thorndike began by posing a question concerning what we can assume about the mental life of a cat as he learns to run to his milk dish upon hearing, “’kitty, kitty!’”:

80
Does the kitten feel ‘sound of call, memory-image of milk in a saucer in the kitchen, thought of running into the house, a feeling, finally, of ‘I will run in’? Does he perhaps feel only the sound of the bell and an impulse to run…? (2)

The different implications of these two proposed explanations are clear once we remember the behaviorist’s agenda to purge psychology of any concern with mind and consciousness. If we say that the kitten learns to run to the milk dish because our call prompts the cat’s memory-image of milk followed by the thought of running to it, we have assumed cat-mind. But if we say that the kitten learns to run to the milk dish because our call prompts his impulse\textsuperscript{33} to run -- an impulse which has been “stamped in” by past pairings of the call with the dish of milk -- we have bypassed any mention of cat-consciousness in our discussion of cat-learning.

To test his two explanations, Thorndike placed cats (in addition to dogs and chicks) into puzzle boxes. Each box contained a latch that opened when bumped or pressed, with food placed immediately outside. The cats instinctively struggled against confinement: Except for one old and one “uncommonly sluggish” cat, Thorndike reported, “the vigor with which it struggles is extraordinary” (13). Inevitably, one of these random, vigorous, instinctual movements triggered the mechanism, opening the door and allowing the cat access to food. Thorndike repeated the experiments until the cats either tired or mastered the latches, usually within a period of about three hours. Rewarded by food on the outside, Thorndike noted that the cats repeated these successful movements more quickly each subsequent time they were placed in the same box, an improvement that Thorndike represented in hand-drawn “time-curves. (Rorshak: Does it amuse or distress you to know that you compare yourself to a vigorously struggling cat when you refer to your own “learning curve”?)
But had the cats created mental representations of the workings of the latch in order to undo them more quickly? Did their learning involve *understanding*, and thus, *mind*? In the absence of the cats’ introspective reports to the contrary, Thorndike said no. To support the position that cats had no mind -- or, at the very least, that their minds were not necessary for their learning -- Thorndike put the same cats in the same boxes a week after the first set of trials. Thorndike wanted to know if the cats would immediately open the boxes the second week, or if they’d have to learn how to undo the latches all over again. He labeled the second week of time curves, “memory.”

Thorndike posited that if the cats had formed and held understandings of the latch in memory, they would have immediately unlatched the latch during the second week of trials. But they didn’t. Thorndike’s hand-drawn “time-curves” are difficult to follow; there is no key provided, and the labels are somewhat inconsistent. But, as Thorndike noted, the time-curves for the second week were similar in shape to the curves from the first, suggesting that the cats had learned all over again the second week using the trial-error method. However, the time-curves weren’t identical from first to second weeks: while similar in shape, they were shorter in duration, indicating that the cats did learn more quickly the second time around.

If the cats didn’t carry a visual image of the workings of the latch in mind from one week of trials to another (which would have resulted in immediate success the second week), what accounted for their quicker learning curve the second week? Thorndike explained that the first day of experiments had left a “vague body-sense impression” in the cats. This body-sense impression allowed the cats to start their instinctual escape behaviors in the right area of the box during the second week; instead of clawing and
struggling completely randomly, they clawed and struggled in the area of the latch. But they still had to rely on trial, error, and reinforcement to relearn the latch.

I have no desire to insist on the visual conceptual capacity of cats, or to contest the idea that sense-impressions play a part in human learning and behavior. But Thorndike’s insistence that cat-learning has nothing to do with memory or mind was part of a larger project to dismiss the very construct of human mind and consciousness. In fact, Thorndike’s work on this dissertation, along with Pavlov’s work with animals would lay the groundwork for Watson’s behaviorism. This project would have profound effects on learning and educational measurement theory.

In 1913, the same year that Watson’s Behaviorist Manifesto was published, Teacher’s College published the second volume of Thorndike’s psychology textbook written for teachers in training. After an introduction outlining the elementary laws of behavior he had developed from his work with cats (The Laws of Readiness, Exercise, and Effect), Thorndike begins the substance of the book with this sentence: “The complexities of human learning will in the end be best understood if at first we avoid them, examining rather the behavior of the lower animals ...” (6) Thorndike spends the rest of this chapter discussing his own experiments on chicks and cats, along with experiments on turtles conducted by none other than Robert Yerkes.

When Thorndike did finally turn the discussion to human learning, his consideration of such terms as “interest” or “will” were a bit more complex than Watson’s, who essentially preferred not to consider such terms at all. Perhaps one of the dangers of working in classrooms (all that mollycoddling!), was that students resisted being controlled more than rats, cats, and infants; the issue of interest for human learners, in
other words, was perhaps more difficult to ignore in the classroom than in the zoo, laboratory, or nursery. Despite his acknowledgement of them, Thorndike’s consideration of such terms as interest seems conflicted: he affirms their importance, but then limits our sense of them (and proposes to measure them) in such a way that they become almost unrecognizable:

“No one probably doubts that interest in the exercise of a function -- liking to add, or typewrite, or learn nonsense series, or whatever the work may be -- favors improvement at it. Such statements...appeal to our common sense as probably true, though they have not been fully verified by actually comparing learning with and learning without intrinsic interest in the matter learned*. (219)

First, notice the tasks themselves: adding, typewriting, learning a nonsense series. It is hard to imagine who might be interested in these tasks (after the novelty of a typewriter has worn off) in the absence of larger goals that these activities serve. It is rather like asking a man whether he is interested in blinking, and then, when he says that he is not, proclaiming that he is not interested in watching the movie that he is blinking to moisten his eyes in order to watch. Like Taylor’s assistant documenting the separate motions necessary for a man to move pig iron, Thorndike has broken up the tasks in which the learner might be interested. The concept of a learner’s interest deteriorates when the tasks themselves break down too far.

The concept of interest deteriorates further when Thorndike externalizes it. On one hand, Thorndike implies the need for more research “comparing learning with and learning without intrinsic interest in the matter learned.” Intrinsic and interest, all in the same sentence! On the other hand, in the footnote attached to this call for more research, notice how the phrase “intrinsic interest” shifts to “intrinsic incentives” (and how
“intrinsic incentives” are simply commands to do the task with a reason) without

cmment:

Wright ['06] has given evidence to show that the amount of physical work achieved
is very easily modified by slight intrinsic incentives, even when the workers are
adults all presumably trying to do their best. If such a one is told to work with the
dynamometer as hard as he can and as long as he can for no special end, he does not
do as well, other things being equal, as when some specific task is set...(219)

Work hard with a dynamometer, an instrument measuring engine output? Again, the
task is meaningless by itself: Work as hard as you can at blinking! When there is a
larger purpose for the work, it comes in the form of a task that is set for the worker, who
is expected to internalize it. There is little intrinsic anything measured in this study. In
the end, even with his limited acknowledgement of factors such as interest and attitude,
Thorndike’s version of learning is behaviorist: he seeks to explain all learning as the
result of “bonds” the learner makes that are either rewarded or punished.

The learning studies that Thorndike describes in his textbook reflect (and create) an
impoverished view of human learning. First, there’s the chapter’s worth of cat, chick,
and turtle studies. But most of the studies of human learning Thorndike cites later seem
to have been designed with the stopwatch in mind. For example, in one study, students
were asked to add numbers. Groups were given five-, ten-, or fifteen-minute breaks, then
asked to add again. The research question: Which break length led to greater
improvement? This was a study from which Thorndike claimed that the issue of interest
could be “inferred” (220-22).

Hearkening back to Thorndike’s declaration in his dissertation that you’ll never find
a better psychological subject than a hungry cat, Thorndike reports on one study about
the effects of hunger on learning and performance. Test subjects were asked to add
numbers on days during which they alternately ate nothing and ate normally. Predictably, test subjects’ performance on the days without food suffered. However, tests taken the day after the hungry days showed that improvements had been made, gains similar to those made on well-fed days. Thorndike concluded, “…it still appears that the permanent effect of the practice without food was as great as that with” (223).

Impoverished view of learning, indeed.

Thorndike never suggested that, if students learned as much on hungry days, we should only feed them every other day in the name of efficiency. But the study itself and Thorndike’s commentary on it exemplify the focus in the behaviorist model of learning on inputs and outputs -- a focus shared by the industrial model of efficiency. Both models attempted to control the behavior of the humans involved (workers/managers and students/teachers) by counting inputs, measuring outputs, timing the interval between the two, and then using rewards and punishments to achieve the fastest return on investments. With little concern for what happened in between the input and the output -- the experience of the individual whose output would be measured -- the behaviorist model of learning treated the learner as a machine on the production line, just as the industrial model of production treated the worker as a machine at the same time that it sought to replace him with machines.

Without the intervention of a psychological learning theory that insisted on attention to the experience of those involved in the learning, education for efficiency was free to become nothing but “materialistic, prosaic, and utilitarian.” Efficiency no longer existed to serve the human interests represented by education. Instead, the human
beings involved in education existed to serve the goal of efficiency, which, in turn, served the economic bottom line. The pupil -- and education itself -- had a new master.
CHAPTER IV

EDUCATION’S NEW MASTER

“Schools are tremendously inefficient, not because everyone does not work, but because there is no accepted ideal...”

--Harrington Emerson’s speech to the High School Teachers Association (HSTA) of New York City on December 9, 1911

It was a Saturday morning, but members of the High School Teachers Association (HSTA) of New York City weren’t sleeping in. Once a month from fall to spring, representatives from the nation’s largest school system gathered at the School of Commerce at 10 a.m. for a general meeting, followed by department meetings at 11:30. Here, teachers shared pedagogical ideas and engaged experts from nearby colleges and universities. Topics of discussion for department meetings between 1911-13 ranged from “A Latin Play Given By Children in England” to, “A New Use for an Old Wall Map” (No 40 Oct 1913).

Members of the HSTA biology department were intensely engaged in their profession. NYC biology teachers were well positioned to take a leading role on a national level; with access to university experts and contacts at local publishing powerhouses, they not only published, but were also at the forefront of controversial conversations such as sex education. In 1906, the biologist Maurice Bigelow from nearby Teachers College had argued that the student will “arrive at the truth” of “these deeper and more delicate questions” about sex if teachers would simply leave the student “alone with his thoughts” (qtd. in Pauly, 678). But by 1911, Bigelow had changed his position and joined Henry Linville, chair of the HSTA biology department and a teacher at Jamaica High School, in helping to host the nation’s first public discussion of sex education.
At 11:30 on December 9, 1911, biology teachers were scheduled to review four new textbooks “in relation to recent developments in biology teaching,” and one of these textbooks had been written by one of their own: George Hunter, a teacher at Clinton Dewitt High School in Manhattan. However, the 10 a.m. general meeting that preceded department meetings would bring HSTA biology teachers out of their professional comfort zones. Comfortable making and translating scientific knowledge for schoolchildren, they would nonetheless embrace the work set for them that morning: making and translating a powerful educational metaphor.

The featured speaker at December 9th’s general meeting wasn’t an educational expert. While he had taught at the university for six years, he made very few references to schools in his presentation that morning. Harrington Emerson was, after all, an efficiency engineer, an early “disciple” of Frederick Taylor’s. Two months earlier, the president of HSTA had “made a brief address, stating that the chief work of the year would be the study of High School Efficiency, urging all who were willing to serve on the committee...to make known their preferences” (HSTA Bulletin, Oct 21, 1911). Harrington Emerson, then, was the second presenter in a series of speakers arranged by the HSTA Efficiency Committee.

Emerson would not advocate the model of efficiency that Charles Eliot had described in 1908; he wouldn’t implore teachers to inspire and guide students’ intense and inventive thinking or suggest that they encourage students to develop a keen sense of smell. Nor would he insist, like Dewey, that the child’s experiences not be wasted. Instead, Emerson would present an industrial, profit-driven model of efficiency. The industrial terms and principles he outlined that morning would not only show up
repeatedly in future HSTA bulletins, but also guide educational reform efforts for the next 100 years.

Using examples almost entirely taken from factories, Emerson presented eleven principles of efficiency, all of which depended upon the leadership of a “strong able executive.” The first four principles were noble but rather vague: common sense, competent counsel, discipline, and fair dealing. The last seven were practical and immediately applicable: standard records, planning, standard conditions, standardized operations, standard instructions, standard schedules, and efficiency rewards.

According to HSTA secretary Agnes Carr, Emerson’s speech was a big hit:

That the subject was of vital interest was amply proved by the eager attention of the audience, and the evident reluctance with which, at last, with a rising vote of thanks to Mr. Emerson, the Association adjourned...(Mr. Emerson left for Bermuda that afternoon...) (HSTA Bulletin, No 40)

As five of Emerson’s principles indicated, the effect and means of efficiency was standardization: a standard ideal would direct all the organization’s energies through standardized records, conditions, operations, instructions, and schedules. And the coordination of all of these standardized efforts depended on the controlling influence of the “strong able executive” and the centerpiece of any behaviorist scheme: the efficiency reward.

THE STANDARDIZATION COMMITTEE: DEMOCRACY AND EFFICIENCY

AT ODDS

In coming months, HSTA members would grapple with exactly what these scientific principles of management would mean for education. The idea of a “strong able executive” proved a bit difficult for some to swallow. As one HSTA member
argued, the administrator, the teacher, and the student himself are, at various times, the executive. But everyone seemed to agree with Emerson that one cause of the school’s inefficiency was the lack of agreement on an accepted ideal. With the easily measureable goal of greater profit, a business had a distinct advantage over a school in terms of getting started on its drive towards ever-greater efficiency. To jump-start the movement towards more efficiency in the schools, Professor Hanus from Harvard was invited to define the aim of education during a general meeting in May of 1912. His first principle underlying “our whole public school activity” was this: to “train efficient citizens.”

As John Dewey had been arguing for some time, one of the central aims of education is democracy. And, fortunately or unfortunately (depending on your perspective) democracy itself isn’t efficient. The work of sharing, contesting, negotiating, and sometimes even transforming the different aims and experiences of individuals through community takes a lifetime. Knowing that democracy couldn’t be defended on the basis of efficiency, Dewey argued in 1936 that the only reasonable defense of democracy rested on the quality of experience it offers the most possible people:

Can we find any reason (to prefer democracy) that does not ultimately come down to the belief that democratic social arrangements promote a better quality of human experience, one which is more widely accessible and enjoyed...? Does not the principle of regard for individual freedom and for decency and kindliness of human relations come back in the end to the conviction that these things are tributary to a higher quality of experience on the part of the greatest number than are methods of repression and coercion or force? (Experience and Education 25-6)

A month after the HSTA president’s call for the study of efficiency, Dr. Stevens would acknowledge the incompatibility of democracy and efficiency during an HSTA discussion. As secretary Agnes Carr reported, ~
Dr. Stevens closed the discussion with a word of sympathy for democracy in the
schools, but with the added statement that in spite of the best efforts in that
direction, the constant tendency of a big organization is towards centralization.
(HSTA Bulletin, No 40)

**THE PUPIL’S NEW MASTER: THE CAPITALIST**

Even if the aim of education couldn’t be settled immediately, the standardization
could get underway as soon as possible, and other parts of the metaphor of the factory
could be applied. The biology department took this work seriously, holding a special
conference dedicated to the subject and reporting their conclusions to the entire
Association in 1912. On behalf of the biology teachers at Morris High School in the
Bronx, Mrs. Pingaey presented a paper on the application of principles of scientific
management to education. In it, she redefined the participants in school in terms of the
factory:

A. Purpose or object of “Scientific Management.”
   1. To increase the efficiency of the laborer, i.e., the pupil.
   2. To increase quality of the product, i.e. the pupil.
   (HSTA Bulletin No 47)

After learning that the pupil is synonymous with both the product and the laborer, Mrs.
Pingaey’s audience learned that the goal of the pupil’s education was no longer his own
edification, the good of his community, or the cause of democracy; the aim of his
education was now to “increase the amount of output and value to the capitalist.” Thus,
the redefinition of pupil to laborer or product forced a corresponding shift in the aim of
education. The pupil had a new master. It wasn’t knowledge and wisdom. It wasn’t
even the schoolmaster. The pupil now served the capitalist.

The public had a complicated relationship with the capitalist. Certainly, the
depression of the late 1800’s had intensified resentment of the Have’s by the Have-Not’s.
In 1896, William Jennings Bryan had launched his presidential campaign with the “Cross of Gold” speech in Chicago, in which he protested the hierarchy that put the capitalist at the top:

The man who is employed for wages is as much a business man as his employer...the farmer who goes forth in the morning and toils all day—who begins in the spring and toils all summer—and who by the application of brain and muscle to the natural resources of the country creates wealth, is as much a business man as the man who goes upon the board of trade and bets upon the price of grain; the miners who go down a thousand feet into the earth, or climb two thousand feet upon the cliffs, and bring forth from their hiding places the precious metals to be poured into the channels of trade are as much business men as the few financial magnates who, in a back room, corner the money of the world.

Despite Jennings’ protests, the capitalist was still a popular figure, viewed as a man who had made good by virtue of his hard work and intelligence. The word “virtue” here is apt: As Max Weber pointed out in 1905 (though his work wouldn’t be translated into English until several decades later), Benjamin Franklin himself had equated money with morality:

...money is the *prolific, generating nature*. Money can beget money, and its offspring can beget more, and so on. Five shillings turned is six, turned again is seven and threepence, and so on, till it becomes a hundred pounds. The more there is of it, the more it produces every turning, so that the profits rise quicker and quicker. He that kills a breeding sow, destroys all her offspring to the thousandth generation. He that murders a crown, destroys all that it might have produced, even scores of pounds. (Franklin, qtd. in Weber 13)

In Franklin’s formulation, the capitalist, who didn’t necessarily make a product with his own hands but used money to “beget” more money, was just as productive and therefore just as important to society as the skilled craftsman or farmer -- and perhaps even more so, because his industry generated even more wealth than the craftsman or farmer’s.

A year after Mrs. Pingaey identified the capitalist as education’s new master, Woodrow Wilson would give the capitalist an even greater position of prominence in
American life during a campaign speech: “...the masters of the government of the United States are the combined capitalists and manufacturers of the United States: (57). Still, educators such as John Dewey were leery of the effect of the industrial revolution on schools -- not so much because of the great gap between those who have a “corner on the money of the world” and everyone else, but because he believed that only by engaging in activities with social and personal meaning would the child learn. A speech Dewey gave in 1899 begins with his acknowledgment that the industrial revolution is the main social change with which schools must grapple. But then, he articulates the problem that this poses:

Plato somewhere speaks of the slave as one who in his actions does not express his own ideas, but those of some other man. It is our social problem now, even more urgent than in the time of Plato, that method, purpose, understanding, shall exist in the consciousness of the one who does the work, that his activity shall have meaning to himself...How many of the employees are today mere appendages to the machines which they operate! This may be due in part to the machine itself, or to the regime which lays so much stress upon the products of the machine...

To the extent that workers in the industrial model were “slaves” in a system that separated work from meaning and social activity, Dewey believed that the industrial model was inappropriate for schoolsxiv.

While many balked at the implications of equating the capitalist’s concerns with the educator’s, the capitalist was still a popular enough figure (or, perhaps better said, a powerful enough figure) for the analogy to have real consequences -- and sticking power -- in education.

In telling the story of the how the industrial efficiency model made its way into the HSTA, I don’t mean to suggest that this was the first or even the most important entrance -- although there is some reason to think that the HSTA had some national
influence, positioned close to other centers of educational influence such as Teachers College. Instead, I want to show particular actors -- the HSTA as an organization, individuals such as Mrs. Pingaey and Mr. Harrington, and ideas -- coming into association with each other for the first time. My goal in telling this story is to remind us that the economic metaphor that shapes education today is, in fact, a social network that human actors play their part in maintaining and strengthening. Of course, Mrs. Pingaey herself wasn’t responsible for the translation of industrial business concepts into education; it was an association that many educators were happy to make. As Callahan posits, the concepts of a high-status group are often sought out by the low status group.

While the enthusiastic acceptance of this metaphor is understandable, here’s what putting the capitalist metaphorically in charge of education meant for schools. The capitalist doesn’t much care how the work gets done. (Often, he doesn’t much care what the work is, as long as it profits him.) Nor need he care about the experience of those doing the work. He need only care that the work gets done, as efficiently as possible. (“...it still appears that the permanent effect of the practice without food was as great as that with...”) And, with a vested interest in the factory’s progress but little to do with its day-to-day operations, the capitalist demands frequent reports on profits. Of course, that’s where the application of the business model to schools got tricky: a representative case of the product (pupils? their learning?) couldn’t very well be shipped to the capitalist along with a report of the number of such cases produced and sold at what cost each month.
That's where educational measurement -- and, thus, Thorndike and Yerkes -- came in.

IDIOTS, IMBECILES, AND MORONS: THE PSYCHOLOGICAL SORTING MACHINE

With their experience in World War I measuring the minds of the almost two million recruits, Yerkes and Thorndike were uniquely poised to help educators generate the reports they now owed the capitalist-taxpayer. Just as Watson was eager to apply behaviorism to parenting and advertising, Yerkes had been eager to apply behaviorism to the war effort. As president of the American Psychological Association (APA) at the start of WWI, Yerkes had formed twelve committees to explore psychological applications to the war effort. He was particularly eager to apply intelligence testing -- a relatively new subfield of psychology. He named himself the head of the committee on intelligence testing, convincing the army to let him design and administer a test to reject recruits whose intelligence was low and identify recruits who should be considered for Officer Training School.

The testing committee created a test and ran a trial in two months, a stunning feat, given that standardized testing was still in its infancy, and large scale standardized testing hadn’t yet been born. Two committee members, Lewis Terman and Henry Goddard, had some previous experience with standardized testing procedures. A decade earlier, French psychologists Alfred Binet and Theodore Simon had developed a test to identify schoolchildren as advanced on the one hand, or “idiot, imbecile, and moron” (266) on the other. Terman and Goddard had both been involved in importing the Simon-Binet test
from France; after the war, Goddard would oversee an English translation of Simon and
Binet’s work.

Several factors should have made the Simon-Binet test more difficult to employ
for military use than Yerkes’ committee found it. It is true that Simon and Binet’s work
was driven by many of the values that would prove central to the large scale standardized
testing movement. As they explained: “We have nothing to do with [the child’s] past
history or with his future…” (37); and, “We have aimed to make all our tests simple,
rapid, convenient, precise…” (41). Still, the Simon-Binet test was created to be
administered individually by a professional whose pleasant interaction with the child and
whose interpretation of each answer was integral to the results.

In fact, Simon and Binet had warned those administering the test that providing a
simple rating of each answer was worse than insufficient: “this must not be tolerated; it
encourages negligence and even fraud” (298). Instead, the ratings should be
accompanied by copious notes, and only “experience” would tell the professional which
details were important to include, so that “another experimenter may be able to judge the
results for himself” (296). And, in a lengthy and relatively thoughtful exploration of a
teacher’s judgment, Simon and Binet emphasized the “incontestable superiority of
observation over the test” (311), noting that,

…[the teacher’s view of the child] presents one very great advantage. It is based
on long observations…if the facts observed have not each a great value, on the
other hand they are numerous, diverse, and when needful they correct one
another. Herein lies the incontestable superiority of observation over the test.
(311)
And yet, Simon and Binet dispensed with the teachers’ judgment with the following complaints: “But on the other hand, what indecision in the observation! What errors! One rarely arrives at certainty and never at a measure. So much for observations” (311). Interested as they were in making quick judgments about millions of men, “certainty” and “measures” were among the top priorities for Yerkes and the members of his testing committee. As Yerkes and Yokum recounted shortly after the war ended:

...the test should be so arranged that the scoring could be done rapidly...also, this arrangement should be so simple that relatively inexpert assistance could be used in scoring the large number of papers. (3)

Toward this end, Yerkes and his committee adopted Arthur Otis’ new multiple-choice format that didn’t rely on interaction or human judgment or interpretation during administration or scoring. The 50-minute tests could be given to five hundred recruits at a time.

The IBM 805 (the precursor to the Scan Tron that looked like an old-fashioned sewing machine table) wouldn’t be available until 1936, when the New York Regents would pilot it in city schools. So, during World War I, the AAT had to be scored by hand. But scoring a multiple-choice test was still quicker than scoring the Simon-Binet test, which had required an expert to make a subjective judgment for each answer of every test administered. Yerkes reported that “small staffs” were sometimes able to score 2,000 tests a day, using “stencils” of the correct answers placed over each recruit’s score sheet.

To properly place recruits who might be illiterate but nonetheless intelligent, the committee created a visual version of the AAT -- the Army Beta Test (ABT). After conducting a trial of the ABT and the AAT with groups of military personnel and
schoolchildren, the committee sent the results to Edward Thorndike and a “statistical team” at Columbia for a check of “validity, reliability, and significance” (Yerkes and Yokum 5). As a direct result of Yerkes’ quick work with the psychological testing committee, 7,800 recruits were recommended for discharge during World War I, 10,014 men were recommended for labor battalions, and 41,000 became officers.

But the AAT and the ABT weren’t Yerkes’ only contributions to the war effort on behalf of psychology. Yerkes also headed a committee on the “control of morale” in the military, framing the problem and the solution in behaviorist terms. In addition, he created and oversaw the administration of “rating scales” that were used for promotion within the military. Yerkes thought quite highly of all these efforts, crediting his psychological sorting mechanisms with helping to win the war (Gould).

“SO MUCH FOR OBSERVATIONS”: FOUR LETTERS IN WHICH LIEUTENANT FURGESON OBJECTS TO HIS IMPERSONAL RATING

Not everyone was as enthusiastic about these tests and scales as Yerkes. Just one year before sending AAT “blanks” to the Gilman School at Watson’s request, Yerkes received a series of increasingly urgent letters from George O. Furgeson Jr, 1st Lieutenant and Chief Psychological Examiner. In the first letter, dated November 8, 1918, Furgeson complained about the manner in which he had been rated for (and denied) promotion. In the last letter sent a little more than a month later, Furgeson requested discharge -- frustrated that his appeals had been unsatisfactorily resolved.

There were other letters in Yerkes’ collection protesting decisions made on the basis of his tests and rating scales. But, perhaps because he was a psychological
examiner himself, Furgeson’s complaints are particularly articulate and poignant. His first criticism of the rating scale, in a letter written to Yerkes on November 8, 1918, is humorously indirect:

When, on Oct. 10, in response to your request, I asked the Camp Surgeon to rate me, he remarked that he was afraid he was one of the older officers who thought the rating scale worthless. In fact, he used profanity in speaking of it.

Furgeson’s direct criticisms of the use of the rating scale involved how impersonally they were applied. Later in this letter, Furgeson told Yerkes what had happened when he met Major Todd, the new officer responsible for rating him:

...this officer, an entire stranger to me and to psychology, had visited me in my office for less than ten minutes, engaged me in general conversation, complimented me on the condition of my newly oiled floors, and departed without stating his business or learning anything of mine. I have just now looked him up in order to ask him to learn something of our work and to tell him the consequences of his rating of me. To my surprise, I found that he had rated me without making any inquiries of people who know me or the psychological work here...

The very next day, Furgeson wrote Yerkes again, to tell him the results of his “re-rating” by the local medical personnel board, which included Major Todd, a fan of Furgeson’s newly oiled floors. Furgeson protested again:

All three officers are entirely unknown to me except for the brief visits paid by them, and their positions are such that they have not come into contact with my work. Major Owensby...did not favor the psychological work. His only professional question to me was, “How can a man have manic-depressive insanity and still get a high score on your tests?

On the same day, Furgeson wrote a memorandum to the Camp Surgeon objecting to his rating, reporting,

[The Camp Surgeon] was quite agreeable personally, but informed me that since the board’s report had been made, he could not change it, although he thought the use of the rating scale an absurdity...
Finally, Furgeson wrote to Yerkes on December 23, 1918, requesting discharge for the second time,

...my rank is too low for successful work in the highly cooperative position to which I am assigned... I believe the Camp Surgeon has unjustly deprived me of promotion...

I could find nothing in Yerkes’ papers to indicate the conclusion to poor Furgeson’s case.

Furgeson’s objection -- *Important decisions are being made about my future by men who do not know me!* -- was impossible for Yerkes’ psychological sorting machine to address satisfactorily. The very reason for the creation of that machine, and its major selling point, was that it didn’t require much in the way of human interaction or judgment.

Enabled by behaviorism, the industrial revolution had been eroding the value placed on human interaction and experience for years. The Great War would strike another blow against human experience, judgment, and interaction; there simply wasn’t time to take it into consideration! Once the crisis was over, however, the industrial testing complex wasn’t about to put humans back in charge.

**“GREAT WILL BE OUR GOOD FORTUNE”: IN WHICH WARTIME LESSONS IN HUMAN ENGINEERING ARE APPLIED TO PEACETIME CIVIL INSTITUTIONS**

Despite Lieutenant Furgeson’s objections to the impersonal rating that denied him promotion, Yerkes was quite sure that his impersonal testing format would be useful in every conceivable setting. He expressed no hesitation in 1919 about giving AAT “blanks” to the Gilman School, where Watson’s son was a student. That same year, he proclaimed:
Great will be our good fortune if the lesson in human engineering which the war has taught is carried over directly and effectively into our civil institutions and activities. (qtd. in Yerkes and Yoakum xiii)

A year later, Yerkes and Clarence Yoakum published Army Mental Tests with the authorization of the War Department. Apparently, the demand for their tests outside the military was great:

During the past few months the Office of the Surgeon General of the Army and the National Research Council have been besieged with requests for information concerning the methods of psychological examining and for the printed materials used in the United States Army. To meet this demand it has seemed advisable to prepare this little book, which, in addition to the ‘Examiner’s guide,’ presents information concerning the results of psychological examining in the Army and indications of the possible uses of similar methods in education and industry. (iii)

What did these military psychological examiners imagine that this test, designed to sort adults into officer training school or labor battalions, could offer children and educators?

The first proposed application of the tests to education had nothing to do with helping children to learn: it had to do with excluding certain high school graduates from college. Of course, Yerkes, Yoakum, and the rest of the testing crew wouldn’t quite put it that way. They preferred to focus on how grateful students would be to be placed, and even excluded, by the quick magic of testing. Consider how Lt. Col. Bingham -- one of the seven original members of Yerkes’ psychological testing committee -- put the matter for the Association of College Registrars in 1919:

Is it not possible toexclude from college a larger proportion of the students who now come to college and fail, who leave with the brand of failure upon them, having wasted their own time and their fathers' money? (qtd. in Yerkes and Yoakum 176)

Almost a century later, despite Bingham’s optimistic phrasing, it is difficult to imagine a student expressing gratitude to the SAT for preventing the waste of his time and his father’s money on being denied admission to the college of his choice.
To understand the vision of society that psychological testing proponents were pushing, it is helpful to back up and consider two smaller but related questions. First, what did psychological testing experts imagine that their tests could offer the work of learning in schools? Secondly, what was the relationship between Watson’s rejection of mind and Yerkes’ attempt to measure it?

Lewis Terman would call upon the industrial metaphor to explain his answer to the first question about the benefits of testing to schools: “Without knowledge of the quality of its raw material a school has to work more or less in the dark” (Terman, qtd. in Yerkes and Yoakum 160). Eight years earlier, Mrs. Pingaey had made a similar metaphoric leap: she, too, had referred to pupils as “raw material” in her speech on behalf of the biology teachers at Morris High School. But she had done so in the course of identifying a “difficulty in the way of making exact applications of scientific principles,” pointing out that, when it comes to schools, “…poor raw material (the student) cannot be exchanged for good…” (HSTA Bulletin 47).

It is impossible ascertain Mrs. Pingaey’s attitude, here: had she offered this line with a wink that served to undermine the analogy without taking an overt stand against it? I’d like to think so, but nothing else in her speech outline suggested that she described this difficulty with the analogy in order to erode it. More likely, it had been an effort to establish solidarity with her fellow teachers, all of whom, at various times, must have
wished they could exchange little Johnny (who, half a century later, still wouldn’t be able
to read and write!) for someone more imbued with scholastic ambition and aptitude.
Perhaps this line had even elicited a laugh.

But Terman wasn’t noting any difficulties in applying the industrial model to
schools. In his view, the very act of sorting students with intelligence tests -- labeling
some quality raw material and some poor raw material -- would enlighten schools.
Before we can explore how Terman and other psychological testing enthusiasts thought
this sorting could benefit schools, we should pause to note how Terman’s invocation of
the industrial model here (pupils as raw materials) helps explain the relationship between
very strange bedfellows -- Watson’s behaviorism and Yerkes’ intelligence tests. The
relationship between Watson’s rejection of mind and Yerkes’ measurements of it, in fact,
helps us get clear about how Watson and Yerkes joined forces by using these tests to
restructure society and establish a new ruling class of technocratic elites.

Despite Watson and Yerkes’ longstanding friendship and professional
collaborations, we might expect Watson to have objected to Yerkes’ tests. After all,
Watson rejected the very concept of mind, but Yerkes’ tests purported to measure it.
Furthermore, Yerkes viewed intelligence as a fixed genetic inheritance, but Watson
considered the infant a blank slate, conditioned by environment. And while Yerkes used
the results of the AAT to validate the eugenicist viewpoint, Watson was theoretically
opposed to eugenics. Still, Watson never objected to intelligence tests, except when
Yerkes administered them to primates. When it came to their human applications,
Watson actually ordered them up for his son’s school.
The tie binding together Yerkes’ intelligence tests and Watson’s behaviorism had at least two strands. The first was Watson and Yerkes’ shared vision of an industrial model of society, with psychologists in charge as “human engineers.” Toward this end, Yerkes wanted to use mental tests to make high-stakes, large-scale decisions about every aspect of a person’s life. Intelligence tests, then, were part of a larger project to efficiently sort people into their proper place, with as little human judgment and interaction as possible. It was to be a modern-day Republic, with psychologists scientifically ascertaining whether a person was gold, silver, or bronze.

Yerkes’ pronouncement in 1919 that his wartime mental sorting machine was an experiment in “human engineering” mirrored the language of one of Watson’s mentors in Chicago, Jacques Loeb. Loeb, a biologist, believed that a science was not a science if it simply attempted to describe nature. Instead, he believed that sciences needed to control nature much as engineers controlled physical materials. Thus, Loeb’s professional ambition as a biologist was to control organisms, and the very beginnings of life itself. While working with Watson in Chicago, he manipulated sea urchin eggs in his attempt to create life without spermatozoa. Later, he would argue for a view of organisms as machines in his 1913, article, “Mechanistic Conception of Life.” Yerkes’ view that psychological testing was human engineering, then, appealed to Watson’s sense (inspired by Loeb) that the science of psychology should be used not just to describe man’s behavior, but also to control it -- along with society itself.

The second strand holding together Yerkes’ mental tests with Watson’s rejection of mind was the dramatic way in which World War I advanced the behaviorist cause. Even though Watson’s “Behaviorist Manifesto” in 1913 was retroactively declared the
birth of behaviorism, the truth is that the article didn’t generate much buzz when it was published (Wozniak). It would take Yerkes’ applications of psychology to the war effort several years later to draw widespread attention to the behaviorists’ efforts. In other words, Watson might have opposed mental testing. But Yerkes was using mental testing as part of a larger project with which Watson aligned himself. Perhaps Watson found that, if it alienates your allies, theoretical purity is overrated.

Mental tests had been the gears that had made Yerkes’ military sorting machine go round. But their use in the military had been nothing more than a test run, an experiment proving what was possible on a larger scale. In order to carry out their larger purpose, these tests needed to be implemented from the beginning -- in schools. College admissions were the first step toward that end, but they needed to start earlier, to transform schools into an educational bureaucracy. In this bureaucracy, interactions between educators and students would be governed by procedures worked out in advance, procedures that could only be efficiently implemented only if children were sorted into the proper groups. According to Yerkes,

It therefore is proposed that children should be classified in accordance with mental ability either as they enter school or shortly thereafter and that mental ability should thereafter be taken into account in connection with their educational treatment. (191)

Testing would facilitate the efficient sorting of children into homogeneous groups, each of which would then receive its own predetermined educational treatment. These homogenous groups would not resemble the democracy into which they would later emerge -- that is, not until the factory model would eventually transform society itself into homogenous groupings, a goal eagerly pursued by the eugenicsists.

The shape and implications of this vision for education and society become clearer
in contrast with Dewey’s vision. For Dewey, democracy, diversity, education, and experience were inextricably linked. The goal of a democratic social arrangement was the greatest quality of experience for the greatest number of citizens. And education served this greater goal. Students were formed, and learning happened, within learning-communities that were as diverse as the democracy into which they would later emerge as well-educated citizens, ready to find and make meaning in their work and lives. And the teacher’s role in this process was to structure the learning community in ways that would lead to “educative experiences” for students -- experiences that not only resulted in learning in the moment, but that also made future learning more likely.

We can imagine, then, Dewey’s response to the “hunger” study cited in Thorndike’s 1913 psychology textbook for teachers. Test subjects -- hungry but busy adding numbers without any purpose -- might indeed have shown improvement when their adding skills were measured the following day. But Dewey would have asked this question: Would this experience of adding without purpose or food be likely to “open up” the test subjects to future adding experiences, to future mathematical learning? If the answer was ‘no,’ Dewey would have labeled this a miseducative experience and banned it and other experiments like it from the “laboratory” school he helped to run at the University of Chicago.

Society and education in the behaviorist scheme had drastically different goals with accordingly different means. While key terms in Dewey’s scheme were *experience*, *democracy*, *diverse communities*, and *interaction*, key terms in the behaviorist scheme were *testing*, *control*, *homogenous groups*, and *educational treatments*. Instead of educative interactions, the behaviorist version of education relied on one-way actions:
“educational treatments” applied after educational measurements had been administered (with as little interaction between the tester and the tested as possible). Yerkes and Terman may have had the best of intentions: in their own mental schema, education can occur only when students are put in the right place to receive the right treatments -- and the “right place” is always with students who are just like you: idiots in one room, imbeciles in another, and future leaders of America in the next. And to Yerkes, Brigham, and other eugenicists, these labels just happened to coincide with particular ethnicities.

True to their ambition to engineer every aspect of society, Terman, Brigham, and Yerkes weren’t content to limit the sorting function of tests to schools, college admissions, or the military. They believed that their tests could be used to place workers within a specific business, divining a person’s true place in the organizational hierarchy. Here’s Yerkes on the use of tests to sort workers within a particular industry:

Before the war mental engineering was a dream; to-day it exists, and its effective development is amply assured. For the present, at least, it is probable that if three grades of intellect were distinguished in industry, as has been suggested for the school, a very great gain would be made in degree of fitness of the individual for his task, and in his resulting contentment and efficiency.

Note the echoes of Plato’s Republic -- no more or less than three grades of intellect within a given industry, contributing to the contentment of the individual (who has found his place) and the common good (efficiency) within the industry. But the dream went further: the applications of mental testing weren’t just useful for sorting employees within a particular business (laborers here, managers there). Mental testing could be used
in an even more fundamental way: to sort everyone into the appropriate profession in the first place.

Again, the basis for these determinations would be scientific: Yerkes correlated the stated professions of almost 20,000 men with their AAT scores. He reported that barbers had scored an average of C- on the AAT, while accountants had scored in the B range. The only profession whose members scored an A average? The very profession in whose image Loeb, Watson, and Yerkes sought to remake their own: Engineers (Yerkes, qtd. in Yerkes and Yokum 198). In other words, the educational sorting machine would feed an industrial societal sorting machine. And psychologists would be the human engineers in charge of it all.

REPORTING TO STAKEHOLDERS WITH THORNDIKE’S RATING SCALES

With potentially contentious decisions about opportunity and resource allocation handed over to tests that rendered human judgment, interpretation, and democratic deliberation unnecessary, society would function like a Platonic corporation that existed for its ever increasing profit margins, obsessed with its own bottom line, the ever-important and ever-increasing GNP. To make this corporation work as efficiently as possible, mental tests would place each member of society in the appropriate track from elementary school through college, with the dream of vocational placements on the horizon. But mental tests only measured the quality of raw material for the purposes of sorting it. Schools still needed tests that could be used to report progress to its stakeholders -- taxing educational capitalists, otherwise known as the public.
Thorndike was happy to help.

Thorndike was obsessed with rating scales and measurements. He created, published, and distributed them for handwriting, composition, geography, arithmetic, spelling, reasoning, and various facets of reading. Thorndike begins *Measurement of Ability in Reading* by outlining the pressing need for these scales:

The work is being continued and a much more satisfactory report will be made later; but the need for even rough means of measuring school achievement of reading is so pressing, and so many requests have come for information about these new scales, that I am giving now this preliminary account of the work, so far as it has gone. It is obvious that educational science and educational practice alike need more objective, more accurate, and more convenient measures...Any progress toward measuring how well a child can read with something of the objectivity, precision, commensurability, and convenience which characterize our measurements of how tall he is, how much he can lift with his back or squeeze with his hand, or how acute his vision is, would be of great help in grading, promoting, testing the value of methods of teaching and in every other case where we need to know ourselves and to inform others how well an individual or a class or a school population can read. (1-2)

Two big assumptions in two short paragraphs: educational measurement would someday be as simple and accurate as measuring how tall a child is; and any measurement scale, however rough and preliminary, is better than no scale at all.

Thorndike's main concern was that the measurement be easy to administer and to score. Perhaps this explains the following test of the ability to comprehend a sentence:

E. In these two lines draw a line under every 5 that comes just after a 2, unless the 2 comes just after a 9. If that is the case, draw a line under the next figure after the 5:

5 3 6 2 5 4 1 7 4 2 5 7 6 5 4 9 2 5 3 8 6 1 2 5
4 7 3 5 2 3 9 2 5 8 4 7 9 2 5 6 1 2 5 7 4 8 5 6

This test may, indeed, measure something: the student's tolerance for following directions with no perceived purpose, for instance. But it is difficult to follow Thorndike
in asserting the benefits of such a “reading” scale even as he acknowledges its limitations:

Future work will improve the scale greatly... but even as it stands, the resulting comparison of two classes will be much fairer and more exact than probably even the most capable supervisor of reading would arrive at by ten hours of oral questioning. And the results will be readily comparable with thousands of others obtained with other classes by other supervisors, and will be at once understood by anybody who knows the scale -- a most desirable feature, which no ordinary school examination possesses. Consequently... its use even now will do much more good than harm...(57)

Nowhere in this volume did Thorndike explain how these measurements would help instructors help students learn. In fact, he asserted that the results are most often more accurate for large groups of students than for individuals, especially in the scale of visual vocabulary. Their chief value, then, was less in helping individual students learn and more in comparing groups of students with each other, the basis for reports to educational stakeholders.

Teachers have probably always looked for evidence of student learning -- signs that their instruction has some kind of effect. But it is important to note that educational measurement was tied up in a larger project to “increase the amount of output and the value to the capitalist” by providing the basis for reports designed to inform stockholders about how their money was being used.

But what would happen when an educational crisis was declared on the basis of these reports? What should be done with a lax industry not meeting its quality control quotas? And here, we circle back to Watson, who offered a simple formula for controlling human behavior: reward and punish. Take Watson’s behaviorism, throw in Thorndike and Yerkes’ educational measurement, fold them into Taylor’s industrial management process, and you’ve just cooked up the educational accountability
movement of the twenty-first century. In the educational accountability structure, schools are a business answerable to stakeholders through reports generated by educational measurement. High performing schools are given Emerson's "efficiency reward," declared to be meeting AYP, and left alone to continually improve their test scores on their own. Low performing schools (or, in the case of Linda Rief's middle school, high performing schools that hit the AYP ceiling) are punished through threats and sanctions, including state takeover. This system of accountability, and the reports on which they depend, have turned the complex intellectual, social, emotional, and relational work of education into a constant ticker-tape of test scores scrolling at the bottom of the national consciousness, as stakeholders (and economists such as Milton Friedman) wonder whether public education is worth the investment and whether it might be more profitable to simply privatize education once and for all.
CHAPTER V

CARL IN THE BEHAVIORIST MATRIX: MEASURE, REWARD, AND PUNISH

"Mom, how many errors did I make?"
-- Ben, age 5, reading a book aloud at home for the first time.

It’s almost time to return to Amy and Carl at the alternative middle school in Michigan, Amy wishing she could send the MEAP manual hurtling through the air to land next to Carl’s crumpled up essay, Carl himself somewhere alongside the highway, walking anywhere as long as it takes him away from that test. I’ve represented this moment as a breakdown -- precipitated by an assessment tool that dismisses mind, consciousness, meaning, and experience. I’ve traced the origin of this tool’s design back to the emergence of the Holy Behaviorist Trinity of Watson, Thorndike, and Yerkes. But, from Carl’s point of view, this moment is hardly a breakdown -- it is just one of many such moments that led slowly and un-dramatically to his placement at the alternative school.

It wouldn’t be much of a stretch to imagine the ways in which Carl might have become well acquainted with Watson, Yerkes, Thorndike, and Taylor -- though he never knew their names -- in the years leading up to this moment. We can’t ask Carl about it; he took his own life when he was in high school. But after teaching for seven years in Carl’s district and ushering my two sons through three of its elementary schools, I can offer informed speculation. There will be nothing extraordinary about what I’ll describe. Similar stories are lived out in ordinary schools across the country every day, because behaviorism has shaped our sense of the ordinary when it comes to all aspects of schooling.
In kindergarten, Carl would have met Thorndike, Yerkes, through DIBELS, a reading test reminiscent of Thorndike’s “nonsense series,” Yerkes’ commitment to “objective” test conditions, and Taylor’s fondness for the stopwatch and determination to fracture every action into itty-bitty parts. The Dynamic Indicators of Basic English Literacy Series (DIBELS) includes various so-called diagnostic “probes.” From K-5th grade in this district, teachers were required to administer the DIBELS fluency probe – a requirement that took so much time that high school students were often imported to administer this probe to students, one by one, in the school hallways. During the DIBELS fluency probe, our five year old Carl would have been told to read a passage of nonsense syllables out loud as quickly as he could -- sim, lut, etc. -- while the examiner recorded his “errors” on a clipboard, starting and stopping a stopwatch at one minute. The number of errors Carl made would have been subtracted from the number of syllables he attempted. This number would have been divided by the total number of nonsense syllables possible, and, voila! Carl’s reading fluency rate would have been determined, recorded, and then retested the following month.

Any experienced reader might pause here to question: isn’t there a difference between the rate at which you can read a meaningful sentence and the rate at which you can read nonsense syllables? And doesn’t fluency imply much more than speed -- including expression, for which it is necessary to understand the meaning of the text? How can reading fluency be represented by the rate at which you can pronounce nonsense-syllables?

But the concern represented in the DIBELS fluency probe is the same concern that Yerkes and Thorndike had established in creating and administering the AAT:
fairness = objectivity = standardization. In Thorndike and Yerkes’ formulation, this concern was limited to the actual room, materials, and time allotted for the test. The playing field wouldn’t be level unless each man had the same amount of time in the same kind of room to complete the same test. In later years, this concern would broaden. The test questions themselves should maintain a level playing field, without disadvantaging test-takers on the basis of background knowledge. For example, reading comprehension passages including mention of silos and combines might not be fair for urban kids, while passages including mention of subways and high-rises might not be fair for rural kids.

But by the time researchers at the University of Portland created the DIBELS battery of probes, this concern had extended even further. Children’s vocabularies differ not only by the number of words they know, but also by the actual words they know. Especially at the earliest stages of reading, it is easier for children to decode words that they’ve heard before than to decode words they haven’t. If test developers used real words in the fluency probe, it would disadvantage children who hadn’t ever come into contact with those words before.

The standardized solution to a problem of standardization? Don’t use actual words. Use nonsense syllables that all children will have had an equal chance of not knowing. Ironically, I share this concern for the effect of background knowledge on the fairness of standardized tests. But when that concern is not used to dismantle the illusion of objectivity in standardized testing, the solutions become... well, nonsensical.

An experienced reader might look at the DIBELS fluency probe and dismiss it out of hand, thus diminishing any negative effects on her conceptualization of what good reading is: this doesn’t test my reading fluency! But think about the experience of the
small child, not yet an expert, eager to please. The child has spent his life inferring what is important to the adults in his life; his survival has depended on it! And it isn’t difficult for him to infer two things so far about what’s important to the adults in school: tests are important, and what is in this reading test is what adults think good readers do. Good readers read fast. Good readers read without any errors. And, most devastatingly, good readers don’t bother with meaning.

Some will still protest: it’s just a test! Kids know what reading is! But consider the experience of my youngest son, Ben, who spent his first five years immersed in literacy at home -- not just in the form of the bedtime story, but, most importantly, as he watched his parents and his older brother connect to others and make sense of the ideas shaping their lives through reading and writing. Of anyone, Ben should have been immune from any reductionist message about reading.

But when Ben sat down next to me in first grade to read an entire book out loud to me for the first time, we didn’t have the Literacy Hallmark Moment I’d anticipated. We didn’t talk about the connections between himself and the character in the book (they both had a pet bird and the name, Ben!). We didn’t go look for another book to read together. We didn’t even pause to celebrate his accomplishment. Instead, Ben closed the book and asked, “Mom, how many errors did I make?”

Errors? I’d never used that term with my son. Mistakes, maybe. But I never say the word error -- mostly because I can’t. The r’s disappear into the back of my throat, like swallowing a growl. Ben hadn’t learned that word from me. He’d learned it from watching the teacher subtract his errors while “reading” nonsense syllables during the
DIBELS reading fluency probe that his teachers were required to administer once a month from K-5.

Teachers themselves aren’t immune from the messages this probe sends about what good readers do -- and what good reading teachers ask their students to do. I’ve spoken with teachers all over the country who report that they regularly have students “practice” reading nonsense syllables in preparation for DIBELS. Since scores on these assessments are often posted in the teachers’ lounge and used to evaluate teachers, teachers feel forced into this particularly perverted form of reading test prep.

Despite the relatively rich picture of reading I’d tried to give Ben since birth at home, he didn’t enjoy reading or see himself as a reader until we moved out of the state and his fifth grade teacher let students read what they liked and refused to administer DIBELS. But this, apparently, was an aberration. When Ben’s teacher retired unexpectedly mid-year, I told the principal what a positive experience Ben had had in her classroom. He leaned in to confide what a problem this teacher was: she never followed the program! It wouldn’t have made any difference to point out that the program itself had never served my son’s learning. The program hadn’t served Carl’s learning, either. More vulnerable to reductionist messages about the nature of reading than my son, Carl’s view of literacy itself would have been shaped by Taylor’s stopwatch and Thorndike’s reading scales and “nonsense series.”

Carl’s introduction to Watson would have taken place through The Matrix. The Matrix is the centerpiece of the elementary schools’ behavior management program, Positive Behavior Instructional Support (PBIS). Here’s how PBIS begins: teachers walk through all areas of the school, discussing the behaviors that they want to see students
exhibit in each area. These behaviors are listed in a school-wide Behavior Matrix and taught to the students. A system of rewards and punishments is then implemented to enforce it.

A Sample PBIS Matrix

<table>
<thead>
<tr>
<th>In the bathroom...</th>
<th>In the hallway between classes...</th>
<th>Etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students will...</td>
<td>Students will...</td>
<td></td>
</tr>
<tr>
<td>Flush the toilet</td>
<td>Stand in line quietly</td>
<td></td>
</tr>
<tr>
<td>Wash hands with soap—without spilling soap on the counter</td>
<td>Place pointer finger of right hand on lips (&quot;Shhhh...&quot;)</td>
<td></td>
</tr>
<tr>
<td>Dry hands—without leaving paper towels on the floor</td>
<td>Place left palm against left side</td>
<td></td>
</tr>
<tr>
<td>Exit bathroom quickly</td>
<td>Walk only when given teacher’s signal</td>
<td></td>
</tr>
</tbody>
</table>

A professor of literacy who consults with schools on the Navajo nation watched teachers in a Bureau of Indian Affairs (BIA) boarding school (yes, they still exist!) construct their PBIS Matrix in 2010. With tears in her eyes, she told me what teachers were writing into the “hallway” column of the Matrix: In the hallway of this BIA school, students -- mostly Navajo children -- were to stand and walk with their arms crossed tightly against their chests. The symbolism of this posture, conjuring up the dark history of these boarding schools where native children were separated from their families to strip them of their language and culture, was apparently lost on these teachers. But that’s no real surprise. Watson’s behaviorism doesn’t recognize the layers of meaning in a symbol. All
that matters in the Behaviorist Matrix is identifying, measuring, and then rewarding or punishing behavior.

In Carl's school, students wouldn't have been required to walk with arms crossed over their chests. But the behaviors written into The Matrix at Carl's school for the bathroom (flush, wash, dry, exit) and the hallway (right finger to lips, left palm against leg) would have been monitored. In fourth grade, my oldest son, Isaiah, proudly announced he was a school safety and got to wear a cool fluorescent orange sash. Back in the day (or, at least, back in my day), the school safety helped little kids cross the street before and after school. I assumed that's what he would be doing. But, no. After several weeks of unenthusiastic responses to my inquiries about safety duty, my son told me what he was doing -- enforcing the Matrix.

Stationed outside the bathroom during lunch and break, Isaiah asked any child entering the bathroom to sign the sheet on his clipboard -- name and time. When the child was finished, Isaiah went into the bathroom after him to check that the toilet had been flushed, that no soap was on the counter, and that no paper towels were on the floor. He was to report children whose bathroom behavior didn't measure up to the powers-that-be, who were in charge of doling out (or taking away) "I caught you being good!" points.

Children who had accumulated the requisite "I caught you being good!" points at the end of the year were allowed to go to a carnival on the playground, complete with ice cream, games, and big, pillowy, moon jump. Children without the requisite number of points? Well, they couldn't go, a fact that teachers and administrators reminded the children of constantly throughout the year.
I was briefly trained in PBIS before the high school teachers rejected the program. During my training, I protested that the program seemed behaviorist. The trainer begged to differ -- “This is POSITIVE behavior support.” To him, the program’s emphasis on rewards rather than punishments exempted it from the charge of behaviorism.

Apparently, no one at Corporate PBIS Headquarters got the memo about *Punished by Rewards*, in which Alfie Kohn shows how rewards are the flip side of punishment in the behaviorist scheme. Indeed, it seems cruel to claim that the program is “positive” when many children fail to “earn” the rewards; dollars to donuts that every child who wasn’t allowed to attend the PBIS carnival felt punished.

Carl’s behavior in the bathroom and the hallway wouldn’t have provided the only occasions for elaborately behaviorist point distribution and reward systems. Like thousands of schools, Carl’s elementary used Accelerated Reader (AR). The school had bought the Accelerated Reader (AR) library -- normal books with point values attached (based on “difficulty”) -- and the 10 question, computerized, multiple-choice “comprehension” test that accompanied each book. A child picks an AR book, reads it, takes the test, and either moves to the next book, or has to reread the book and retake the test. Once she passes the test, the child is awarded however many points the book was “worth.” These points can translate into grades, but the program also encourages schools to establish inter- and intra-class AR competitions, with individuals and entire classes working to accumulate points to win [insert any prize here].

One day, Isaiah came home from third grade crying angrily. The only phrase I could make out of his snuffling was “stuffed cougar.” When he finally settled down, he explained that his class had just lost a school-wide AR competition. The class
accumulating the most collective AR points had been promised a stuffed cougar. My
older son, a voracious reader, had chosen to read many books that weren’t on the official
AR list. Because books not on the AR list had no points attached to them, my son was
dragging the class AR total down. After they failed to win the stuffed cougar, his
classmates met up with him on the playground, accusing him of causing their
disappointment.

Several weeks later, just after I’d managed to calm my son down and convince
him that reading, not points, matter, Isaiah came home from school upset again. At issue
this time were slushies. In the morning, the teacher had announced that the two best
readers in the class (as determined by AR points) would be given a 32-ounce slushie after
lunch. True to her word, she appeared after lunch with two huge fluorescent blue
(radioactive?) icy treats, presenting them in front of all the children to the two readers
with the most AR points. When someone protested that it wasn’t fair, she asserted that it
most certainly was; after all, these two students had earned the most points!

There was nothing evil about this third grade teacher. She was a caring person.
But she, like most of the teachers in the school, had bought (or, at least, not protested)
AR’s behaviorist mantra. In their hands, Watson’s theory of behavior and Thorndike’s
theory of teaching and learning had become a cynical theory of motivation: readers will
only read if you coerce them with points and prizes. And points and prizes can only be
fairly awarded on the basis of “objective” multiple-choice tests.

Years later, here’s how my sons remember experiencing these point systems.
Isaiah started giving the “I caught you being good!” points back to the teacher.
teacher, befuddled, insisted that he keep them, he gave them away to other students. Now 15, Isaiah recalls,

I hated those points. They’re a bad idea because when you’re older, you feel like the only way you can learn is by having a donut at the end of a dark hallway. So, I always gave [the points] away because I thought they were stupid. I wanted to learn. I didn’t want to be shepherded.

Ben was anxious and indignant for his entire first grade year about the spring carnival. Now 12, he explains,

I didn’t like it because if you were a nice kid, but you weren’t doing everything as perfectly as possible to get those stupid points, you couldn’t go to that stupid carnival, and it really made me mad. One kid, when he was standing in line, he put his finger to his lips, stood up super straight, and the teacher said, “You get five [points], young man!” I was being quiet, but I wasn’t performing, so I didn’t get any points.

Unlike Carl, my sons had help understanding their feelings as they became entangled in the Behaviorist Matrix. Early on, I told them I didn’t care how many AR or “I caught you being good!” points they got; I cared that they engaged in their reading and treated others with kindness. I assured them that their resentment at this system of rewards and punishments didn’t indicate that they were bad, but that the system was. So they came to hate the points, not themselves.

But imagine kids like Carl, who don’t have any help disentangling themselves from the Behaviorist Matrix. Their resentment and anxiety at the system of rewards and punishments turns to resistance from the very learning that the system was supposed to engage them in. This resistance takes several forms, from anger (directed inward or outward) to withdrawal. The theory of behaviorism gives teachers no help in understanding children’s resistance to this type of behavioral control. A teacher’s only
option with a resistant or disengaged learner is to accelerate the punishments and
rewards, which, in turn, increases the resistance or withdrawal.

As MacDougall might point out, a theory of human behavior, learning, and
motivation that refuses to acknowledge the role of “‘motive,’ ‘purpose,’ ‘intention,’
‘goal,’ ‘desire,’ ‘valuing,’ ‘striving,’ ‘willing,’ ‘hoping,’ and ‘responsibility’” (69) is
useless for dealing with children such as Carl. And it is far worse than useless; it is
paralyzing to educative interests.

And, so, finally. Here we have Carl as an 8th grader, in a warehouse of
“alternative” students just like him, all of them angry or withdrawn victims of Watson’s
time of behavior, Taylor’s stopwatch, Thorndike’s theory of learning and measurement
tools, and Yerkes’ large scale standardized testing structures. And here we have Amy,
who has been working triple-time to assure her students that she takes their minds, their
consciousness, and their experiences seriously. It’s been slow going -- they’re skeptical
of her promises, of her commitment to relational teaching. But there have been moments
when Amy thinks they’re starting to trust her. Maybe soon they’ll be able to get down to
the work of forming healthy relationships with learning outside the Behaviorist Matrix.

But it is only October, so early in the school year. These moments of connection
and relationship are too fragile to survive the full-on behaviorist attack of MEAP week.
Despite Amy’s efforts, Thorndike and Yerkes have inserted themselves once again into
Carl’s school life, this time in the form of a disingenuous inquiry about Carl’s friendships
that Amy would have known better than to make. Thorndike and Yerkes don’t have the
courage (or the longevity) to face Carl or his classmates themselves; the MEAP script
forces Amy to do their dirty work. They’ve separated Amy and Carl’s experiences from
learning in their attempt to measure it. The results are tragic but predictable. Carl simultaneously explodes and withdraws.
CHAPTER VI

INTROSPECTION AND EXPERIENCE GO UNDERCOVER AND SHOW UP TO THE COMPOSITION AND RHETORIC PARTY IN NEW PEDAGOGICAL CLOTHES

“For humanists, no good idea ever dies. It just puts on new clothes for new times.”
--David Foster

So far, I’ve been telling a story of winners and losers, employing military terms inspired by Watson’s “battle of behaviorism.” My story ends violently, on a bloody battlefield littered with Watson, Yerkes, and Thorndike’s victims. Over there is Titchener, mind itself skewered through with the words, “as unscientific as soul.” Here are Watson’s children -- Polly and Little John -- struggling to form attachments, damned by their father if they do and damned by their hearts if they don’t. Across from a defeated Lt. George Furgeson, denied promotion by men who did not know him, we have Linda Rief’s middle school, judged not by the New Hampshire community members who know it, but by the AYP formulas that condemn it and the computer program algorithms that determine the “root causes” of its failure. And centerfield from where I stand, we have Amy, her knowledge of Carl’s experiences overwritten by the words in a MEAP manual.

We’ve seen the effects of behaviorist ideas on those whose lives were immediately touched by Watson, Yerkes, and Thorndike -- Little Albert, Polly, Lt. Furgeson, and scores of hungry cats. But we also know that similar effects continued to play out long after Watson, Yerkes, and Thorndike were dead and gone. Familiar to the point of being invisible, the behaviorist infrastructure of educational accountability works like Robert Moses’ low Long Island overpasses: higher standards and the tests that enforce them create smaller and more difficult-to-open gates to opportunity and

125
resources. Students, teachers, and schools become trapped like Thorndike’s vigorously
struggling cats when their scores don’t trigger the latches.

This infrastructure doesn’t just parcel out opportunity and resources, restricting
access to those who have the least, and consolidating wealth amongst those who have the
most. It also shapes the relationships between community members and educators,
between educators and children, between children and learning. It reconfigures the
school as a business, treating students and their learning as products, with both teachers
and students as laborers whose productivity must be quantified, documented, and
managed through inducements and disincentives. Promising to restore transparency and
inspire confidence in our public institutions on the one hand, it creates distance,
dermines democracy, and foments mistrust on the other. And it maintains stability by
convincing everyone that they’ve earned what they’ve gotten -- or failed to get.

Like Amy, who knows that the way to reach Carl is not to hold grades or test
scores over his head but to form a mentoring relationship with him, many teachers within
this system close their classroom doors and refuse to let their relationships be shaped by
Watson and his crew -- as long as no administrator or literacy coach interrupts to assure
“fidelity” to the commercial programs that districts often buy in an effort to increase test
scoresxv. And surely, this kind of resistance means something: to the teachers, whose
refusal to parrot the behaviorist script affirms their humanity, and to their students, for
whom every moment of healthy connection is important.

But no matter what these teachers and students know and believe, they work
within a doggedly behaviorist infrastructure of accountability. And the only thing that
replicates itself as quickly as standardization within this structure designed to control
man's actions, is the call for compliance. And so, teacher evaluations are increasingly linked to student test scores, despite a lack of evidence that this system of evaluation works, even on its own terms. And, most recently, lest teacher educators in colleges of education try to teach their pre-service teachers that a focus on test scores can undermine long-term learning, the Department of Education has piloted a program that proposes to evaluate teacher educators by the test scores of their students' students. While this is a voluntary grant now, it foreshadows the next mandate. There is no room for conscientious objectors in a system that denies the presence of anything like a conscience.

While this story has its share of victims, it is a story that can't be told strictly in terms of winners and losers. Watson may have won the battle of behaviorism. And, without a doubt, Yerkes won the contract for the human engineering of the schools, a contract that federal education secretaries Rod Paige, Margaret Spellings, and Arnie Duncan have been more than happy to honor. But, as David Wallace points out, "[f]or humanists, no good idea ever dies. It just puts on new clothes for new times" (456).

"REPORTS OF MY DEATH HAVE BEEN GREATLY EXAGGERATED":

INTROSPECTION LIVES

Despite Watson's battle, introspection -- and experience itself -- are alive and well. In support of the hypothesis that it survived rather than staging a miraculous resurrection, historian of psychology Alan Costall argues that the simple, naïve version of introspection that Watson vanquished was a bit like a straw man™: introspection in the early twentieth century hardly held the naïve monopoly on psychology that Watson
claimed it had. It isn’t my intent to tell a comprehensive history of introspection and the related constructs of *mind, consciousness* or *experience* after Watson’s attack, but simply to point to several locations in psychology and rhetoric and composition studies where they survived and even thrived. In fact, the three locations in psychology I’ll identify -- Freudian psychology, Jacobsen’s “progressive relaxation,” and cognitive psychology’s think-aloud protocols -- were, perhaps, the breeding grounds for the process movement that shaped composition studies in the 1960’s and 70’s.

The most obvious place to look for consciousness in the first half of the twentieth century was in Freudian psychology. Freud’s theories revolved around consciousness and its related constructs of unconsciousness, and pre-consciousness\textsuperscript{XVII}. While Titchener had focused on consciousness for the purpose of generating theoretical psychological knowledge, Freud focused on consciousness for therapeutic purposes – to treat neuroses. Treatment in Freud’s psychotherapeutic model began with a form of intro- and retro­spection. The patient’s description of his thoughts, feeling, memories, and associations were the materials for analysis, as the analyst attempted to help the patient resolve various urges or conflicts at different levels of consciousness.

The great paradox, as Mark Rilling describes it in “John Watson’s Paradoxical Struggle With Freud,” was that Watson himself helped secure Freud’s place in the American psyche. Rilling claims that Watson was fascinated with Freudian concepts. He not only accepted many of them, but also spent much time trying to explain them in his own terms even as he actively vilified him. Thus, Watson appropriated Freud’s work into his own, but then obscured his debt to Freud by dropping any reference to him in addition to using his name as a put-down.
Rilling describes three phases of Watson’s attempt to grapple with Freud. In the first phase, Watson tried, less successfully, to explain Freud’s concepts in terms of William James’ concept of a “habit.” Later, during the most “creative” part of his career, Watson framed psychoanalytic concepts in terms of stimulus and response. In fact, Rilling argues that the “Little Albert” experiment -- and Watson’s choice to focus his research on the infant more generally -- was actually Watson’s attempt to explain Freud’s observation of emotional displacement or transference using the terms of Pavlovian conditioning: in attempting to condition fear into infants, Watson was trying to show how mothers displace their own fears onto their children. The third phase of Watson’s career was marked by contempt for Freud, (he referred to psychoanalysis as “voodooism” in Behaviorism (24)), but Rilling claims that Watson himself was largely responsible for the “Americanization” of Freud (302).

Just as Freud had used intro- and retro-spection therapeutically, one of Titchener’s students, Edmund Jacobsen, would pair introspection with the behaviorist’s focus on measurable behavior for treating anxiety. Unlike the analyst in Freud’s therapeutic model, however, proponents of “progressive relaxation” (also known as “biofeedback”) would make no effort to find or make meaning from the patient’s introspection. Instead, Jacobsen’s work would depend upon the phenomenon that led Dewey to observe in 1891: “...when introspection starts, the anger ceases...” (8).

“Nervousness” was considered a uniquely modern American malady in the late eighteenth and early nineteenth centuries, the price paid for the country’s entrance into the industrial age. William James, along with several presidents, suffered from nervousness and cultivated it as part of their identities as “fully modern individual[s]”
 Jacobsen, a self-proclaimed sufferer of nervousness himself, would combine Titchener's introspection with Watson's focus on measurable behavior to create a new treatment for anxiety.

After earning his degree in psychology, Jacobsen had worked for a year with Titchener at Cornell, perfecting the method of introspection, and helping Titchener work out his position on "imageless thought," a popular psychological controversy of the time. But when Jacobsen left psychology to become a doctor in 1920, he found himself the victim of anti-semitism. Unable to climb the normal medical career ladder at a hospital, he opened a private practice. He'd noticed the calming effect of introspection under Titchener's tutelage, and the feeling of control it had given him. He began teaching patients to focus intensely on their own sensations while tensing and then relaxing various muscles. He called this, "progressive relaxation."

As his interest in this treatment grew, Jacobsen began experimenting with various medical measurement instruments. He'd first intended to document the effect of relaxation on physiological responses such as blood pressure and pulse. But he noticed that these instruments could interact with the relaxing subjects, who could use the measurements to attempt to control their own physiological states. He built special chambers and rooms with hidden measurement instruments in which patients could take their cure, hooked up to various electrodes, monitoring their relaxation via electromyograms. Though not uncontroversial, progressive relaxation caught on, and Jacobsen even treated members of the Rockefeller family.

Progressive relaxation was a strange marriage, indeed: the introspector's private, subjective realm of experience wedded to the behaviorist's measurement of muscular
movements. Despite its “hocus pocus” trappings of consciousness, the technique still embodied Loeb’s credo that science should be used to control nature. Patients might be ruminating on their own private sensations, but they were controlling them through the use of scientific measurement instruments: “the perfected self as experiment” (Kroker 103).

But introspection after Watson wouldn’t just be confined to clinical uses; cognitive psychology would once more use it as a method for generating psychological knowledge. In the 1950’s, cognitive psychology reintroduced a form of introspection — “think-aloud protocols” — to shed light on mental functions and structures. While cognitive psychology’s founders proclaimed themselves the “revolutionary alternative to mechanistic, stimulus-response behaviourism” (Costall 637), Alan Costall argues that their work was shaped by Watson’s behaviorism more than they realize. As Costall puts it:

> In fact, cognitivism is no more than the “flip-side” of stimulus–response behaviourism, where “mental processes” are defined as whatever “is left over after one tries to stuff all psychological phenomena into the S–R box” (Reed, 1997a, p. 267). (qtd. in Costall 638)

According to Costall, cognitive psychology never truly challenged Watson’s exclusive focus on stimulus and response, sometimes obscuring this allegiance by substituting the language of computer processing for mental processing — inputs (stimulus) and outputs (response). This observation helps explain the criticism leveled in the 1980’s at compositionist Linda Flower and her coauthor John Hayes, who used think-aloud protocols to draw a parallel between composing and cognitive processes (see Berlin, “Ideology...”; Bizell; North). Despite Costall’s convincing argument that cognitive psychology was not the liberation of mind and consciousness from behaviorism that it
claimed to be, practitioners of the think-aloud protocol certainly embodied the spirit of Titchener’s Observers, giving voice to their thoughts and feelings as they engaged in an activity or interacted with a stimulus.

**THE WRITING PROCESS MOVEMENT: “EXPERIENCE SQUARED”**

The field of Composition Studies has many Genesis Stories. My interest in entering this colloquium of stories is less in claiming an authoritative genealogy than in tracing the ways in which the field has cultivated an interest in the key term of my scholarly interest: *experience*. Located within an American educational system whose concerns were shaped largely by Watson, Yerkes, and Thorndike’s behaviorism, the strong theme of experience within the Composition Studies is all the more remarkable.

The composition course exposed instructors to students’ written accounts of their experiences -- long before Peter Elbow and Donald Murray would make their mark on the field. While Sharon Crowley refers to the “daily themes” assigned by Harvard composition instructors (such as Barrett Wendell) as the university’s way of policing student subjectivities (74), it is difficult to read Thomas Newkirk’s description of Wendell’s English 12 course in the late 1800’s and not to be struck by how interested Wendell was in the lived experiences of the young men in his class. First, there was Wendell’s “personal invitation to students in his opening remarks” (“The Politics of Intimacy” 119) at the beginning of the 1888-9 school year:

> So far as I can stimulate you and you one another feel what life is; to feel the inevitable inadequacy of all expression thereof; to feel, more and more, the beauty for all its inadequacy...we shall do well...(Notes on Lectures, October 3, 1888).” (Wendell, qtd. in Newkirk 119)
This invitation establishes the experience of life and the difficulties of its expression as a major goal of the course. In addition, Wendell solicited descriptions of his students’ experiences in the “dailies” -- one page themes written six days a week, “focusing on what made that day in some way particular.” Experiences described in these dailies ranged wildly:

“...watching an amputation at the medical school, the joys of the ukelele, the novelty of typewriting and electric lights...roommate problems...watching a drunk being carried off in a wheelbarrow...” (Newkirk 121)

Beyond soliciting students’ accounts of their daily experiences, Wendell solicited their experiences of the course itself: criticisms of work assigned, teaching methods employed, and feedback given. In criticizing Wendell’s feedback in 1886, one student’s insight about the connection between interest and quality sounded much like Kenneth Macrorie’s a century later. In arguing with Wendell’s judgment (“Your work is exceptionally uneven, from whence I infer that you do not always take pains”), the student protested:

I think the cause of the unevenness in the work is not so much lack of pains as lack of interest in the subject...a great deal of work is done unconsciously on a subject you are interested in. But this ought not to be called “pains” for the very reason that it is unconscious work. The amount of conscious effort on every theme but one has been about the same. (1886, “Themes”) (qtd. in Newkirk 123)

While Wendell didn’t make students’ experiences of the writing process central to his theory of composition in the same way that later writing process theorists would, he positioned himself as close as possible to the experiences of the young men he taught.

Almost a century later, what was true for Wendell in practice became true in theory for writing process pioneers such as Donald Murray, Peter Elbow, and Janet Emig,
to whom the concept of experience was truly generative. Here’s Peter Elbow on the
presence of experience within the process movement of the 1960’s, 70’s, and 80’s:

What these figures had in common -- and what seems to me to characterize that
moment in the history of composition -- was a burgeoning interest in the
experience of writing. There was a mood of excitement about talking about what
actually happens as we and our students write. Thus, there was lots of first person
writing...‘process’ means experience (*Everyone Can Write* 65).

As compositionist and long-time colleague of Donald Murray, Thomas Newkirk, puts it,
the process movement was “experience squared”: writing about experience and talk about
the experience of writing (email).

After spending time with Titchener, whose method for understanding mind and
consciousness was to ask individuals to talk about their experiences, it is difficult not to
read the writing process movement as a reclamation (or, a reincarnation?) of the losers of
Watson’s war. Consider two juxtaposed passages: one from Titchener’s nineteenth
century psychology textbook, next to one from Murray’s 1970 article, “The Interior
View: One Writer’s View of Composition”:

A mental process is a process which can form part of the experience of one
person only...Not only does the mental process go on inside of you, it is so
entirely inside of you that you are the only person who can ever get at it and
observe it. (Titchener 9-10)

And,

At the moment of writing the writer has a fundamental aloneness...I have found
that at the center of the process I am alone with the blank page, struggling to
discover what I know so that I can know what to say” (Murray 22)

But, despite certain similarities in their focus and method, writing process leaders weren’t
drawing on Titchener’s work: his name never appears in the writing process literature.
When the word “introspection” occasionally appears, for example, in Sondra Perl’s work,
it is used as a general synonym for reflection rather than a gesture to Titchener’s
psychological method. But, for early process movement leaders, reflection wasn’t simply a habit of mind: talk about experience was also a tool for writing process theorists to generate professional knowledge. Murray and Elbow, along with other leaders of the writing process movement, were building their entire theory of composition and their practice of teaching composition around introspection -- their own, and their students’.

“The Interior View: One Writer’s View of Writing,” published by Donald Murray in 1970, is as good a place as any to begin our examination of the writing process movement’s concern with experience. Murray’s title itself indicates the movement’s defining tension and central concern: in a field defined by an “exterior view” of writing (study and analysis of published products), Murray advocated a turn inward, towards an awareness and discussion of the experience of writing as it happens, as it is in-process. That fact that Murray, a columnist for the Boston Globe, would write about his experience of writing in 1970 was not entirely remarkable. Writers had written about their experiences of writing long before he did, and his published work is peppered with their insights: in “The Interior View” alone, Murray quotes no less than 10 authors who write about their experience of writing, including Goethe, Spender, and William Carlos Williams. What was remarkable, perhaps, was that Murray was not just writing as a writer, but as a writing teacher, and he was beginning to construct a theory of how we compose and how we could teach composition from his introspection and the collected introspection of generations of great writers.

Writing pedagogy had principally been shaped by what Murray called the “exterior view of writing, principally examining what has been written or studying patterns which have evolved by the analysis of what has been published.” (21)
didn’t dismiss this view as useless, but explained his own view differently: “I do not see writing from the exterior view but from within my own mind and my own emotions as I try to write every single day of my life.” (21). Three years after Murray published “The Exterior View of Writing,” Peter Elbow would publish *Writing Without Teachers*, in which he warned readers that his advice to writers was based on his own experience:

> Though much or all of this may be in other books -- some of which I have probably read -- it seems to me my main source is my own experience. I admit to making universal generalizations upon a sample of one. Consider yourself warned. (16)

Elbow’s generalizations were the product of his introspection about his own writing process. He described how he came to the practice of freewriting: when he got stuck while writing, he would, “...take out a fresh sheet of paper and simply try to collect evidence: babble everything I felt, when it started, and what kind of writing and mood and weather had been going on” (18). Similarly, when Elbow successfully broke through his writing block,

> I would often stop and try to say afterwards what I thought happened. I recommend this practice. If you keep your own data, you may be able to build your own theory of how you can succeed in writing since my theory of how I can succeed may not work for you. (ibid)

Here -- *you may be able to build your own theory of how you can succeed* -- Elbow made the democratizing, humanizing turn that characterized the writing process movement.

That Elbow and Murray used their experience as writers as the basis for their own theory of composition was departure enough. They could have made their mark on the field on the basis of this move alone, reserving introspection for their use as theorists -- arriving at generalized principles that student writers should then learn to observe.
But Elbow and Murray were deeply concerned with the experience of each student. In fact, they proposed a form of introspection as the means by which each student writer might “learn the process for himself” (Murray 15). Here’s Murray again:

> There is no one way to write and there is no one way for the student to learn to write. We must accept the individual student and appreciate his individualness….ultimately he [the student] has to learn the process for himself. (24-25)

Later, Murray would suggest in “Teaching the Other Self: The Writer’s First Reader” that the point of the writing conference was not to examine the student text, but to interact with the writer’s “other self.” According to Murray, in writing and then reading his own writing, the writer forms two distinct selves: the first self who writes; and the “other self” who reads, counsels, advises, and navigates the territory mapped out by the writing for the first self. The other self also “articulates the process of writing” (Murray 142). Lest we confuse an articulation of the process of writing with a purely behavioral description of what the writer does, Murray assured us that the teacher must first acknowledge and respond to the writer’s descriptions of his feelings as he writes (145). The writer needs this other self to develop and grow, and the teacher can help make this growth possible simply by encouraging, expecting, and listening to the other self speak. In our Titchener-inspired terms, then, Murray held that the teacher’s job in the conference was to interact with the self created by the student’s introspection.

While Newkirk calls the writing process movement “experience squared,” I might call it “experience cubed.” Experience supplied Elbow and Murray’s method for generating theoretical knowledge. And experience became the subject of their students’ writing. But the concept of experience, in the tradition of John Dewey, also shaped their conception of their teaching: “We are coaches, encouragers, developers, creators of
environments in which our students can experience the writing process for themselves” (Murray 5).

FROM INTROSPECTION TO FREUD: LAYERS OF SELF

In the process movement, we see strains of the various forms that introspection took in Freudian psychology, “progressive relaxation,” and cognitive psychology. There’s a fascinating focus on meditation and self-help reminiscent of Jacobsen’s work in the early process movement (see Berlin, Peary). And Linda Flower’s collaboration with cognitive psychologist John Hayes introduced the field of composition studies to think-aloud protocols. But I’m going to focus here on Freud’s influence on the early process movement. Murray, Elbow, and Emig never drew directly on Titchener’s work, perhaps because Watson had won that battle. But they drew directly on Freudian language and concepts. Three years after Murray proposed nurturing the student writer’s “other self,” Janet Emig described the multiple selves -- or multiple layers of self -- that must be attended to by the writing teacher. Her version of Murray’s “other self” had a Freudian twist: the writing teacher must nurture the student’s unconscious self.

Rather than dividing the writer into two separate selves as Murray did, she divided the self into layers. These layers first take the form of skin in her opening startling and wonderful image of the writer who has dutifully produced “the conscious student theme” (46):

…the theme seems to have been written from one layer of the self -- the ectoderm only, with student involvement with his own thought and language moving down an unhappy scale from sporadic engagement to abject diffidence. (46)
Emig quickly dropped the skin analogy -- we hear nothing further about the ectoderm or endoderm -- but in this line, she vividly planted the idea of the layered self, some of the layers exposed on the surface, and some submerged underneath.

Freud never viewed the conscious or unconscious as having different locations within the body in any literal sense. But popular imagination did. Even now we talk of “uncovering” our unconscious thoughts, of “peeling the onion” of our selves in therapy, of repressed or recovered memories which implies a place where the forgotten memories have been stored, held under the surface, hidden from our conscious self, which lies at the surface. Like Murray, who wanted to see a shift from the “exterior view” of composition to the “interior view,” Emig argued that traditional writing instruction doesn’t allow the student to “consult this [unconscious] part of the self” and “conspires against his inwardly attending” (46). She discussed how author Rudyard Kipling personified the “unconscious part of the writing self into daemons” (49) and how Amy Lowell described dropping a simple topic for a poem “‘into the subconscious much as one drops a letter into the mailbox. Six months later...the poem...was ‘there’” (Lowell, qtd. in Emig 52). According to Emig, writing teachers needed to encourage inward attending -- journeys to the unconscious -- if students were to write papers that went beyond (or below) the “surface scrapings” produced by a traditional overemphasis on the surface -- the control of the conscious mind.

Five years after Emig’s “The Uses of the Unconscious in Composing,” Murray would make his plea for the role of procrastination in composing on the grounds that procrastination allowed the subconscious to do its work. In “Write Before Writing,” Murray explained why writers procrastinate:
They sharpen well-pointed pencils and go out to buy more blank paper, rearrange offices, wander through libraries and bookstores, chop wood, walk, drive, make unnecessary calls, nap, daydream, and try not ‘consciously’ to think about what they are going to write so they can think subconsciously about it. (376)

To Murray, Elbow, and Emig, one legitimate subject of composition studies was the writing self, and that self was divided: into the writing self and the other self; into the unconscious or subconscious and conscious; or into the id, ego, and superego. Emig and Murray emphasized the subconscious mind, unconscious mind, or the id as a corrective to the overemphasis in traditional writing instruction on the conscious mind. The writing teacher, according to writing process leaders, needed to acknowledge the student’s experience and various levels of consciousness in the process of writing; furthermore, the writing teacher needed to structure classroom activities in order to make room for and work with the students’ consciousness(es) and experiences of writing.

The centrality of self and experience in the early process movement was later represented as naïve and simplistic by members of the social turn, especially James Berlin and David Bartholomae (see Berlin, “Rhetoric and Ideology”; Berlin, *Rhetoric and Reality*; Bartholomae, “Interchanges...”; Bartholomae, “Inventing...”). While Berlin and Bartholomae’s critiques prompted a necessary examination of the complexity of self-hood and the ways in which experience is ideologically constructed, I believe these critiques missed or dismissed a defining gesture in the field: asking an individual to represent the way she experiences something is a humanizing and democratizing act, an act that assumes the rights to self-hood and self-representation. The focus on experience by early founders in the field of composition studies such as Elbow, Murray, and Emig may not have been sufficient, as critics such as Berlin or Bartholomae pointed out. But
this focus on experience is, I contend, necessary. It is a focus and concern that I would extend to writing assessment.
My critique of the field of writing assessment involves a truly troubling contradiction. I will be accusing writing assessment scholars of failing to confront behaviorism at the same time that I am asserting that their work has been characterized by resistance to behaviorism’s many reductions. The concept of an infrastructure -- which perpetuates its originating values by structuring our interactions, choices and experiences -- has helped me to reconcile this apparent contradiction.

ON BICYCLES, HUMMERS, AND MULECARTS

I didn’t think much about the role of infrastructure in my life until I was transferred from the high school where I taught, which was half a mile from my house, to the district’s alternative education school, which was six miles away. Conscious of my carbon emissions, I didn’t want to drive to work. But my rural city didn’t have public transportation, so I thought maybe I would bike. I wanted more exercise anyway, but found time for exercise in short supply with my job and young family. Thrilled with the prospect of simultaneously taking care of my commute and daily workout, I bought a road bike.

But biking to work proved neither enjoyable nor safe. I had neglected to consider the infrastructure. My new school was located on a highway with no sidewalks or bike lanes. I never saw bikers on that stretch of road, so drivers wouldn’t even think to watch
for me. I chanced it several mornings -- after all, I’d bought this great new bike! But I
was miserable. When I wasn’t coughing out exhaust or cursing the small-business tax
loophole that encouraged real estate agents in my community to buy Hummers, I was
imagining the safest way to land in a ditch: go limp on impact, or try to roll? Defeated, I
“decided” to drive to and from work every day.

Now that I walk the mile-long footpath between my apartment and my university
office, I realize how many choices and experiences were actually shaped by my
“decision” to drive to the alternative school. My walk takes me through a small space
of woods, and I frequently stop to watch a strangely colored and ever-growing family
of deer that roams the island on which the university is built. I arrive at work in the
morning feeling relaxed and awake.

I also think much more carefully the night before about what I want to bring with
me, knowing I won’t be able to drive home in the middle of the day. And thinking
through what I’ll need for the following day always reminds me to pack gym clothes
for a visit to the recreation center near my office. I work out more often now simply
because I’ve planned it the night before. I pack a sandwich, because cobbling together a
lunch from vending machines won’t supply the necessary energy to work out, and I
won’t be able to drive to purchase something healthy. In short, the infrastructure,
which I rarely take the time to think about, shapes my choices, my ability to enact my
values, and my experience of my life.

Infrastructure, as I’m using the term here, structures our movement through our
environment, as well as the technologies we create to build, maintain, and navigate it.
Like oceans or mountains, existing infrastructures often feel neutral and natural -- part of
the landscape -- to the individuals whose movements, activities, and decisions they shape. Within an infrastructure, an individual may feel that his choices are inevitable: of course he must choose a car to drive on a highway! Likewise, the spread of these supporting technologies expands the infrastructure itself, crowding out alternatives: the more cars we purchase, the more lanes of highway we must build. Before long, alternative infrastructures begin to disappear through neglect. I am reminded of this every time I merge onto I-95 from my home in Orono, Maine. At the end of the entrance ramp there is a sign that I have seen posted in no other state: “Prohibited: animals -- led, ridden, or driven.” I have no desire to ride or lead a mule from my home in Orono to pick up cannoli from my favorite café in Bangor. But the sign reminds me of a time when the possibility that two technologies requiring separate infrastructures might tragically collide demanded this warning. But eventually, highways became our landscape, and alternative technologies -- bikes or mule carts -- went the way of nostalgia.

Because infrastructures feel like landscape, we often don’t think about their originating values. It makes no sense, after all, for a biker to stand at the shore of an ocean in his way and “disagree” with the values that called the ocean into being. But an infrastructure and its supporting technologies, unlike the ocean or the mountain, result from human decisions and therefore embody human values. Users of the infrastructure need not share these originating values in order to participate in their effects, but they do. Remember, for instance, those low Long Island overpasses, which still embody Robert Moses’ infatuation with middle-class car-culture, wrapped up in his social and racial prejudices. The contemporary car owner on Long Island who plots his daily commute under Moses’ overpasses and the bus driver who knows to avoid them likely have no
intention of perpetrating social and racial inequality. But their activities -- structured by the infrastructure -- still position them within a stratified social and material network. And their participation in this network stabilizes and extends it.

This conception of infrastructure can make us feel helpless and hopeless. Doesn’t it deny our sense of ourselves as principled decision-makers? After all, the alternative technologies presented us within the infrastructure sometimes aren’t real alternatives: Ford or Chevy? But once we understand an infrastructure’s originating values and the technologies, choices, and experiences it structures, we are presented with a new series of choices. These choices involve profound responsibilities -- so profound, in fact, that we sometimes avoid them through convincing ourselves of their impossibility, preventing our understandings by predicting that whatever follows them will be deemed “impractical.”

It is true that I have not set out here to create immediately marketable assessment technologies. But, at the same time that I don’t claim my work is immediately “practical,” I still claim that it fits within the pragmatist tradition in two important ways: I have tried to trace the practical consequences of our ideas about assessment through the human and non-human “actors” in our social networks; and I have used my understandings, which lead to no immediately marketable technologies, to do precisely what William James claimed is the practical cash-value of our ideas -- “to help us get into more satisfactory relations with other parts of our experience” (“What Pragmatism Means”). Through my work here, I aim to get in better relation with my experience as a teacher, a writer, and someone deeply interested in the experience of others. I aim to face, as much as I am able, the profound responsibilities that sometimes follow these
understandings. Any despair involved in these understandings cannot compete with the despair I have already felt from being asked to assess writing in a way that ignores -- in theory and practice -- my experiences.

AN UNLIKELY RELATIONAL TRIANGLE: COMPOSITION STUDIES, EDUCATIONAL MEASUREMENT, AND BEHAVIORISM

I have promised to tell a story of writing assessment's resistances to behaviorism - resistances that have been heroic but which have ultimately accepted and perpetuated behaviorism's dismissal of experience. However, resistance implies connection. So my story of resistance begins with a complex relational triangle: composition studies, educational measurement, and behaviorism. Writing assessment, that anomalous subfield within composition studies, was one byproduct of the tension from this tangled trio.

Composition, as a field of study, was emerging in relationship with the entrance examinations first given at Harvard in the 1870's. Sharon Crowley claims that, by proving just how many incoming freshmen couldn't write, the exams necessitated the first-year composition course and the discipline that emerged around it. In 1888, Harvard’s Adams Sherman Hill complained,

Between 1873 and 1874, I read several thousand compositions written in the examination room...Of these a hundred, perhaps—to make a generous estimate—, were creditable to the writer...some of the compositions are a great deal worse than the mass, and some a little better; but in other respects there is a dead level...(1888, 12-13) (qtd. in Crowley 68).

As Crowley dramatically put it, “the exam breeds a course” (72); the deficiencies revealed by these exams led to the universal requirement that Crowley depicts as composition’s lifeblood -- and its bane (66, 72-4).
The first-year composition course does not represent the entirety of the field’s investments. Interest in composition across the lifespan and across settings has played an important role in the field. But the first-year composition course does represent an origin of sorts: it was the cause around which disparate interests and funding streams merged. Telling the story of composition studies by starting with the first-year writing course allows me to show composition’s entanglement with the very idea of assessment — and the behaviorist infrastructure that had defined assessment’s concerns, forms, and uses.

While the composition course was emerging in relationship to the entrance examination, the entrance examination was taking shape in relationship to educational measurement. The College Entrance Examination Board (CEEB) and Educational Testing Services (ETS) embodied this relationship: the CEEB had been created in 1900 as part of an effort to standardize entrance exams across various institutions (Lemann 28-9); the CEEB, in turn, created Educational Testing Services (ETS) in 1948 (Lemann 65), to be a “new national agency to control all the leading tests” (Lemann 61). Besides controlling all the leading tests, ETS sponsored many important educational measurement innovations and scholars. The format of the entrance examination followed ETS and educational measurement through various incarnations: from essay tests scored by local institutions to centrally scored multiple-choice grammar tests, to centrally (holistically) scored essay tests in combination with multiple choice grammar tests.

To complete our tangled triangle, educational measurement itself was emerging in relationship with behaviorism. First, and most obviously, behaviorism provided the blueprint for the shape that the examinations would take: the behaviorist’s stimulus-response experiment. Like the experimental behaviorist, the educational measurement
scientist created and administered a stimulus (the essay prompt or multiple-choice question) in a controlled environment in order to elicit a response behavior (the test-taker’s answer/s) that could then be measured. Through the influence of the CEEB and ETS, this blueprint was followed more and more closely -- from the loosely and locally standardized stimulus of each university’s essay prompt to the centrally standardized and easily measurable multiple-choice grammar question of the SAT. The constructs created and elaborated by the field of educational measurement were developed to give theoretical flesh to this basic behaviorist blueprint.

Furthermore, while behaviorism didn’t lead to the entrance exam (it had made its debut on the university scene several decades before Watson’s Manifesto), the logic of behaviorism affirmed the exam, ensuring that it prevailed over competing admissions systems. It is easy to look back at entrance examinations as inevitable -- the only logical solution to the growing problem of too many qualified applicants. But several other solutions could have been tried, and at least one was: in the early 1870’s, the University of Michigan introduced a “certification” or accreditation system by which it would accept graduates from any high school that it certified, referred to as a “Diploma School” (Williams 17). Faculty from the university would visit the high school, examine its curriculum and facilities and attend classes. (During his time at the University of Michigan, John Dewey participated in such visits.) Many schools -- especially from the Midwest -- participated in this certification system.

But once it grew in popularity, behaviorism provided the logic that helped to authorize one system over the other. Compared to the entrance exam, the certification system was low on the behaviorist hierarchy. No standard “prompt” elicited an easily
measurable behavior in the certification process. Furthermore, conducted as they were in the naturalistic setting of the functioning high school, the certifiers' observations and judgments hardly took place under controlled conditions. Entrance examinations, on the other hand, provided a standard prompt given under controlled conditions, making the judgment itself easier to standardize. The logic of behaviorism, then, was part of the environment that "naturally selected" examinations over certification. While vestiges of the certification system remain in the form of high school accrediting bodies, graduation from an accredited high school is sometimes a pre-requisite, but never a guarantee of admissions as it was designed to be.

While the accreditation system didn't fare well on the behaviorist hierarchy, one radically different admissions system was never tried on a large scale because it wasn't on the behaviorist hierarchy at all: To solve the problem of a mismatch between available resources and demand, universities could have instituted a lottery. But a lottery doesn't follow the logic of behaviorism, even badly. In the logic of behaviorism, limited university spots and scholarships present occasions for control in two ways: determining which kind of people are groomed for leadership; and controlling the efforts of those preparing to compete for the limited spots. But the lottery system renders limited university spots impotent for the purpose of social control. In the lottery system, there is no stimulus, there is no behavioral response to measure, and there are no rewards or punishments to dole out -- and thus, no control to exert.
FROM THE EPISCOPACY TO THE MERITOCRACY: EDUCATIONAL
MEASUREMENT HELPS PRESERVE THE ELITE IN A DEMOCRACY

The system of resource allocation that did complement behaviorism perfectly was Yerkes’ vision of a society controlled by Human Engineers. Behaviorism was the perfect theoretical construct for creating and maintaining a meritocracy, defined here as rule by the most intelligent people, as determined by IQ scores. The meritocracy had been prefigured in Yerkes’ twin interests: IQ testing and Human Engineering. By midcentury, the meritocracy was eagerly pursued by Harvard president James Conant and ETS’ first president, Henry Chauncey.

As Nicholas Lemann tells the story, James Conant objected to the stranglehold of the New England “episcopacy” -- families who had descended from the British nobles who had landed on the east coast in the seventeenth century and who had run the United States ever since. Conant wanted a different kind of elite, a “natural aristocracy” who would rule by virtue of their intelligence rather than their family’s position. Conant had found the phrase “natural aristocracy” in one of Thomas Jefferson’s letters to John Adams and enthusiastically took it out of context -- it had never been a key term for Jefferson -- to support his vision for creating a new American elite using IQ tests.

After Conant got rid of Harvard’s athletic scholarships, he recruited Henry Chauncey to establish a new scholarship program that would use the SAT to find and recruit talented students from far outside the Episcopacy, in the Midwest. Although Chauncey was part of the Episcopacy himself, he was also a firm believer in the meritocracy. The term, “meritocracy,” had first been used by the British sociologist
Michael Young to describe, “rule not so much by the people as by the cleverest people” (Young, qtd. in Lemann 117).

While Chauncey believed in Young’s “rule by the cleverest people,” he seems not to have realized -- or minded -- that Young had used the term satirically. In Young’s satire, “Rise of the Meritocracy,” published in 1958, rule by the cleverest people did create an explosion of wealth for Great Britain. But the clever elite quickly turned into an inherited elite (since the elite intermarried and passed on their merit, measured in IQ, to their children), and became so arrogant and controlling that the lower classes rose up in bloody revolt.

Satire or not, Chauncey knew, without a shadow of a doubt, that testing was the single most important mechanism in the gears of the meritocracy-making machine. As an assistant dean at Harvard, Chauncey had proposed a Census of Abilities: multiple-choice tests (mainly the SAT) would be used to catalog the human resources of every person in the country. After Chauncey established Conant’s scholarship program, Conant invited Chauncey to further his meritocratic program by becoming the first president of Educational Testing Services (ETS). Tests, centralized in ETS, the nation’s new testing agency, would discover Conant’s “natural aristocracy,” sorting them into the educational tracks that would groom them for leadership. Everyone else would be sorted away from these particular leadership tracks, of course: if Brigham and Yerkes had had their way, some would have been selectively bred right out of citizenship or existence altogether.

It is interesting to note that, decades after the eugenicist intentions of early testing architects and proponents have largely been forgotten, the contemporary educational accountability system touts racial equality as its raison d’etre: what unites these two
seemingly irreconcilable agendas is their implicit acknowledgement of educational measurement’s role in sorting, grouping, and apportioning resources. Responsibility for the consequences of this project -- inequality -- would be displaced onto the objective testing apparatus and onto individuals, who either deserved or didn’t deserve what came their way. The meritocracy would be defended by those who benefitted most from it. And, it would be at least passively supported by almost everyone else -- either by those who thought they could benefit from it if they just “had what it takes” and worked hard enough, or by those who blamed themselves for not being smart or hardworking enough to make it.

Thus, we have what I have been referring to as the behaviorist infrastructure: an elaborate tracking system, much like a network of roads leading to different clusters of opportunities and resources (or lack thereof). At frequently placed testing tollbooths, built and maintained by educational measurement, we pause in the movement of our lives to be told which of the roads fanning out from the tollbooth are, in fact, available to us, according to our objectively tested abilities and needs. Our relationships, experiences, choices, wealth, achievements, and influence are shaped, in large part, by our movement in and through this infrastructure.

Surely this infrastructure has been touted as democratic and egalitarian; both James Conant and Henry Chauncey firmly believed that it was. Women were admitted to Harvard’s medical and law schools under Conant’s presidency, and Conant and Chauncey’s new scholarship program gave Midwesterners outside the Episcopacy
opportunities they wouldn’t have had otherwise. In Chauncey and Conant’s view, they were extending invitations to join the elite further than they’d even been extended before. But, they were still preserving the concept of an “elite”: the infrastructure they were helping to create was egalitarian only in the sense that they encouraged everyone to use it. Once in it, however, the tracks or roads themselves were (and are) anything but democratic and egalitarian; only a few of the roads led to wealth and influence. Everyone is invited to compete for admission to the same few roads, but there are only, ultimately, a few of them, with even fewer decision-making destinations at their end.

And, as Nicholas Lemann points out, this was an infrastructure that was created by the elite: it was never proposed for democratic deliberation. At first glance, the idea might strike us as something that the public would have approved anyway. Who would object to rule by the most qualified, intelligent people? Isn’t the alternative rule by unqualified and unintelligent people? As Dewey would remind us, however, the promise of democracy isn’t the best decisions, most efficiently made. The promise of democracy is that, by investing power equally in ordinary people, our experience of our lives is transformed, making us all extraordinary. Rule by the cleverest people, then, runs the risk of making the rest of us alternatively passive and, as the ending of Young’s story illustrates, very, very angry.

While segregation and inequality were anything but “unintended” consequences of this behaviorist infrastructure (equality cannot be the result of a system designed to select an elite), the composition course was an unintended consequence. The problem with installing tollbooths -- mechanisms that stop traffic to sort it -- is that they sometimes create traffic jams. The finding of the first written entrance examinations --
that most students couldn’t write! -- had the potential to create a massive bottleneck. This was a consequence, unintended or not, that needed to be averted. The tollbooth’s role, after all, was to leave each route comfortably full of qualified travelers, not completely empty.

Composition courses, then, were constructed as temporary overflow lots on the side of admissions examination tollbooths. The first year composition course would maintain a flow of tuition-paying travelers down various roads to their objectively determined portions of the American Dream, without disregarding the findings (and thus undermining the integrity) of the testing tollbooth. In this way, the composition course emerged as an unintended consequence and integral part of the behaviorist infrastructure.

Herein lies the source of a great, hopeful paradox: despite its debt to behaviorism -- which dismisses if not vilifies the concept of experience -- composition instructors were positioned to be intimately involved with the experiences of their students. While most academic fields at the university were moving away from students and towards disciplinary research agendas, the discipline of composition studies was created for and because of students. Despite the troubling assumption of deficit that underlay this purpose, composition studies not only concerned itself with more students because of the universal requirement, but it also concerned itself more with the students than any other discipline outside of colleges and departments of education. While composition instructors have never had a corner on “caring” about their students, they have been uniquely positioned to take their students’ experiences into account -- in both theory and practice -- a dynamic we saw this illustrated vividly in Chapter 6.
THE PARADOXICAL BIRTH OF A SUBFIELD

So, here we have a discipline -- composition -- that owes its existence to the emergence of behaviorism, educational measurement, and social engineering. And yet, composition instructors were positioned so close to students that they perhaps could not help but make experience one the field’s central (perhaps foundational) concepts. It was from this tension that a relative anomaly was born: an academic field with a subfield devoted to assessment -- a subfield both linked to and resistant of the field of educational measurement. Writing assessment was born to ameliorate its own effects.

Outside of the field of education, there are no disciplines in higher education with comparable assessment subfields. Surely, accreditation has sparked an interest in student assessment across fields. But this focus is relatively recent within most fields, and not integral to their formation. In contrast, composition’s interest in assessment, at least indirectly, began at the beginning. The educational crisis that led to the establishment of first-year composition justified or was “discovered” by large-scale assessments. These assessments had emerged to fill a critical role in the behaviorist infrastructure. Composition has always been more than this point of origin, of course, but identifying it again here may help us to understand why, despite its best efforts, writing assessment has had so much trouble extricating itself from behaviorism.

In a direct sense, the concerns of educational measurement would enter into the discussions in the field in Braddock, Lloyd-Jones and Shoer’s 1963 report on “Research in Written Composition.” Signaling the formation of a field of study responsible for generating scholarly knowledge rather than just a motley crew of instructors responsible for improving Johnny’s writing skills, Braddock, Lloyd-Jones and Schoen’s report...
established writing assessment (or, evaluation, as it was called at the time) as a central
care. If instructors of composition were going to do more than try to improve student
writing, they would need to generate knowledge about the improvement of student
writing. And if they were going to conduct studies of writing improvement that would
move the state of research in the field beyond “dreams, prejudices, and makeshift
operations” (5), they would need serious scientific tools and procedures to gauge that
improvement. What good was a study comparing two teaching methods, after all, if there
were no way to measure the writing produced by students exposed to the two different
treatments?

Early in the development of the field, then, assessment surfaced as a major
research concern. What better way to establish legitimacy for a fledgling field than by
relying on the knowledge and methods of an established one? Braddock, Lloyd-Jones,
and Schoen called on ETS researcher Paul Diederich to review and recommend the
composition studies they reviewed in their report (3). Their report is filled with the
language of controlled experiments — subheadings under “Suggested Research Methods”
include, “the writer variable,” “the assignment variable,” the rater variable,” “the
colleague variable,” “controlling of variables,” etc (Table of Contents). And the key
terms of educational measurement — “reliability” and “validity” — show up in the report
50 times outside of references and footnotes.

In addition to the call from within the field of composition for writing assessment
as a tool of research, demands for writing assessment came hard and fast from outside the
field: from the university, from the public and from politicians. As Linda Adler Kassner
puts it, “writing is everyone’s business.” Long before Braddock, Lloyd-Jones, and
Schoen’s report -- and before the written entrance examination -- the public had been making judgments about each other on the basis of language use, using accents, dialects, and issues of correctness as a quick way to place others on a social hierarchy. And assessments of language proficiency have been political wedges in discussions about who should be included in -- and excluded from -- the political process in the United States: from literacy tests designed to exclude non-white voters to legislation about bilingual education.

Composition faculty bore the burden of these concerns, expected not only to implement literacy competency tests, but also to deal with the curricular and human fallout from them. Twenty years after Braddock, Lloyd-Jones, and Schoen’s internal call for attention to assessment, and several decades before accreditation would put assessment on everyone’s disciplinary agenda, the inaugural issue of “Notes from the National Testing Network in Writing” (NTNW) in 1982 shows a young field grappling with a myriad of external assessment demands. In the first paragraph of this first issue of Notes, Karen Greenburg, Harvey Wiener, and Richard Donovan link the establishment of the NTN to these external demands:

During the past five years, programs for assessing students’ writing competency have mushroomed at colleges and universities across the country. At the present time, thirty-five states have passed or are actively considering legislation that mandates statewide testing of students’ writing skills. The results of the tests currently in place have a decided influence on educational, administrative, economic, and political decisions affecting the lives of thousand of students. Both those who create the writing tests and those who use them must examine repeatedly the merits and the drawback of current models and methods of evaluating students’ writing abilities. (1)

Assessment, in other words, became an issue that compositionists could not avoid. This has become a trope: arguments about writing assessment often begin with the assertion
that assessment is an issue that composition instructors can or should no longer avoid (see White, ). In a humorous variation on this trope, the assessment scholar will begin his presentation by asserting that he never wanted to be an assessment specialist, or still doesn’t see himself as one. After a laugh all around, the speaker then delivers his treatise on assessment, having allied himself with the audience’s leeriness or even distaste for the topic.

Embedded in this trope is the tension -- and defensive position -- that characterizes the field of writing assessment\textsuperscript{xix}. At the turn of the century, writing assessment was taken out of the classroom to become a key mechanism in the behaviorist infrastructure, its purposes, theoretical underpinnings, and basic design blueprints set and defined by educational measurement. This version of writing assessment was then thrust back into the field of composition to be grappled with by instructors, scholars, and writing program directors whose approach to students was anything but behaviorist. As writing assessment became a research concern and then an accountability demand, composition would first draw directly on knowledge and procedures from the field of educational measurement -- and then spend most of its energy trying both to resist and reconcile itself to them. These efforts at resistance and reconciliation would create the subfield of writing assessment within composition studies, separate from but connected to educational measurement’s theories and purposes.
CHAPTER VIII
RESISTANCE AND RECONCILIATION: HOLISTIC SCORING AND PORTFOLIO ASSESSMENT

Writing assessment demands cooperation between the humanists who teach writing and the scientists who measure outcomes.

--Barbara Weaver, 1987

While compositionists had called on the knowledge and procedures of educational measurement in response to internal and external calls for assessment, that knowledge and those procedures sometimes proved distressing to "the humanists who teach writing" (Weaver 44). And so, a subfield emerged -- a subfield that would both do the work thrust upon it by the behaviorist infrastructure and resist it -- or, more recently, "reframe" it (see Adler-Kassner and O'Neill). This subfield of writing assessment is now thriving, boasting its own journals, conferences, and scholars. Writing assessment specialists within composition studies have created their own assessment procedures -- portfolios, directed self-placement, expert reader placement, and Dynamic Criterion Mapping -- that resist some aspects of behaviorism even as they accommodate its demands.

And, as signaled by Brian Huot's oft-cited "Towards a Theory of Writing Assessment" and (Re)Articulating Writing Assessment for Teaching and Learning, scholars in this subfield are attempting to articulate their own theory. In 1996, Huot described an epistemological discomfort with most writing assessment procedures, procedures he describes as "missing the mark" ("Towards..." 550). According to Huot, these procedures, and the theories that support them, come from a "positivist epistemology," the belief that we can isolate an external "'reality -- in this case student writing ability" ("Towards..." 550). A limited concept of "reliability" seems to be the real problem here for Huot. The demand that raters agree on the quality of the text
assumes that the quality of the text is an external, fixed reality that raters need simply be
trained to recognize.

In large-scale assessment, individual matters of context and rhetoric are to be
overcome in favor of producing a "true" measure of student ability whose validity
can only be established through technical and statistical rigor. These beliefs and
assumptions put enormous faith in the technology of testing, things like the
development of scoring guidelines or rubrics, the training of raters, the scores
papers receive, and the statistical calculation of interrater reliability.
("Towards..." 550)

A theory of writing assessment driven by a limited concept of reliability cannot do
justice, in Huot’s estimation, to what we know about “social construction” and “written
communication in a postmodern age” (“Towards...” 550). Thus, as early as 1996, Huot
asserted the need for a new theory:

Few important or long lasting changes can occur in the way we assess student
writing outside of the classroom unless we attempt to change the theory which
drives our practices and attitudes toward assessment. ("Towards..." 551)

Although we’ll examine later how and why Huot attached his new theory -- wittingly and
unwittingly -- to educational measurement’s purposes and terms, Huot’s resistance to
reliability brings us, at long last, to story of writing assessments’ resistances. In trying to
articulate a theory of writing assessment separate from the theory supporting “scoring
guidelines or rubrics” and “the training of raters,” Huot was resisting another of writing
assessment’s resistances: holistic scoring. Holistic scoring, championed within
composition studies by Edward White, was itself a resistance to the multiple-choice
grammar test.

Holistic scoring is as good a place as any to begin this story of writing
assessment’s resistances and reconciliations. By working our way from holistic scoring
back to Huot’s attempt to build a theory of writing assessment, we’ll be able to see
several additional resistances: portfolio assessment in this chapter; and William Smith’s “expert reader” placement procedure, Huot’s principles of contextual and local assessment, and Bob Broad’s Dynamic Criterion Mapping in Chapter 9. All of these resistances have been important, because, among other things, they have given “humanist writing teachers” ways to evade or delay educational measurement’s worst reductions. But these resistances have failed to identify and account for the concept of “experience” in assessment because, ultimately, all of them fail to fundamentally challenge assessment’s behaviorist purposes and designs. This failure is no accident, a product of scholars’ desire to reconcile their work to the institutional imperative, which has its own allegiance to the behaviorist infrastructure. Consequently, the theory Huot is attempting to construct around these resistances is at danger of collapsing in on itself because of its failure to put the concept of experience at its core.

**HOLISTIC SCORING: RESISTING THE MULTIPLE CHOICE GRAMMAR TEST**

The multiple choice grammar test as a measure of writing ability -- which had largely replaced the essay test in admissions testing by midcentury and later in placement, proficiency, and exit exams by midcentury -- was essentially the Hummer of the behaviorist infrastructure. It was the most extreme incarnation of all of educational measurement’s behaviorist impulses: its love of reliability, its distaste for subjectivity, its blind faith in correlation, its disregard for its own effects on classroom practice. In responding to the protests of teachers who worried about the multiple grammar test’s tendency to squeeze the act of writing out of the curriculum altogether, ETS helped to
assemble the assessment infrastructure’s version of the Prius: holistic scoring, a kinder, gentler, more environmentally responsible technology to navigate the meritocratic highway.

ETS conducted the studies that would pave the way for holistic scoring, a procedure that brought actual student writing back onto the table by training scorers to read essays for the same textual factors or qualities and thus producing high-enough reliability rates (Diederich, French, and Carlton, 1961; Godshalk et al, 1966). These studies have been thoroughly summarized elsewhere (Huot, 2002; Broad, 2004; Lynne, 2005; Wilson, 2006). What interests me here is the initial enthusiasm within composition studies for holistic scoring, embodied in Edward White’s 1984 article, “Holisticism.”

The rhetoric at the beginning of White’s article soared,

To proceed holistically is to see things as units, as completes, as wholes, and to do so is to oppose the dominant tendency of our time, the analytic spirit, which breaks things down into constituent parts in order to see how they work. Analytic reductionism...works very well with machinery or other objects, but less well with art forms or life forms. (400)

While White never mentioned humanism directly, he invoked it in his first paragraph by conjuring Michelangelo’s expression and defense of humanism, David.

A table leg is much the same whether attached to the table or not; my leg or that of Michelangelo’s David changes meaning drastically when detached. Holisticism argues against reductionism and denies that the whole is only the sum of its parts. (400)

Readers such as myself inclined to follow White in his impassioned indictment of analytic reductionism and celebration of humanistic holism might be unsettled by where he takes us next. There we are Florence, marveling at David’s marvelous leg, when we’re suddenly whisked away to Princeton, NJ, home of ETS. According to White,
The early development of what we now call holistic scoring took place wholly under the auspices of the Educational Testing Service, and ETS deserves considerable credit for sponsoring the research and developing the techniques which have led to the present state of the art. (401)

The present state of the art, to White, meant holistic scoring. Apparently, David human form :: ETS' holistic scoring : humanistic writing assessment. 

At first, White's inspirational tone didn't falter at this change of scenery: he declared “the advent of holistic scoring” at ETS a “manifestation” of the “this spirit” (400), explaining how this once obscure “new concept in testing and (thus) teaching writing has triumphed” (401) during his eleven year involvement in writing assessment. Before long, though, White's description of ETS' holistic scoring procedure began to sound anything but humanistic: readers were “calibrated” during “controlled” reading sessions which served to “eliminate” issues of reader preference, experience, situational context, and other such “extraneous variables” (404).

My synopsis of the holistic scoring procedure developed by ETS:

1. A “carefully developed and precise” (403) prompt is given to test takers.
2. Scoring session leaders meet before scorers gather to read a sample of the essays, develop descriptors at various performance levels, combine these with the traits or characteristics to create a rubric or scoring guide, and then identify papers to use as anchors or exemplars for each performance level.
3. Scorers practice with these papers until they reach agreement on the scores, a process that serves to calibrate readers and establish the anchors.
4. Leaders circulate during the reading to ensure reader consistency, and recalibrate readers whose scores show signs of “drifting” (405).
5. A two-point disagreement in scores between readers is solved by a third reader.

6. Records are kept to identify the most reliable scorers, who may serve as leaders for the next scoring session.

Despite White’s celebration of the wholeness of artistic and human forms early in the article, these “controlled essay readings” sounded like a procedure designed to turn readers into machines, carefully aligned for the ETS assembly line: *essays in, scores out; essays in, scores out; essays in, scores out.*

In his defense, White wasn’t thinking about readers and their readings as the organic wholes to be protected from analytic reductionism; he was looking at the essay as an organic whole. The indirect testing of writing didn’t consider the essay at all, much less view it as whole. Instead, multiple-choice technology claimed to test various sub-skills of writing, extracted from the performance of writing altogether.

In White’s view, then, holistic scoring was responsible for bringing the whole essay back onto the table for evaluators’ consideration — metaphorically and literally. Furthermore, White pointed out that holistic scoring avoided the pitfalls created by two other direct writing assessment methods: “general impression scoring,” which White worried would leave students vulnerable to the abuse of the grammar and convention nazis; and “analytic scoring,” which fractures writing into sub-skills to be considered and scored separately within a single essay.

Although holistic scoring and the scoring guides and rubrics associated with holistic scoring are widely used today, holistic scoring was challenged by several assessment scholars in the 1990’s (Huot, 1990; Williamson and Huot, 1993; Broad,
1994). The enthusiasm of White’s support for holistic scoring was matched only, perhaps, by the enthusiasm of Bob Broad’s critique of it. Between White’s enthusiasm in 1984 and Broad’s critique ten years later, another of writing assessment’s resistances had emerged: portfolio assessment. The values that led to and recommended portfolio assessment helped Broad to shed light on what he found wrong with holistic scoring, and so we’ll engage the portfolio movement and its resistances before coming back to Broad’s critique of holistic scoring.

PORTFOLIOS: RESISTING THE REDUCTION OF THE ONE-SHOT WRITING EXAM

In their earliest forms, binders, folders, and folios were probably used more for organization than for teaching or assessment purposes. In 1946, Paul Kies and Albert Kitzhaber (the two known authors of the “Syllabus for Freshman Composition…” at Washington State College) suggested that instructors could require that students use a “manila cover” to collect their themes:

For ease in handling, the instructor may wish to direct his students to hand in all papers flat, not folded. Such handling saves considerable time in grading. A heavy file-pocket with bellow edges is a convenient carrier for one or two batches of themes. Or, if the instructor chooses, each student may be directed to buy a manila cover and to bind in it all his themes for the term, handing in the entire booklet each time with the new theme on top; this has the advantage of keeping his previous work recurrently before him and before the instructor. (3)

Although the reflective, pedagogical purposes of later portfolios weren’t mentioned explicitly, the stage was set: by “keeping his previous work recurrently before him and before the instructor,” both the student and the instructor would be able to reflect on each
piece of work in the context of a body of work, noting growth and the student’s range of topic, genre, purpose, and style.

By the 1980’s, portfolios were being discussed and used for teaching and grading purposes. Portfolios were affirmed by values and concerns already in the field. Portfolios had the potential to showcase “process” even while they collected “products”: drafts, feedback, revisions, and reflections included in portfolios showed the writer at work in ways that finished drafts, submitted individually, did not. The reflective element of the portfolio encouraged students to be metacognitive about their learning. Decisions about what to include in the portfolio -- when left to students -- gave students ownership of their work.

And, significantly for a field fundamentally interested in teaching, the portfolio made the teacher’s pedagogy visible in a way that other grading or assessment systems did not. As Roberta Herter from Henry Ford High School in Detroit put it in 1998,

That long look back on a semester or a school year is so much fuller with a stack of portfolios... The students’ portfolios give me their most significant moments as learners; in some way, these moments mirror both the presence and absence of teaching. Inevitably, they become a site of my inquiry, reflection, and change. (qtd. in Tierney and Clark 477)

In addition to the reflective pedagogical potential of portfolios, writing instructors believed that the collection of work in a portfolio gave a better snapshot of student achievement and learning within a single course than the sum total of individual, graded assignments.

Perhaps portfolios had potential for reconciling large-scale assessment with compositionists’ concerns. After all, the work in the portfolio could capture students' strengths in ways that the one-shot writing exam did not, and the portfolio’s emphasis on
process, student reflection, and student ownership might have the added benefit of
couraging better large-scale pedagogy. But when portfolios began to be used for the
purposes of large-scale assessment, they presented challenges for educational
measurement -- challenges that educational measurement would solve either by changing
portfolios or dropping them altogether.

Portfolios confounded educational measurement's desire for straightforward
reliability. As Diederich, Carlton, and French's 1961 study had illustrated, getting
"esteemed readers" to agree on the quality of a single essay was difficult enough. But
imagine trying to get these same esteemed readers to agree on the value of a portfolio
containing a variety of pieces in various genres in different stages of the writing process.
And then, if your goal is to "fairly" withhold or award resources and opportunities,
imagine trying to figure out the relative worth of a large batch of portfolios when the
contents of each weren't standardized, let alone the conditions under which that writing
was produced! It was enough to send a psychometrician into an uncontrolled variable
seizure.

In a sense, the use of portfolios for large-scale assessment shifted the focus (at first)
from scoring to collection methods. When ETS put essays back onto the table for
consideration with the invention of holistic scoring, it originally took for granted the
methods to be used for collecting the essays to be scored: to avoid contamination,
writing samples would be collected under standardized laboratory conditions, including
common prompts, time constraints, and no interaction between writers. Within this basic
framework, there might be some kinks to work out: Which essay prompts were more or
less biased?; How much time was enough? But with the basic sample collection method
in place, reliable scoring was the important puzzle to be worked out. The use of portfolios for assessment, however, defied laboratory sampling techniques, bringing the complicated world of the classroom -- with all its interactions and relationships -- back into focus.

Many instructors and scholars in composition studies not only allowed the complexity of classroom writing contexts back into the assessment picture, but also celebrated it. Composition’s interest in qualitative research methods prepared it to embrace this complexity. The involvement of the researcher in the complex, “naturalistic” world of the research subject was critical to the validity of the data in many qualitative research methods. Shirley Brice Heath had made Ethnographic Hall of Fame for spending over ten years in the households of two communities, documenting the culturally-diverse versions of literacy that children from each community brought with them to school. By the early 1990’s, feminist notions of collaborative, participatory research that blurred the boundaries between the observer and the observed were an important part of conversations about research in composition; composition scholars such as Patricia Sullivan and Gesa Kirsch were exploring subjectivity, positionality, and reciprocity as givens to be explored rather than threats to be expunged.

While assessment itself wouldn’t be framed as ‘research’ until at least a decade later, there are some indications that portfolio proponents were drawing on the values expressed in discussions of qualitative research as they defended and promoted portfolios. In 1992, Kathleen Blake Yancey’s description of portfolio assessment mirrors the language of participant research:

In portfolio pedagogy, assessment is seen as a process in which all the parties are bona fide participants, and in which the person who is being assessed is more than
While Yancey’s connection between portfolios and qualitative research was made from firmly within the composition community, Roberta Camp, who worked for ETS but straddled the educational measurement, composition and K-12 educational communities, had gestured toward this connection seven years earlier.

In 1985, Roberta Camp presented a session on portfolios with Pat Belanoff, referring to the “naturalistic setting” from which some writing samples could be drawn even in large-scale portfolio assessment (Camp and Belanoff 2). Camp would undoubtedly have been reading the work of Egon Guba and Yvonna Lincoln, education researchers who had been writing about naturalistic evaluation since the late 1970’s. Camp may not yet have read Naturalistic Evaluation, published the same year as Camp’s session with Belanoff. But Guba and Lincoln had published “Epistemological and Methodological Basis of Naturalistic Inquiry” three years earlier, in which they contrasted the “rationalistic” evaluation paradigm with the “naturalistic” evaluation paradigm.

In this essay, Guba and Lincoln described reality in a naturalistic paradigm as “multiple, intangible, divergent and holistic.” Truth statements in this paradigm were “context-bound” and focused “on difference” (237)\textsuperscript{x}. Later writing assessment scholars - - Brian Huot, Bob Broad, Patricia Lynne, and Chris Gallagher -- would frequently rely on Guba and Lincoln’s contrasts: context-bound subjects rather than context-less objects, a focus on difference rather than agreement, and hermeneutic dialogue and values rather than hierarchical, standards-based judgments.

Predictably, the contextual complexity of the naturalistic setting revealed in the
portfolio posed problems for educational measurement. It wasn’t just that this relatively new science hadn’t caught up to the complex nature of real contexts. This is a critical point: The extraction of an individual’s abilities or achievements from his context was critical to the mission of educational measurement. After all, the role of educational measurement has been to generate the objective (mostly numerical) basis for awarding or withholding resources in a meritocratic system. How fair would it be to award more resources (admissions, scholarships, better classes, better chances to land a more prestigious/well-paid job) to someone whose context had set him up to achieve more? While it was never framed this way at the time, the contextual complexity of the portfolio called the meritocratic myth into question.

It was a myth so engrained that portfolio proponents themselves didn’t identify or confront it directly. But that’s no surprise. The shock value of the opening story in Malcolm Gladwell’s *Outliers* depends upon our wholesale acceptance of this myth -- that individuals deserve, through innate talent and hard work, the opportunities and resources they are awarded. Gladwell’s story begins with the most seemingly meritocratic of institutions: Canadian hockey. According to the meritocratic myth, Gladwell explains, the best hockey players, no matter their background, will rise to the top in a system that lets everyone compete but rewards excellent performance. And Little League hockey in Canada seems to be just such a system: any child can play, and, supposedly, any child has an opportunity to rise to the top if he is more talented and hard-working than the others. But Gladwell recounts the odd observation that exploded this assumption: the top Canadian hockey players tend to be born in the first three months of the year, and player rankings follow a backwards regression, with lower ranked players more likely to
have fall birthdays.

The explanation for this observation is simple -- if you can get past the idea that the top players must have been born with more talent. The difference in size and coordination between a five year-old born in December and a five year-old born eight months later is substantial. But both five year olds would play on the same Little League Hockey team. Since coaches single out more “talented” players for extra coaching and playing opportunities, older players (who generally happen to be bigger, stronger, and more-coordinated) are “made” better even as they are labeled as “more talented.” In other words, the success of premiere Canadian hockey players has more to do with their birth-month (which no one “deserves”) than anyone would like to acknowledge.

Educational measurement operates on the same myth as the Canadian hockey feeder system. Its measurements purport to isolate individual ability, achievement, or performance, and then these measurements are used to justify decisions made by society’s public institutions about who gets what. This is our basic mechanism for making the rich richer, the more educated more educated, and the influential more influential. In this way, meritocratic systems in general and educational measurement in particular perpetuate the achievement gap, touting equality and justice all the while.

The complexity of some portfolio assessment challenged the illusion of context-less performance or ability: portfolios showed the individual’s ability and achievement taking shape within the classroom context, a context that would be different for each student. Did the writing samples, produced in the classroom over weeks or even months, say more about the teaching than the student’s writing ability, achievement, or performance? This question, celebrated by portfolio proponents for its pedagogical possibilities, led to the
standardization of the portfolio’s contents for the purposes of large-scale assessment.

Despite the efforts of educational measurement scholars such as Roberta Camp or Pamela Moss, who pushed ETS to accept portfolios in the name of validity, educational measurement’s concern for reliability had made the portfolio’s complexity a liability.

And once the logic of educational measurement applied to portfolios, they became far less useful or interesting to instructors. These concerns are reflected in Barbara Grant and Sandra Murphy’s title question in 1996: “Portfolio Approaches to Assessment: Breakthrough or More of the Same?” In contrasting the concerns of curriculum developers at district, state and university levels, Grant and Murphy concluded that positivism exerted the most influence over state and university curricula, while constructivism exerted the most influence at the district level. The values of positivism influenced “…decisions about how portfolio folders will be filled,” including, “the degree to which the contents of portfolios are specified in advance as well as the degree to which the conditions for writing are the same for all students” (291).

Grant and Murphy worried that the further that portfolios got from the classroom, the more standardized and the less useful for teaching they became. And, indeed, portfolio assessment at the state level had often had the effect of stripping portfolios of everything that had made them different and appealing to classroom teachers, denying teachers and students ownership. Often, the standardization of portfolio contexts had the effect of standardizing the classrooms themselves. Once the logic of educational measurement applied to them, portfolios were fairly quickly dropped from state assessment systems: according to this logic they were costly, time-consuming, and subject to complaints about reliability.
Portfolios were mulecarts on the assessment interstate. From the point of view of educational measurement, they generally gunked up the flow of traffic. They just didn’t slow things down (in terms of the time it spent to collect and score them), but they were also a dangerous technology because they were unreliable: they couldn’t do what the behaviorist infrastructure was designed to do -- to award and withhold resources without creating any doubt about the fairness of such decisions.

PORTFOLIOS: MULECARTS ON THE ASSESSMENT INTERSTATE

From the point of view of “the humanists who teach writing,” many of whom were leery of the technologies created by educational measurement, the tendency of portfolios to gunk up the flow of traffic on the behaviorist infrastructure was simultaneously a curse and one of its great virtues. On the first hand, by gunking up the works of the behaviorist meritocratic assessment infrastructure, portfolios bought the humanists some time to make portfolios their own -- to celebrate rather than to worry about their complexity. On the other hand, if portfolios weren’t accepted by the infrastructure, the widespread support for them within composition studies would be undermined. On the third hand, as Grant and Murphy’s question (“Portfolio Assessment: Breakthrough, or More of the Same?”) illustrated, the infrastructure’s acceptance of them might require that they undergo fundamental changes in their structure, such as the standardization and reductive scoring of their contents. It would be like putting a gas engine into the mulecart -- instead of decreasing our use of fossil fuels by slowing down traffic and perhaps even necessitating an alternate infrastructure, the meritocratic behaviorist infrastructure’s acceptance of the mulecart would eventually transform the mulecart into the
Hummer it had prided itself on not being.

In the terms of my analysis, it was the infrastructure’s addition of the holistic scoring gas engine to the portfolio mulecart that Bob Broad was protesting in “Portfolio Scoring: A Contradiction in Terms.” But in Broad’s own terms, it the collision of the “context-bound” portfolio with educational measurement’s search for the “context-less” reality of the writing’s true score that prompted his spirited objection to holistic scoring.

Broad began his 1994 critique by noting the irony of attempting to quantify the portfolio, either as a whole, or piece by collected piece:

...while our work in portfolio assessment has made the ground fertile for context, choice, difference, and multiplicity in the area of students' evaluative input, we persist in stripping context and quashing difference when it comes to our evaluative output. (304)

Broad blamed the psychometric legacy of a simple concept of reliability, as Huot would two years later:

I propose that quantification and the demand for agreement intertwine theoretically and reinforce each other practically...those twin practices contradict our best theoretical insights...but also lead us to abuse our colleagues and ourselves. (302)

A strong accusation -- that holistic scoring “leads[s] us to abuse our colleagues...” Broad was talking about what happens when colleagues meet together to read student essays: divergent readings are suppressed because of holistic scoring’s demand for reader agreement. This suppression -- and the demand for agreement that drives it -- qualified, in Broad’s thinking, as abuse.

Broad’s deep concern for readers in this argument is striking. In discussing holistic scoring, White seemed interested in controlling readers, referring to the “calibration” process that leads to “controlled readings” (204). But Broad listened and
gave voice to readers' experiences of the holistic scoring session. He even conducted an interview study of participants' experiences of these sessions, noting that some of them felt coerced and abused -- not a majority, but enough to warrant our concern:

Interviewees use words like oppressed, silenced, intimidated, frustrated, angry depressed, and insulted...if these are the feelings of some of the people whom we attempt to draw into our 'interpretive community,' we might want to begin asking ourselves just what sort of community it is...(308)

White didn't disagree with Broad's contention that some teachers don't experience the controlled and controlling scoring session as professionally affirming: In *Teaching and Assessing Writing*, White acknowledged that when he spoke about assessment, he often asked participants to raise their hands to indicate either a positive or negative experience with holistic scoring sessions, noting more negative experiences than positive (291). But he used this evidence not to investigate the nature of the complaints or to condemn the practice, but to call for two safeguards: holistic scoring sessions should be controlled by faculty, not administration or some outside agency, and session leaders should maintain a degree of “sensitivity and good humor” (“Holisticism” 405). Of course, these safeguards left the assumptions on which holistic scoring is built intact: they assumed the possibility and desirability of evaluative agreement; and left quantification, its practical and theoretical twin, safe from judgment.

While scholars such as Huot and O'Neill would echo Broad's objection to the context-stripping tendency of reliability, Broad's expressed concern for the experience of readers in the assessment process (though this concern is slightly different from a concern for readers’ experience, a difference I'll return to later) is most striking. However, Broad's use of the concept “context” as a key pivot-point in his criticism of holistic scoring (portfolios show more context about a student and his writing abilities than a one
shot writing exam, while holistic scoring strips the reader’s judgment of its context) helps me to explain a shortcoming of contemporary writing assessment theory. In order to explore this shortcoming, I’ll need to critically examine one of contemporary writing assessment’s guiding principles and a foundation in the development of Huot’s new theory, *context-sensitive*, and its corollary, *locally-based*.

I want to make it clear from the outset that I think that both principles, context-sensitive and locally-based, are invaluable parts of any new theory of writing assessment. Huot himself viewed his principles as “preliminary” (“Toward…” 561) when he first introduced them, so it is no argument to call these principles incomplete. While the writing assessment principles outlined by Huot can buy oppressed compositionists time and space in which to resist bad imposed assessment, they provide little guidance for how to construct and evaluate the “locally-controlled” and “site-based” assessments we might imagine and implement. I will argue that the concept of experience, as a “principle of principles,” may provide that guidance.
CHAPTER IV
GETTING DOWN AND DIRTY WITH EXPERIENCE: VISIONS OF ASSESSMENT EMERGE FROM THE VAN-RIDE HOME FROM CHURCH

“No one can hear what you’re trying to say when you flail your hands around like that!”
--Nancy Carol Rozell Wilson

Brian Huot’s five preliminary principles for writing assessment have become a mantra among writing assessment scholars within composition studies. Huot’s principles -- that writing assessment should be *site-based*, *locally-controlled*, *context-sensitive*, *rhetorically-based*, and *accessible* -- first appeared in “Towards a New Theory of Writing Assessment” and were later published in *Re)Articulating Writing Assessment for Teaching and Learning*. They are important principles, and I have counted myself among the faithful. In fact, I literally armed myself with *Re)Articulating Writing Assessment* in anticipation of my first rubric-norming session as a high school teacher, clutching the book to my chest as I walked into the room to score eighth grade essays. I set the book prominently and pointedly on the table in front of me when the session began and pounded on it a few times when I argued with the session leader about the rubric we were required to use. Exploring why I’d smuggled Huot’s book into this norming session, why I was thumping emphatically on its cover, and why I’d later feel disappointed in the way these principles were being used by Huot and others helps me to explain the promise and limits of these principles.

**VIVE LE RESISTANCE!: MY USE OF HUOT’S PRINCIPLES**

To understand my enthusiasm for Huot’s principles, you have to understand the
context of this rubric-norming session and the nature of my participation in it. Teachers from across five districts in two counties had gathered together for a late-fall in-service under the direction of the regional Intermediate School District (ISD), which had recently begun providing our professional development. No one had consulted teachers about this shift, but we hadn’t (yet) protested: the ISD always provided great snacks. Earlier in September, I’d joined a small group of teachers assembled by the ISD. Our mission? To standardize the ninth grade curriculum and assessments across the five districts.

I was opposed to this effort. In my own district, I treasured the freedom I had to experiment with curriculum, pedagogy, and assessment. I met frequently with dedicated and knowledgeable colleagues to share assignments, units, and materials. But we also taught different sections of the same course differently based on our interests and expertise. I wanted to preserve this freedom. We can laugh at my naivete now, but I had joined this group with the hopes of stopping it.

Most of my colleagues didn’t even know this ISD curriculum group existed, so I began reporting back to them. They grew increasingly alarmed by what they saw as its intrusion on our self-governing processes. We had our own curriculum council to vet and approve all curriculum proposals, and the ISD seemed to be circumventing this. In fact, unbeknownst to any of us, our local school board had already voted to accept the ISD’s curriculum -- before it had been fully created. Our district superintendent was, after all, married to the ISD superintendent. While I was in the process of successfully mobilizing my colleagues to protest this intrusion, I hadn’t gotten much traction within this ISD group.

I had just learned that similar ISD grade-level groups had already standardized
4th-8th grade ELA curriculum and assessments across the two counties: eighth grade ELA teachers, for example, were now required to administer eleven common assessments each semester. When I saw that the eighth grade teachers were scheduled on the fall in-service day to score the common writing assessment they’d just collected, I requested permission to attend. I wanted to see what was coming the high school’s way. And I wanted to gather fodder for my objections to the ISD’s common assessment scheme, which involved rubrics.

Several months earlier, I’d been asked to write *Rethinking Rubrics*. In anticipation for the “alternatives” section I’d promised Heinemann that I’d write, I was experimenting with assessment in all my classes. But if the ISD had its way and I were forced to use their common rubrics, my experimentation with alternatives would be pointless. I’d recently stumbled upon the field of writing assessment through Huot’s book. While I hadn’t yet read it, I scanned ahead the night before the rubric-norming session. Huot’s principles -- that assessment should be *site-based, locally-controlled, context-sensitive, rhetorically-based* and *accessible* -- stood out on page 95 because they were italicized and indented, so I read them carefully. I felt like I’d hit the assessment revolution jackpot: these principles were perfect tools of resistance.

I could make a compelling argument (to myself, anyway) that the ISD’s common assessments violated all of Huot’s principles. They did not arise from needs identified by each department, so they couldn’t be called *site-based*. I didn’t consider it at the time, but the ISD could have legitimately called itself a site, too. They had needs, too: to justify the funding they’d received from the state for their professional development work with us. The data they’d get from the rubrics we’d use would help them do just that. But
since Huot’s principles didn’t dictate the boundaries of each site, I felt justified in declaring my department as the site that mattered. Furthermore, I thought I could use Huot’s principle of *locally-controlled* assessment to insist that my colleagues and I -- not the ISD -- should control the design administration, interpretation, and use of any assessments we decided to use.

More importantly to me, the ISD’s assessments didn’t seem *context-sensitive* because they hadn’t arisen from work that individual teachers were doing. Again, Huot’s principles didn’t dictate where contextual lines should be drawn, so I happily drew them around my own classroom. Even if the common rubrics mapped easily onto what some teachers in my department were doing, they most certainly were not sensitive to the context of *my* classroom instruction.

I wanted to highlight rather than hide readers’ disagreements in my classroom. I would later learn that my interest in reader disagreement was shared by those involved in distributive evaluation, which begins with the premise that readers read differently and thus “rejects ‘true’ scores and accepts individual variation in scores, rates, or other values placed on the same text” (Whithaus 1). I hadn’t yet encountered this literature, but my experience as the daughter of an evangelical minister pointed me there already. This might seem counterintuitive: evangelicals in the 70’s and 80’s are not known for their rejection of Objective Truth (as manifest in the Bible or an Essay’s True Score), or for their acceptance of diverse values. But the evangelical tradition I’d grown up in harbored a deep interest in experience. Truth, in fact, was revealed *through* it. Experience was a double-edged sword because it could sometimes mislead, but it was important to pay attention to because that’s where you might find God.
GETTING DOWN AND DIRTY WITH EXPERIENCE

An evangelical preacher like my father had to get down and dirty with human experience. Truth surely existed, but simply saying it on Sunday morning wouldn’t do the trick. The sermon needed to reveal Truth to audience members -- with all their human imperfections -- through an experience that moved the heart, the mind, and the soul. My father (whose epistemology has changed significantly over the years) is a gifted speaker. His sermons were a family affair, and I was involved in every stage of their production throughout my formative years. Every Sunday, I’d witness my father perform the composition I’d watched (and occasionally helped) him think through and pound into shape during the week. I’d study the audience while he spoke, noting nods of engagement or slumps of boredom, learning to distinguish between various reasons for a listener’s furrowed brow: confusion, disagreement, revelation. During the social hour after church, my siblings and I would make a beeline for the donut table, where we’d overhear congregants’ first assessments of the sermon.

But what followed the service -- the Family Sermon Debrief -- was what really shaped my understanding of writing assessment. The Debrief involved, in roughly equal parts, my father’s explanation of what he was trying to say and my mother’s explanation of what people actually heard. My mother was my father’s biggest fan and his best critic. While she never questioned my father’s message, she was the congregation’s self-appointed spokeswoman for How the People Experienced the Message.

The Debrief would officially begin on the 20-minute van ride home when my dad would ask whether some point he’d been trying to make had been clear. “Ken,” my mother would shake her head in irritation, “No one can hear what you’re trying to say
when you flail your hands around like that.” Or, “If you don’t give people enough time to
turn to the bible passage with you, they’re busy trying to find it while you’re reading it --
by the time you’ve moved on to your point, they’re lost.” I hear my mother’s voice in my
head even now when I ask my students to turn with me to a passage in the class text: if
anything, I wait a moment too long.

These criticisms were never based on criteria or rules for good sermons. (Except,
perhaps, when my father announced the Rule of Threes. People could only handle three
major points a sermon, he declared one Sunday afternoon -- and no surprise, that: lists
always work best in threes; three is the number of symmetry; and God Himself manifests
in three!) Instead, my mother’s assessments were most often offered in the form of
explanations about human experience, vigorously debated and constantly revised. As my
father puts it now, “Call it sleeping with your sermon. We’d all write better sermons if
we had to sleep with at least one person hearing them.” I’m not planning to turn this into
the Sleeping-With Assessment Principle (SWAP), but you get the point: assessment was
an ongoing conversation, enmeshed in relationship.

The Debrief wouldn’t end when we’d pull into the driveway. Throughout the
week, I’d listen to and sometimes participate in conversations about specific audience
members: Mrs. Jones, a new mother, had just been diagnosed with cancer, and that might
affect how her husband heard next week’s sermon on suffering. I learned to listen
rhetorically, then -- to hear how phrases, stories, formulations, and appeals had been
composed with particular congregants in mind. Before I would ever confront a pile of
student essays, I had learned powerful lessons about writing assessment: assessment
hinges on descriptions of the writer’s intent and the audience’s experience of the text; a
person's experience of the text depends upon her life experiences, beliefs, and personality; and assessment directly and indirectly shapes the writer's process of invention and revision. Assessment, in other words, has to get down and dirty with human experience.

I stepped into my first writing classroom as a high school teacher with a fully formed and well-lived version of assessment, next to which school-assessment turned deathly pale. Long before I had the terms to put it this way, I believed that reader disagreement was the glorious product of human beings' transaction with language. I believed that my classroom practices needed not only to allow for reader disagreement, but also to account for and make use of it. In other words, I didn't just want students to know that readers disagree. I also wanted them to use those conflicting readings to clarify their intentions and to make decisions.

It made no sense for me to put reader disagreement at the center of my pedagogy, and then attempt to hide it in my assessments. By exposing the reality of readers' different readings, I hoped that writing assessment could help writers construct the internalized audience that I'd grown up watching my parents invoke and address -- the "audience invoked, audience addressed" that Lunsford and Ede claim writers construct throughout the writing process. "Assessment for teaching and learning" would be no more than a ritualized nod unless I could find or create assessments that truly supported the existence and uses of reader disagreement by revealing the workings of human experience (See chapters 5 and 8 of Rethinking Rubrics).
MEANWHILE, BACK AT THE RUBRIC NORMING SESSION...

...with Huot’s tools of resistance on the table in front of me. Quite apart from my concerns about whose contexts or needs the rubric emerged from – or who controlled its design, implementation, and interpretation – I believed that no rubric could be *rhetorically-based*. My mother’s assessments had been *rhetorically-based*. She was, in fact, an audience member, responding from within the communicative context that had prompted and been created by the message. Her assessments took shape within conversations that linked the speaker’s purpose and the audience’s experience of the message. And (to round out my list of three points), her assessments actually made the workings of the rhetorical situation visible, rather than simply pronouncing judgments formulated from within it. I would later learn that many compositionists considered rubrics *rhetorically-based* when they were written to reflect the rhetorical situation of the prompt, or a genre’s conventions which had originally been shaped by rhetorical contexts (See Adler-Kassner and O’Neill; Huot, Moore, and O’Neill). But that interpretation of *rhetorically-based* looked awfully pale next to the version of assessment I’d constructed during those van-rides home from church.

My mother’s contributions to these assessment conversations were not the result of internalized and then applied criteria. She had no special training. She hadn’t studied or been taught what made a sermon or a composition good or bad. She’d never even taken a first-year rhetoric or composition course: pregnant at the age of eighteen, she’d dropped out of college before completing her first semester. She was simply fully alive – mentally and emotionally alert and responsive – within the rhetorical situation from which her assessments arose. As Renford Bambrough might put it, she claimed her
human birthright by doing, in a deep and more powerful way, something that everyone
can do: She paid attention to her experience and tried to sit in other people’s Sunday
shoes, too.

Those assessment conversations had been remarkably educative, for everyone
involved. While I’ve been emphasizing my mother’s role in them, these conversations
wouldn’t have been complete without my father, who was the only one who could give
voice to his intentions. And my four siblings and I contributed, too -- my sister Amy still
remembers the sense of glee she felt after doing a little research (pre-iphone days, when
“research” still really meant something) to discover that my father had mispronounced
“assuage” as “a-sewage” twenty one times one Sunday morning. And, while ours were
the only experiences directly informing these conversations, we conjured up others and
dragged them into the van with us, too: Did Mr. Hirsch leave the room because he was
angry about what you were saying -- he’s been into some weird stuff lately about the
morality of music written in ¾ time, what’s that all about? -- or did he just need to go to
the bathroom?

These conversations, with all their invocation of human experience, didn’t just
help my father compose the following week’s sermon. The understandings that emerged
from them -- about how people work, about how language works, and about how
language and people work together -- built the foundation of my writing and my teaching
life. It was a foundation that would prove strong enough to withstand the extraordinary
pressure that would be exerted on it by educational and assessment systems created to
deny human experience. The administrator present at the rubric-norming session would
later refer to this as my “damn stubbornness.” I prefer, “strong foundation.”
My version of rhetorically-based assessment needed to make the actual *workings* of the rhetorical situation visible -- the interplay of intention and audience experience that had animated the family van every Sunday afternoon. I’d never seen a rubric accommodate a writer’s intentions or a particular reader’s reaction, much less make it visible. After all, the rubric’s effects matched its intents -- to standardize readers’ readings and reactions for the sake of reliable scoring. It wasn’t just that the rubric most often got the criteria wrong. No matter how many times it was revised, the rubric just wasn’t as elastic as a conversation -- and the interplay at the heart of the rhetorical situation requires an elastic container.

While the ISD’s assessments would be accessible in the sense that Huot promoted -- they would be available to all “stakeholders” -- I chose to understood accessibility in a more individual, experiential, and difficult sense. Just as my father needed my mother’s help to access his audience’s experience of his message, writers needed assessment’s help to access readers’ experience of their texts. At the time, this was the only of Huot’s principles I was aware of interpreting differently than he might. But, based on what I thought he meant by *rhetorically-based* assessments (which was, admittedly, based on shockingly little), I didn’t think he’d mind.

The principles on page 105 of Huot’s book had introduced me to the field of writing assessment, which would become my gateway to the field of composition studies itself. I was (and am) indebted to them. They stood beside me and helped me speak my mind when others were telling me to sit down and shut up. It would take me years to understand how (let alone why) my swift application of these principles wouldn’t be fully supported by Huot or other assessment scholars.
When I finally read (Re)Articulating Writing Assessment cover to cover, I was initially heartened by Huot’s desire to “reframe assessment for its pedagogical value” (4), and layers of my scribbles in the text testify to my excitement as Huot begins to articulate how assessment is (or should be) “fully connected to teaching” (4). My enthusiastic exclamation points appear more and more frequently in the margins as Huot begins to do something in his first three chapters that I have tried to do in this dissertation: to separate the needs of writers from the needs for which assessment has traditionally been articulated by educational measurement and institutions. Sentences such as these are both highlighted and underlined in my dog-eared copy of Huot’s book:

"Writing papers for a grade creates a role for the student in which assessing the value of writing is secondary or moot and the attainment of a specific grade is everything. In this kind of assessment, students are accountable rather than responsible because grades come from a bureaucratic, higher authority over which they exert little or no control. (66)"

So, by the time I reached the section on page 75 titled, “Using Assessment to Teach Writing,” I could barely contain myself. I already had some ideas about how assessment could be used to teach writing, extracted from years of Family Sermon Debriefs! Finally, someone was going to say what I’d been thinking all these years! Imagine my experience, then, when I finally read Huot’s practical suggestions for “mak[ing] assessment an integral part of our pedagogy.” Admiringly, he reports, “Many teachers develop with their students scoring guidelines not unlike those used in holistic scoring sessions” (78). Assessment for teaching and learning meant that you just asked students to do to themselves what ETS had been doing to them for decades? You don’t have to imagine my experience of reading this. The marginalia on page 78, scrawled four years before I’d go to graduate school, articulates it: “Bummer!” and, “anti-climax.” And, my
It seemed that, despite Huot’s critique of writing assessment’s epistemologically naïve reliance on the scoring guides and rubrics used in large-scale standardized testing (Re)Articulating 94, they’re not a problem if they’re locally constructed and controlled. I couldn’t wrap my mind around the difference between my interpretation and application of Huot’s principles and his. Surely, he would have joined me in questioning the rubric the ISD leader was asking eighth grade teachers to use, a rubric that wasn’t context-sensitive because it had been designed by the state to score fourth-grade essays. But, instead of supporting my experimentation with rubric-free assessment in my classes and insisting that any common assessments not get in my way, Huot likely might have done several things. First, he might have asked the ISD: what, precisely, did they want to know from those eighth grade essays? Then, he might have asked the ISD to work with teachers (and, perhaps even students) to design its own rubric to answer those questions. I would have welcomed this clarification of the ISD’s purpose for this assessment, which had never been clear to me. But I still would have balked at using a rubric, designed by anyone -- my colleagues, my students, or myself.

Soon, I would meet a new ally, Bob Broad, who had used Huot’s principles to propose a slightly different and, it seemed to me, promising solution to the problem of the ISD’s common assessments: Dynamic Criterion Mapping (DCM). I would have a similar experience of Broad’s work as I would have of Huot’s (though without quite so terse or negative a judgment): I so enthusiastically wrote my own hopes for assessment into his work that I blinded myself to our differences. This time, however, working out my differences with Broad’s work would help me to circle back to Huot’s principles with a
new understanding of their limitations. “SO BAD!,” after all, is not the most helpful assessment.

I discovered *What We Really Value: Beyond Rubrics in Teaching and Learning* shortly after reading Huot’s book. In it, Broad re-conceives of the rubric as a map. He likens the traditional rubric (with its variations on the same old five to seven factors) to the Vineland Map, the earliest known cartographic representation of North America. While it was an important first step for those who had encountered the North American coastline for the first time, Broad explains, The Vineland Map would be woefully inadequate for use by those of us who *live here now*. Broad suggests that each institution needs to more carefully map its own territory on an on-going basis – an elegant metaphoric expression of Huot’s concern for *locally controlled, site-based, and context-sensitive* assessment. DCM, then, is Broad’s effort to help writing programs construct better -- more detailed and responsive -- maps of their values about writing.

DCM emerged from Broad’s concern for readers, expressed earlier in “Portfolio Scoring: A Contradiction in Terms,” and his interest in qualitative research. For a research project at “City University,” Broad spent a year listening to trios of writing instructors as they discussed first-year students’ portfolios, which they needed to either pass or fail at the end of the year. To provide formative feedback, they gave provisional midterm evaluations that students (and their instructors) could keep in mind as they worked towards the final portfolio submissions.

Because the writing program at City University didn’t use a rubric, Broad was able to hear the values that individuals brought to these conversation without its interference. He found that instructors expressed 89 distinct values about the writing they
read. He divided these values into three categories -- Textual Criteria, Contextual Criteria, and Other Factors -- and then organized values (such as, “promise” or “risk-taking”) around them. The graphic organizer that resulted constitutes his map of values at City’s University’s writing program. Broad is careful to point out that these values are only those expressed at one institution in response to this particular batch of portfolios: the process of research and mapping, which should be done regularly, is as important as what appears on any particular map.

I was so taken with the rhetorical implications of Broad’s research -- the rubrics I was arguing against looked downright silly next to Broad’s map of 89 values! -- that I was soon thumping on the cover of What We Really Value in my on-going argument with the ISD about their common assessment scheme. It would be years before I’d see the implications of DCM as an assessment practice -- and understand my dissatisfaction with it. In DCM as an assessment practice, maps of a writing program’s values (which are products of empirical research) are made available to students, instructors, administrators, and the public in order to publicize the writing program’s values and create accessibility. And -- here’s the focus of most of my dissatisfaction -- values from the map can then be used to evaluate (and even score) student work for purposes of placement, grading, certification, or program evaluation.

In other words, these maps can become rubrics. Since they emerge so much more organically from the writing program’s values, however, they avoid some of the indignities Broad notes and abhors in holistic scoring sessions. And they honor reader disagreements -- at least, in the readers who participate in the conversations that lead to the maps. Since I was looking for allies in my fight against rubrics -- not in my search
for a new program evaluation method, which I wasn’t in a position to need -- it took me a
long time to imagine myself actually participating in DCM as an assessment practice,
and, thus, begin to feel my frustration with it.

I started to catch onto the difference when Broad introduced me at a professional
conference as the only person in the world who disliked rubrics more than he did. I
laughed along with everyone else, but I thought that the humor emerged from the
difference in our personalities: unlike Broad’s, mine was sometimes, well, feisty.
Perhaps, by nature, he was incapable of disliking (or expressing his dislike for)
something as intensely as I often did.

The difference would hit home, though, when I’d read Broad’s second book, an
loved the title, which conjured up the power of the slow-food movement and then layered
it back onto Huot’s principles: *locally-controlled, site-based, accessible* xxii, and *context-
sensitive*. While Broad acknowledges in the introduction that, as Richard Haswell
reminds us, “‘Everywhere people will prefer known brands to locally grown
assessments’” (Haswell qtd. in Broad 3), he still insists that “[a]ll good assessment is
local” (xiv).

The local, organic, slow-food movement struck me as a powerful analogy for
writing assessment. The analogy provides the impetus for change by summoning horrific
images to rally against: biplanes dipping low to spew pesticide over rows of genetically
engineered corporate corn; stacks of warehoused chickens, pumped full of antibiotics to
combat the infections they’ll contract from clawing and shitting on each other; semi-
trucks crisscrossing the country trailed by the sound of air-brakes and clouds of diesel
fumes to bring the products of industrial agriculture to your supermarket shelves. And
the analogy provides the vision for change by calling up images of prosperous, principled
goodness: helmeted yuppie families of 4.2 riding to their local Community Supported
Agriculture (CSA) for this week’s basket of produce -- *what shall we do with all the
arugula this week, dear?* -- hand-picked by beaming, well-educated gentlemanandwoman
farmers who do not exploit migrant workers. In addition to providing the negative
impetus and positive direction for change, the analogy brilliantly reframes two of
assessment’s biggest bugaboos -- casting efficiency in terms of sustainability and
reliability in terms of bio-diversity.

When *Organic Assessment* showed up on my doorstep, I was primed to love it. I’d
recently read Michael Pollan’s *In Defense of Food* and watched the documentary, “Food,
Inc..” Before tearing off the shrink-wrap, then, I’d already composed a mental email to
Michael Moore, suggesting “Assessment, Inc.” as the title of his next film. I had an
opening scene ready to propose which would be as visually gut-wrenching as “Food,
Inc.”’s chicken-house segment: Moore would follow the curriculum director of a large
K-12 district whose job it was to clean anxious youngsters’ vomit from test booklets
before sealing them in plastic baggies to send to the state for scoring. I’d even planned
the outfit I’d wear -- flattering yet thoughtfully subdued -- for my interview, which would
follow the visceral punch of the opening vomit scene. Surrounded by books and plants in
my study, much like Michael Pollan’s in “Food, Inc..,” I’d provide earnest expert analysis
of the industrial testing complex.

In between delusions of grandeur, I was excited to note chapters written by
scholars I knew and whose work I admired: Linda Adler-Kassner, Heidi Estrem, and

192
Anne Marie Harrington. When I finally began reading, I was impressed in particular by Adler-Kassner and Estrem’s inclusion of students and community members in faculty conversations about good writing. It was an idea I wanted to steal. But mostly, I was surprised to see that DCM in many institutions led to fairly traditional-looking rubrics. The exception to “traditional-looking” was one graphic rubric, with each of six values forming one point of a six-sided star.

For the life of me, I couldn’t figure out how this rubric’s shape made it any different: perhaps it implied a different relationship between the factors or values? Perhaps a rubric in the shape of a list implied a hierarchy of values, while the star implied inter-connection? While I imagined that I’d much rather participate in one of the conversations that led to this star-shaped rubric than the eighth grade norming session, I still couldn’t see how the star rubric would better capture and reveal the interplay of intention and audience experience in the communicative situation than the traditionally structured four-point eighth grade rubric or the six-point MEAP rubric. And I definitely couldn’t see how students whose work was assessed with it would experience it any differently. Star-shaped or not, it still communicated different truths about reading, writing, and assessment than what I’d learned about assessment from my family’s Sermon Debriefs.

In his reflection chapter in *Organic Assessment*, Broad grapples with the tendency of DCM to produce fairly traditional rubrics. He’s clearly conflicted, and I sense him withholding judgment. He doesn’t want to enforce one version of DCM, and rightly so. Beyond the natural hesitation to critique one’s own collaborators, who are clearly all on the “right side” of the industrial-commercial/homegrown-local assessment divide, I
wonder if Huot’s principles actually got in the way of his expressed judgment. Huot’s principles, which directly inform Broad’s work, don’t provide the basis for critiquing the star-rubric. These principles clearly condemn the multiple-choice grammar test for not being rhetorically based (there’s no communicative context, writer, or intended audience in the test), and they clearly condemn the MEAP rubric for not being locally controlled. But Huot’s principles don’t give Broad a way to critique the star-rubric. After all, it satisfied them all by being locally-controlled, site-based, contextually-sensitive, accessible, and, many would argue, rhetorically-based.

I re-read Huot’s book with this hunch in mind. That’s when I finally put my finger on the cause for what I sensed to be their limits. For all his expressed concern with students and their learning, Huot builds his principles and new theory around four assessment practices that are not concerned primarily with assessment for writers. Instead, the practices he theorizes are designed primarily to serve the behaviorist infrastructure: placement, exit, and program assessment. Huot explains how his emerging theory is grounded:

“The basis for this theoretical exploration are current practices at universities who have been using assessment procedures unsupported by conventional writing assessment’s reliance on the positivist, epistemological foundations of classical test theory. It is my hope to connect these procedures through their common sets of beliefs and assumptions in order to create the possibility of a theoretical umbrella.”

Huot then describes the four local assessment practices from which he extracts his five principles: Smith’s expert reader placement procedure; Haswell and Wyche’s two-tier placement procedure, Allen’s on-line inter-institutional program assessment; and Durst, Roemer, and Schultz’s course-exit portfolios read by trios of teachers using talk (rather than rubrics) to make decisions.
While compositionists had managed to “infuse” these assessment procedures with values I share (for example, they all evaded quantification), three of the four procedures had been created to serve the institution’s gatekeeping purposes for assessment: two place students into different courses, one certifies students as proficient in course standards. The remaining program assessment serves a system of accountability that uses resources as a behaviorist lever to control the behavior of institutions and programs. Those who design such assessments for these institutional purposes must consider the needs of students -- if they consider them at all -- through the lens of the behaviorist infrastructure, an infrastructure that hadn’t been designed with their experiences in mind in the first place. While I admire the way that those involved in all four of these assessments minimize the damage done by this infrastructure, I do not find these practices a sufficient basis for a new theory of writing assessment.

Let me explain this critique in terms of William Smith’s expert placement procedure. Smith asked teachers of a particular course to read a student’s placement essay not to score it, but to decide whether or not the writer’s needs would be served by that course. I admire the elegance of Smith’s procedure, insofar as it is a brilliant evasion of scoring. In Dewey’s terms, the expert placement procedure doesn’t create a particularly miseducative experience for students. Unless they’re angry or dispirited by the class they’re placed in (and, this is admittedly a big unless), it doesn’t make them less likely to want to write. And it doesn’t tell them untrue stories about the nature of the act of reading and writing: it doesn’t, for instance, tell that them that the goal of writing is to produce a score or that a piece of writing produces the same response in all readers.

But while the expert placement procedure doesn’t tell an untrue story about what
readers and writers do, it doesn’t tell much of a story at all: it doesn’t create a particularly *educative* experience for students. It serves the institutional imperative to sort students into the proper “educational treatment,” and does this in a way that evades quantification, the worst of educational measurement’s reductions. But it doesn’t serve students’ conception of what writing does in readers’ minds.

At the time, I wasn’t impressed with the argument that the expert reader procedure served students by placing them into the course that best serves their needs. This argument rested on a thin conception of writers’ needs and a bureaucratic conception of teaching: as a k-12 teacher, I was accustomed to designing my instruction to serve whoever showed up. I didn’t have, nor did I want, the ability to reject or accept students into my classes. I couldn’t see any reason, for instance, why I couldn’t effectively teach an inter-generational group of writers who would learn with and from each other. While I join Huot in admiring these procedures, all of which cleverly evade quantification, I can’t help but note now how all four procedures accept and extend Yerkes and Watson’s vision of a society that works most efficiently when everyone is sorted – via testing – into their proper place in society. Of course, if I had to use writing assessment to place students into different courses (known in K-12 as “tracking”), I’d rather use Smith’s expert reader procedure than automated or human holistic scoring. But in light of all I hoped assessment would do, the expert reader procedure seemed a weak practice to build a *theory* of writing assessment around, which is exactly what Huot claimed to be doing.

Although I risk belaboring the point, I’ve never heard anyone make it: Huot’s theorizing does not include a single assessment practice that is *primarily* concerned with
the effects of the assessment on students’ understanding of their own writing. The assessments are not designed to help students improve their writing. While they do not directly damage it, they don’t contribute to students’ understanding of the nature of reading and writing. They do not even require that the assessment itself -- as opposed to the decision based on it -- be communicated to students. In other words, as much value as I find in Huot’s principles and the practices upon which they are based, it does not surprise me now that my concern with student and reader experience is not directly represented in them. Huot, of course, is personally concerned with students. But the assessment procedures he chose to theorize wrote students’ experience of assessment out of his theoretical picture.

For years, I resolved my disappointment (Bummer!) by ignoring it, using Huot’s principles and Broad’s research at City University however they suited my purposes. I could get away with this because I was making arguments to people who had never heard of Broad or Huot: no one would know the difference. I had an existential interest in ignoring my disappointment, too. Metaphorically speaking, Huot had been my only ally in that room. He had come as close as anyone had to giving voice to my interest in assessment. I would have felt very lonely to think that I was, in fact, alone.

I’ve only forced myself to confront my differences with writing assessment scholars such as Huot and Broad -- into whose work I’ve unapologetically read my own experiences and purposes -- while writing this dissertation. In the process, I’ve found myself obsessed with a question that does not feel in the least bit rhetorical: how in the world could I have interpreted these principles so differently from scholars I’ve come to admire? My mother died three months ago, but she appears nonetheless next to my
computer screen to answer: “Maja,” she shakes her head in irritation, “Brian and Bob never rode home in that van with us. How could they possibly know what assessment means to you?”
CHAPTER X
TOWARD EXPERIENCE AS A PRINCIPLE OF PRINCIPLES

I’ve come to understand how I misinterpreted Huot when I first read his principles, but I still don’t think he’d object to the reason I’d brought him into rubric-norming session with me: I wanted him to be the colleague you kick under the table and raise your eyebrows at when the consultant says something that you think is a pile of horse-dung. The knowing kick and glance relieve the suffering of silence, thus providing insurance against speaking your mind. And I didn’t want to speak my mind just yet, lest I get myself in too much trouble too soon. I was here, I kept telling myself, only to observe and gather information. Early in the meeting, however, a few eighth grade teachers had expressed difficulty using the rubric, originally created by the state to score fourth-grade essays. Huot’s comforting presence could no longer help me maintain my vow of silence and I couldn’t help myself.

When I finally spoke up to explain why the rubric was causing so much difficulty (it was being used out of context in addition to being badly written on its own terms), many teachers in the room nodded their heads. But the session leader quickly put the kaibosh on my explanations: these teachers had spent a lot of time collecting the writing sample, she asserted angrily, and there was no chance they going to change the rubric now. That’s why I was thumping on Huot’s cover -- it wasn’t just me who was saying this! I was insisting. Experts in the field of writing assessment were on my side!

My use of Huot’s principles in this moment points to what I consider one of their main virtues: they useful to those attempting to resist imposed centralized assessment
Huot’s principles have great value as rhetorical strategies. By insisting on an assessment’s accessibility, you have the chance to analyze it and argue that, like the multiple-choice grammar test, it is not rhetorically-based. And, perhaps most importantly, you can always draw a context around yourself that is more local and thus more sensitive to your own concerns than the context from whence the mandate issues.

Given the emergence of writing assessment within composition studies as an attempt to ameliorate the damaging effects of large-scale writing assessments on students and teachers, it is no surprise that this fledging field’s first articulation of its theory might double as rhetorical resistance strategies. (In fact, the four assessment practices Huot had theorized represented compositionists’ attempts to mitigate the behaviorist curse of quantification.) Of course, Huot’s principles didn’t work as rhetorical strategies for me: the session leader didn’t much care what experts in the field of writing assessment thought, much less if her assessment scheme was at odds with theirs. My arguments, grounded in Huot’s principles, got me un-invited from future ISD curriculum meetings^{xxiv}.

This might be no surprise to Chris Gallagher, who would likely point out that a classroom teacher such as me was the least important stakeholder in this particular assessment scene: in her position at the ISD (which defined its role as “rolling out” state mandates), the session leader was responding to the interests of much more powerful stakeholders whose “neoliberal” assessment agenda demands ever-more “data to feed the market” (“Being There...” 456). In “Being There: (Re)Making the Assessment Scene,” Gallagher argues that,

Until we disencumber ourselves from the stakeholder theory of power and rewrite the assessment scene, we will not be able to exert leadership for assessment, no
matter how mature our assessment expertise. (458)

Gallagher explains that stakeholder theory is best expressed through the work of Edward White, who calls for a “‘dialectic,’” between all those with a vested interest in the outcome of an assessment -- its “stakeholders” (White, qtd. in “Being There…” 459). In the neoliberal assessment scene, Gallagher deems this call highly problematic:

...stakeholder theory implies all interest groups are equal -- equal stakes, equal say -- and it assumes a “marketplace of ideas” in which reasoned arguments among sovereign subjects will carry the day. White doesn’t confront the ways in which neoliberalism disempowers teachers’ “subjective” views vis-à-vis the more “objective” views of technical experts. Nor is he willing to make a case for any “special interest”; indeed, he bends over backward not to privilege one over the others. But in the context of a neoliberal scene, not making a special case amounts to acceding to the current power arrangements that prioritize remote “experts” over teachers and students. Under these conditions, collaboration means little more than serving the needs of other “stakeholders.” (459)

While Huot is sympathetic to the problems with stakeholder theory, Gallagher notes that two of Huot’s formulations create similar problems. While honoring the “pains [Huot takes] to establish that assessment, properly understood, is an educational concern,” Gallagher still worries that Huot’s call for a “unified field” of writing assessment

...accedes to the traditional bifurcation between teachers as content experts and testers as assessment experts. If this construction of respective roles prevails, teachers and students will never have the primary role Huot wishes for them -- at least not in a technocratic culture like ours. (459)

In the same way that that Huot’s “unified field” of writing assessment risks silencing teachers and students because of the unequal power brought to the “collaboration” by educational measurement experts, Gallagher worries that teachers and students are still “beholden” to educational measurement experts and politicians when our validity arguments must speak to their concerns, as well. Huot describes an effective validity argument for a writing assessment as one that:
'addresses technical documentation important to those who work in educational measurement, honors political considerations important to administrative and governmental agencies, and most importantly considers the impact on the educational environment and the consequences for individual students and teachers' (55).’ (Huot, qtd. in “Being There…” 459-60)

In other words, Huot asserts that our work in writing assessment must to speak to all stakeholders. As Gallagher points out,

The phrase “most importantly” in the final clause...does not mitigate the fact that in this scenario, they [students and teachers] are beholden to the demands of these other “stakeholders,” of whom very little is asked. (460)

Gallagher discusses Patricia Lynne’s proposed solution to the problem of writing assessment's subjugation to educational measurement. In *Coming to Terms*, Lynne wants to disentangle writing assessment completely from educational measurement by creating a “‘temporary separation' during which we develop our own terms” (Lynne, qtd.. in Gallagher, 460). She suggests replacing *validity* and *reliability* with terms more in line with compositionists’ values: *meaningfulness* and *ethics*. Gallagher questions how this separation would work in practice, doubting whether compositionists’ rousing arguments for their new terms would have any impact on educational measurement, which would have carried on quite happily during their absence.

Gallagher suggests that compositionists can’t change the terministic screen, as Lynne would have us do, until our *position* in the scene changes. His proposal, then, involves “rewriting the assessment scene.” To help us think about how to do this, he draws on Burke’s distinction in *Grammar of Motives* between *dialectical* and *ultimate* approaches to conflict. In a dialectical approach to conflict (which underlies White’s stakeholder theory), competing ideas are aired “with an eye toward compromise” (461). But since no principles are given for assigning preference to any ideas in a dialectical
approach to conflict, the ideas of those already in positions of power are then privileged and labeled “common-sense.” “Compromise” in such a scene becomes a code for the subjugation of less privileged ideas.

In Gallagher’s Burkean analysis, a dialectical approach to conflict in the neoliberal assessment scene won’t work. The neoliberal assessment scene is not a free marketplace of ideas but an ordered hierarchy. Conflicts in an ordered hierarchy require an ultimate approach to conflict, with a “principle of principles” that works to order the conversation. In the neoliberal assessment scene, market-driven accountability functions already as a “god-term” or “principle of principles,” putting technocratic elites at the top of the hierarchy and students at the bottom (462). He calls for compositionists to argue for an alternative principle of principles to (re)order the conversation and rewrite the scene. The principle he proposes, which must be embedded in the logic of “the network,” is that being there matters (464).

I support Gallagher’s call for a new principle of principles, and I find his entry — being there matters — a compelling contender. There’s something intuitively satisfying about it. It neatly captures Huot’s directive that assessment must be site-based, context-sensitive, and locally-controlled, but it resonates outside the small discourse community that is writing assessment. Eight years after the ISD session leader effectively told me to sit down and shut up, I can imagine facing her squarely, and, to the universal applause of the teachers in the room, declaring, “Look, being there matters!” Of course, Gallagher isn’t suggesting this as an alternative to Huot’s principles, and he’d never claim that the session leader would have backed off if I’d just pulled out the phrase. He’s calling for the scene itself to be re-ordered.
In Gallagher’s re-ordered scene, more would change than the arguments I had offered. The god-term of accountability had already worked to order participants’ purposes and relationships in that room -- the session leader’s purpose was to use teachers’ labor to generate data that would justify the ISD’s funding to state-level bureaucrats, whose purpose in collecting data from the ISD was to justify the state’s use of federal funding to policy-makers, politicians, and taxpayers. If I wouldn’t do what I was told as the least powerful member of that scene, I could be replaced. And, to be fair, so could the session leader. So, I am intrigued by Gallagher’s suggestion that it isn’t just that we need to make our arguments in different ways: our purposes and relationships -- and the principles that structure them -- need to change, too.

Gallagher’s call for a principle of principle that re-orders the assessment scene is worth exploring. But I want to pull his call for a principle of principles unabashedly out of context, using it as a lens through which to view Huot’s principles. While Gallagher proposes a principle of principles that would change everything about what happened in that rubric-norming room, I want to use Gallagher’s principle to think about changing the conversations in a much different room: the room in which compositionists and writing assessment scholars talk amongst ourselves. In this room, people already listen to Huot and his principles. I want to propose a principle by which to (re)interpret and (re)order Huot’s principles.

In a certain sense, being there matters is already a principle of principles in our conversations about writing assessment. It is both embedded in and used to interpret Huot’s principles, in ways that are probably not even necessary to explain for the first three: locally-controlled, context-sensitive, and site-based. It may be a bit of a stretch to
see its presence in the principle of *accessibility*, but bear with me: Huot’s dictum that everything about an assessment must “be available to those whose work is being evaluated” (*Re*)Articulating 105) works to protect students -- who are, after all, *there* in the middle of a given assessment -- from the interests of those, such as ETS, who are *further away*. Ralph Nader’s campaign to force ETS to release SAT test questions reminds us that companies such as ETS have maintained their power, in part, by keeping the machinations of assessment private. (Even today, the inner-working of E-Rater, which will likely be used to score new national common core assessments in K-12, remain a proprietary secret.) By calling for accessibility, then, Huot implies that *being there*, in the middle of an assessment, *matters*.

Similarly, *being there matters* is implicit in the principle that assessments must be *rhetorically-based*. Huot and I might disagree about precisely what it means for an assessment to be rhetorically-based, but we’d likely agree about this: the rhetorical triangle itself represents a “location,” the space in which the rhetor’s message interacts with its audience, whose expectations and concerns have both shaped and been shaped by the situation and the message. Huot’s concern for the rhetorical situation, then, constitutes a concern for those who *are there*, in the place that *matters* for writing assessment, the place in which the writer and the reader interact.

At this point, my argument within this room filled with compositionists and writing assessment scholars could go one of two ways. If *being there matters* is the principle that already orders our conversations, I could push for my vision of writing assessment from within Huot’s principle of *rhetorically-based* assessment. I could argue that only readers and writers are *there*, in a rhetorical situation. And, I could argue that
what it means to be a reader in a rhetorical situation is to have layer upon layer of experiences. Your experience of a text is shaped by the experiences and expectations you bring to it. And these expectations have been shaped, in part, by your own purposes for reading (purposes which have also been shaped by your experiences) and by your experiences with various kinds of texts and rhetorical situations.

I could argue that to be rhetorically-based, an assessment would need to protect and highlight the reader’s experience of the message -- an experience influenced by genre expectations, expectations and needs that brought the reader to the text in the first place, and the reader’s experiences and values. I could use this argument to conclude that no rubric -- whether created by teachers, DCM, or students -- could capture or express these experiences. Part of me is drawn to this argumentative route. I may yet go there.

But I keep asking, of this principle of principles: why does being there matter? I accept that it does. But why? At this point, my readers won’t be surprised by my answer. In Chapters Seven and Eight, I argued that composition emerged as a discipline, in part, because writing assessment revealed widespread deficiencies in applicants’ writing that had to be remediated to keep the behaviorist infrastructure functioning smoothly. This remediation -- in the form of first-year composition classes -- put early compositionists in close proximity to undergraduates and their experiences at a time when professors from other disciplines were moving away from them. By being there, engaging in writing -- the attempt of one human being to express and share experiences with another -- early compositionists were getting down and dirty with experience, and conditions were ripe for the writing process movement. (Process as Peter Elbow reminds us, means experience.) For compositionists, then, being there led to a concern for and exploration
of experience -- this, paradoxically, in a field formed to serve a behaviorist infrastructure that denied the existence of experience as a legitimate construct at all. To directly answer my earlier question, then: why does being there matter? Because being there makes one more likely to make decisions that honor and create better, more educative experiences for others who are there, too.

And so, I would amend Gallagher’s principle of principles: being there matters because it allows us to attend to the experiences of those who are there, too. That probably isn’t pithy enough to qualify as a good principle, but it’s a start.

I want to explore a few of the implications that my entry would have for ordering conversations about writing assessment among compositionists. I’ll do this in three ways. First, I’ll connect the concept of experience to Peter Elbow’s understanding of music and how it reframes the textual criteria of “organization” in temporal-experiential terms. Then, I’ll apply my new principle of experience to the way that Broad chose to conceive of the very concept of “values.” Third, I’ll suggest how this principle could be used to evaluate the assessments we design.

**READING WRITING: AN EXPERIENCE UNFOLDING**

While he's not writing about assessment in “The Music of Form: Rethinking Organization,” Peter Elbow begins to imagine traditional textual features such as organization in the experiential sense I’m advocating. Elbow quotes Fish to point to the temporal dimensions of writing: “‘Everything depends on the temporal dimension’ (Fish 159, qtd. in Elbow 621). Elbow points out that structure in writing is almost always
conceived of as a spatial arrangement, but that writing exists in the reader in temporal
time, an experience unfolding. The time-bound nature of writing (as it exists in the
reader) suggests ways of thinking about organization that make use of music’s “time-
binding” techniques -- often, variations on “itch and scratch” or alternating patterns of
resolution, non-resolution, or partial-resolutions (623). Elbow reports on Gregory
Colomb and June Anne Griffin’s presentation of cognitive science’s findings about how
the mind makes meaning as we read:

Colomb and Griffin argue that the human mind itself has a natural tendency to go
through this process of making mental representations in order to compensate for
the fact that spatial texts are buried in time for readers. The mind tried to extract a
more or less nontemporal essence... Their not-quite-explicit theme is how readers
escape from time into space... their theme is coherence in the reader, not the text -
- 'coherence as a quality of experience' (276). (629)

Elbow is making use of Fish and Colomb and Griffin’s work to point writers to a new
understanding of organization -- binding time in order to organize a reader’s temporal
experience of reading. But what would the reader’s temporal experience of reading mean
for writing assessment? After all, it was Huot who pointed out the “obvious fact” that
assessment begins with the act of reading ((Re)Articulating 113).

Elbow is following a lead that I think has great un-tapped potential for writing
assessment. Our experience of music (in the moment, and over time) highlights the
plasticity of our experience of conventions and features of texts. This understanding of
musical experience can be mapped onto, and used to enrich, our discussion of genre
expectations, which affect our experience of texts. My friends accuse me of reading my
scholarly project into everything, but I couldn’t stop thinking of its implications for
writing assessment when I read Jonah Lehrer’s description of the first performance of
Stravinki’s Rites of Spring. Gertrude Stein attended opening night and described one
gentleman, standing in the aisle and hitting people over the head with his umbrella, so irritated was he by the dissonance Stravinski had created. Police were called to stop the rioting. Lehrer then explores neuroscience’s explanation for why that first audience literally rebelled, and why we don’t when we hear it now.

If Huot is right that assessment begins with the act of reading, and if Elbow (and Louise Rosenblatt and Stanley Fish) are right that the act of reading involves the readers experience of the text as it unfolds in temporal time, then how does this conception of reading change our conception of the act of assessment? If one of the goals of assessment is to create an educative experience for the writer -- an experience during which, among other things, the very act of assessment (and, thus, reading itself) is opened up to the student so that she can then use this conception reading to make decisions about her own writing -- then how must the act of assessment communicate this conception of a reader’s experience to the writer? My intent in this dissertation is not to answer these questions yet, but to pose them as the questions that need asking.

**WHAT WE REALLY VALUE: EXPERIENCES**

I want to demonstrate how adding experience to being there matters as our principle of principles helps us understand values in Broad’s *What We Really Value*: the resulting shift in interpretation both enriches our conversation and creates more educative experiences for writers. Without this addition, Broad (mis)labels values that readers at “City University” have about their experience of reading as values that readers have about Textual Features. This distinction may be almost indiscernible to writers who already have Donald Murray’s “internal view” of the writing process, but it is a
pronounced and crucial distinction for those on the outside trying to look in -- students. It is critical that writing assessment clarify and not add to this confusion.

*What We Really Value* represents a important shift in how writing assessment scholars talk about good writing: from *factors, qualities, or traits* of good writing to *criteria* for good writing to *values a reader has about* good writing. This is an interesting shift: while all the terms are sometimes used to mean exactly the same thing, there are important distinctions to be made, serving different ends. Talking about *factors, qualities,* and *traits* of good writing puts the focus squarely on the text: these are factors or traits of the text. This formulation is so commonsense that it seems almost silly to say it. But note that when the focus is squarely on the text, goodness in the text is reified, as are its component factors, qualities, or traits: ideas, organization, style, etc. This is a completely functional reification, as far as educational measurement is concerned: once goodness is defined in terms of things, these things can be separated out, measured, and added up again to discover the “true worth” of the piece of writing. These are the working assumptions behind the scoring guides and rubrics used in holistic scoring. The first trio of terms -- *factors, qualities,* and *traits* -- perfectly serves the demands put on assessment by the behaviorist infrastructure.

The next term, *criteria,* shifts the focus slightly away from the text. Texts have factors, qualities, or traits, but readers have criteria. Now, we’re not just looking at a text, but at a relationship: between readers, criteria, and texts. Compared to *factors of* good writing, *criteria* for good writing brings readers back into the picture, but doesn’t move towards them very far. The criteria themselves are the important things (or, the important abstractions): the criteria exist outside of the readers, like Plato’s Ideal Chair hovering
over all, guiding raters relentless in their pursuit of the Ideal Text. Good assessors, then, are readers who have internalized the right criteria. (That’s the point of the norming sessions.) In this way, the use of the term *criteria* in writing assessment can still lend itself to quantification -- in fact, ETS co-opted it for the trademarked name of their most stunning dismissal of readers to date: *Criterion*, the web-based computerized writing assessment service that relies on its similarly trademarked E-rater.

The term *value*, however, marks a significant step towards readers. Huot sometimes uses the term in *Rearticulating Writing Assessment*, but it becomes the term of choice when Broad employs it in *What We Really Value*. Like *criteria*, *value* implies a reader. But *value* is less Platonic -- more pluralistic and individual -- than *criteria*. Different readers, and different writing programs, have different values. With good reason, Broad doesn’t want to go too far down the slippery slope of the liberal “isms” -- subjectivism, individualism or pluralism. Instead of writing about an individual reader’s values, he invokes a community of readers, referring to a writing program’s values, or “our” values (what we really value). In fact, in “Portfolio Scoring: A Contradiction in Terms,” Broad sparred with Ed White over his use of Stanley Fish’s “interpretive community”: White had invoked it to justify the norming involved in holistic scoring, while Broad objected that a holistic scoring session constitutes an interpretive community “only if *community* and *value-pluralism* are mutually exclusive concepts” (306).

While Broad doesn’t condone reader calibration, he nevertheless avoids a focus on individual readers and their values in his research: his maps are an amalgam of many readers’ values. While I believe there is great value in this descriptive approach, I also suspect that when we don’t *also* focus on (describe and theorize) the individual reader,
our assessment theory and practices risk depersonalizing readers. Thus, we risk missing the role that experience plays in assessment. Individuals, not groups, have experiences. Perhaps because he is more focused on communities than individuals, Broad labels readers' values as values they have about texts, rather than values they have about their experience of reading texts.

Here's what I mean by this distinction. Broad subdivides his map of "City University's" values into three categories: Textual Criteria, Contextual Criteria, and Other Factors. Textual Criteria is the largest category in Broad's map, and he subdivides it into Textual Features and Textual Qualities. Notice the prominence of the word "textual" throughout. But deep within the subcategory of Textual Quality, Broad makes a parenthetical nod to experience, describing this subcategory as "(aspects of the reading experience)" (33). He's onto something important. I should pause to note that I’m grateful for the care Broad takes as a researcher: he includes many of the individual comments that led to his categories so that we can examine and interpret them ourselves, an opportunity I’m going to take full advantage of here. Broad has this teacher-reader's comment categorized under Textual Feature: Textual Quality,

'Cause the end has that corny, tacked on feeling: *Okay, now's the part where I have to say how important this was.* 'Cause it didn't come through in the beginning, right? (33)

I find this a remarkable comment, not because it is rare -- readers who are also writers and teachers say things like this all the time. Instead, it is remarkable to me because it conjures up layer upon layer of both the reader's and the writer's experiences in ways that resist its double-textual label. In this comment, the reader imagines the writer speaking (*Okay, now's the part where I have to say how important this was*) but gets here
from acknowledging his or her experience of the text -- *the end has that corny, tacked on feeling*. Of course, the end of the text itself doesn’t *have a feeling*: the reader has a feeling when s/he reads it. And the reader’s experience of that ending leads to a sympathetic feeling -- the reader (as a writer and teacher him or herself), remembers one of two experiences (or, perhaps, both). The reader, as a writer, too, remembers the feeling of doubt s/he has when trying to figure out if his or her point has come clearly through her writing. The reader knows, first-hand, this feeling of doubt that leads to the “corny, tacked-on” ending. Or (and): the reader, as a teacher, too, remembers the experience of talking to a student writer about such tacked-on endings, with the student’s description of this moment of doubt. Either way, this reader’s comment emerges from layers and layers of experiences.

Broad rightfully refers to this comment as “(aspects of the reading experience)” but he buries the concept of experience within layers of textual concerns, subsuming it first within parentheses, then within the subcategory of Textual Qualities which is, itself, further subsumed within Textual Features. And, with Textual Features, we’re back in the realm of *things about or things in* the text that can be seen, measured, and then quantified. In this way, readers’ values about experiences become readers’ values about texts. It is little surprise, then, when DCM turns into the star-rubric.

What would happen if we drew this reader’s comment out of the amalgam of City University’s map of values and wrestled “(aspects of reader experience)” out of their parentheses and double-walled tomb of textual categories? It isn’t that textual features would disappear as a useful term. It’s that textual features would be interpreted in light of our principle of principles -- *experience*. Textual features exist, but different readers
I was reminded of this when I taught a week-long graduate course for K-12 teachers at the University of New Hampshire’s Literacy Institutes. Partly inspired by Broad’s invocation of the Slow Food Movement in *Organic Assessment,* I’d titled my course, “Gutsy Assessment,” and my students and I spent the week exploring all the connotations of that phrase, beginning with the obvious implication that teachers who want to engage in meaningful assessment sometimes have to have “guts.” To explore the less obvious implications, I’d asked the class to read “Unhappy Meals” by Michael Pollan (based on *In Defense of Food*) to kick off our discussion of the parallels between writing assessment and the Slow Food Movement. (These parallels are striking and fun. We rewrote a list of Pollan’s sentences to illustrate: Pollan writes, “It was in the 1980s that food began disappearing from the American supermarket, gradually to be replaced by ‘nutrients,’ which are not the same thing.” We recast this as, “It was in the 1960’s that essays disappeared, gradually to be replaced by ‘the factors of writing,’ which are not the same thing.”

The morning after I’d asked the class to read “Unhappy Meals,” I began the day’s discussion by gushing about how “good” Pollan’s writing is: I love how he gives us his dumbly simple thesis up front – “Eat food. Not too much. Mostly plants” – then backs up and circles towards it again, complicating it until its truth and simplicity seem almost irretrievable. I’d introduced the text to use it, not to assess it. But since I’d started the discussion by pronouncing my assessment, Thom took my unintentional bait. “I couldn’t stand it!” he ventured. I was taken aback, and briefly considered backtracking, reminding Thom that the quality of Pollan’s writing wasn’t really the point: I just wanted to explore
the parallels between Pollan’s argument about food and various arguments about writing assessment.

But I’d been spouting off about reader disagreement all week. I took a deep breath, and said, “Okay, tell me more.” Eloquently and humorously, Thom described his experience of reading Pollan’s digressions: he felt both disoriented and impatient with them. Thom’s description of his experience prompted me to reframe my judgment (“so good”) as a description of my experience -- how I loved the feeling of stepping far away from Pollan’s point, coming back at the end with more knowledge and feeling to layer back onto it. While I preferred talking about the “moves” that Pollan had made, we could quite easily label what we were both describing as Textual Features. But that didn’t change the fact that Thom and I valued the experience we had of the same textual features differently.

This is critical knowledge for writers, but also for teachers who are assessing and responding to student writing. That discussion has the potential to help both Thom and me to respond more carefully to student writers. When reading an essay full of digressions, Thom might initially want to encourage the student to get to the point more quickly. Perhaps, however, he’ll remember my experience of such digressions, give his student a copy of “Unhappy Meals,” and say, “Look, some readers love these kind of digressions, some readers hate them. I’m not so fond of them myself, but if you want to get better at using them, check out how Pollan does it, and I’ll do my best to help you figure it out.” Likewise, I have a tendency to encourage all my student writers to write more like Pollan, Montaigne, Virigina Wolf, Tom Newkirk or any of the digressive writers I love to read. But with Thom’s experience of Pollan-esque digressions in mind, I
might make it a point to have mentors available for students who don’t want to write this way. Layering the concept of *experience* onto the principle that *being there matters*, then, reframes Textual Qualities as Qualities of Readers’ Experience and shifts our conversations about values just slightly enough to open up understandings that are as useful for teaching and learning as they are for evaluation.

**ASSESSING OUR ASSESSMENTS**

Once again, I am not yet proposing a new assessment process, tool, or procedure. That’s another book (or two). More importantly, collaborative efforts are required. I am pointing to a new direction for our efforts, guided by a principle of principles that orders our conversations and collaborations: *being there matters because it allows us to attend to the experiences of those who are there, too.*

While I am not ready to use this principle to propose alternative practices, it is not too early to imagine how we might use it to assess our current assessments. This is important: Huot’s principles don’t ask us to consider an assessment’s effects on (or in) a writer’s experience, and thus, give assessment permission to contribute to miseducative experiences. For example, imagine this scenario. Instructors in a large writing program invite researchers at ETS work with them to identify common textual features found in the essays they assign. They develop and publicize measurements for these features. Then each instructor in the program chooses freely to use this technology to grade student writing. The assessment could be considered context-sensitive, locally-controlled, rhetorically-based, accessible, and site-based. Is it a good assessment? If we’re relying only on Huot’s principles alone, then perhaps the answer is yes\textsuperscript{xxv}.
But once we employ the principle of experience, the game changes. To determine whether this locally-controlled and accessible assessment contributes to educative or miseducative experiences for writers (and everyone else who experiences the assessment), we have to ask what relevant constructs are involved in a writer’s development. Let’s look at just one. I’ve written here and elsewhere about how reader disagreement isn’t just a fact of nature to be tolerated: a writer’s understanding of it is crucial for helping her clarify her intentions and build awareness of her words’ effects—a understanding and awareness that are critical as she makes decisions (See “Why I Won’t,” “The View,” “Responsive Assessment,” and Rethinking Rubrics, Chapter 5). We barely need ask the effects of machine essay scoring on a writer’s understanding of reader disagreement. It denies all possibilities for reader disagreement as well as the need for a human reader at all. Thus, in addition to the many other problems with machine essay scoring (see “Professionals…”), those who experience it are less likely to view (and use) reader disagreement as a decision-making tool.

Few writing assessment scholars would disagree with my assessment of AES for grading or scoring student papers—in the classroom, or in high stakes assessments. But let’s shift the conversation briefly to program assessment. Most often, program assessments are not conducted to directly help individual writers to improve, but to: A) inform the writing program so that it can make productive changes; B) shape the practices of the teachers who teach in it; C) justify current funding levels or an increased resources to administrators or external bodies. My thinking about assessment unapologetically focuses on the moment(s) between those who are there in a rhetorical
situation -- the writer, her text, and its reader(s). So, what bearing does a principle focused sharply on this moment have on program assessment?

My answer starts with the acknowledgement that program assessment has a trickle-down effect on classroom practice. The public nature of program assessment -- and how it is most often used to satisfy accreditation requirements or justify resource allocation -- sends a message to its instructors about what the program values. In fact, when problems with an instructor's pedagogy are expressed, program assessment is often posed as the solution: the assumption is that assessment can be used to improve teaching and learning. This isn’t an unfounded assumption: If programs and instructors are ultimately judged by whatever assessment procedures or tools used in the program assessment, instructors naturally bring those procedures or tools into the classroom -- along with whatever views of reading and writing are embedded in them.

So, even if students are not privy to the ways in which particular essays they’ve written are read, interpreted, discussed, ranked, and judged in a program assessment, it still influences the moment between a writer, her text, and its readers. Those involved in program assessment, then, are not exempt from grappling with the experience principle. The questions shift only slightly: what kind of experiences of reading and assessment does the program assessment give to instructors? How do these experiences translate to classroom practices? And what kind experiences do students have of these classroom practices? If an assessment designed to meet programmatic or institutional goals tells us a story about the nature of reading and writing that doesn’t match the story I understand about the transaction between readers, writers, and texts, then, quite simply, I consider it a bad assessment.
This is the project I want to claim for writing assessment: the attempt to create educative experiences for writers by revealing readers’ experiences of reading in ways that help writers clarify their intentions and their sense of language’s effect. There are inherent difficulties in this project, difficulties compounded by the role assessment plays in the behaviorist infrastructure. For example, I suspect that one of the most suitable containers for the writing assessment process is a conversation between readers and writers. A conversation is well-suited to handle the ambiguity, mutually negotiated exchanges, and shifts in understanding, articulation, and judgment that are responsive enough to allow a writer to glimpse a reader’s transaction with her text (see Wilson, “The View from Nowhere” and “Responsive Writing Assessment”). Such conversations may be incapable of the creating the certainty that institutions want to justify the decisions they make about who gets (and who doesn’t get) what. (Although, as Lieutenant Furgeson’s lament reminds us, sometimes a little more conversation might prevent a bad decision from being made.)

I consider it hopeful but naïve to assert that our educative aims for assessment can coexist within the behaviorist assessment infrastructure. If we are interested in education, then, as Dewey reminds us, we have to be interested in experience. And if we are interested in the education of writers, why then, we’re interested in a triple dose of experience: we’re interested in the ways in which writers learn to use language to shape and express and create experiences -- their own, and those of their readers. The behaviorist infrastructure of assessment confuses assessment with reliable measurement in order to create the false certainty that allows us to avoid our natural concern for the experiences of those whose needs we have denied. Experience itself is rejected in this
method of decision-making, with its reductive view of human behavior and motivation designed to control individuals and groups. The effects of behaviorism’s denial of experience on education and assessment have not been much different than its effects on Watson’s children. This should be no surprise: Our educative aims for assessment and the accountability movement’s use of it are directly at odds, just as Watson and the early behaviorists set them up to be. Our hope that the two can be reconciled allows us to avoid challenging this infrastructure, before which we may feel helpless and hopeless.

My insistence that writing assessment must consider and protect experience, then, is not convenient. Neither are a human being’s emotions, mind, consciousness, or sense of purpose in the behaviorist’s laboratory. But if human experience is not worth fighting for, I don’t know what is.
BIBLIOGRAPHY


Furgeson, George O. Letter to Robert M. Yerkes. 8 Nov. 1918. Robert Mearns Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 92, Folder 1763.

———. Letter to Robert M. Yerkes. 9 Nov. 1918. Robert Mearns Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 92, Folder 1763.

———. Memo to the Camp Surgeon. 9 Nov. 1918. Robert Mearns Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 92, Folder 1763.


High School Teachers Association (HSTA) Bulletin.


225


———. Letter to Robert M. Yerkes. 21 May 1912. Robert Meams Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 51, Folder 982.


———. Letter to Robert M. Yerkes. 9 Nov. 1915. Robert Meams Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 51, Folder 986.


———. Letter to Robert M. Yerkes. 29 March 1919. Robert Mearns Yerkes Papers (MS 569), Manuscripts and Archives, Yale University Library. Box 51, Folder 988.


Wilson, Maja. "'There are a lot of really bad teachers out there.': One Educator Responds to the Invitation to Impose Her Progressive Practices on Others.” *Kappan* 92.2 (2010): 51-55.


worth noting that, in general contemporary usage, the word “morale” to “Conference on a month later when the group met a second time. The title of the group changed from “Conference on was able to “get back into the fight” shortly following several injuries and a Purple Heart. And they can suggest that it was for this high-performing NH middle school. In an important sense, high-performing designations. The “rehabilitation” demanded of SINI schools is as bad for low-performing schools as I minimize the effect of AYP calculations on low-performing schools. Low performing schools -- which most often have the least resources to serve the most needy populations -- are devastated by SINI designations. The “rehabilitation” demanded of SINI schools is as bad for low-performing schools as I suggest that it was for this high-performing NH middle school. In an important sense, high-performing schools are sometimes better prepared to handle the fallout of such designations, since they can claim (though they don’t always) that the designation isn’t a true reflection of their performance. And they can make use of community outrage that their “excellent school” has been labeled a failure. Low-performing schools, on the other hand, are often viewed by everyone as getting what is coming to them—creating an impossible environment in which to educate children.

Yerkes’ practical applications of “the new psychology” to the war effort—through his development of large scale standardized testing and his plan to “control” military morale—likely substantiated MacDougall’s fear that behaviorism would not be used to not to further human interests, but to control individual humans in the interests of political and economic power.

In April of 1918, Yerkes was called to an “Informal Conference on Morale” with the Assistant Secretary of War and the Chief of Intelligence to apply the work of psychology in creating a “systematic plan for stimulating and sustaining morale of troops...” (Report of Informal Conference on Morale). It is worth noting that, in general contemporary usage, morale connotes a happy (or unhappy) individual emotional state. But in the first half of the twentieth century, the military definition of morale emphasized collective action (behavior): “the psychological forces within a combat group that compel its members to get into the fight” (Grinker and Spiegel, qtd. in Manning) In this formulation, “psychological forces” may or may not have anything to do with emotions, much less happiness. What matters is group behavior: if the group is compelled to action, its morale, by definition is high. If it hesitates or refuses to get into the fight, its morale is low.

Of course, as MacDougall might point out, the experience of the soldier whose morale is in question matters greatly. MacDougall had treated victims of “shell shock” in the British army during World War I. Unlike some of his colleagues who used “disciplinary” treatments, which were “behavioural” -- “electric shocks, shouted commands, isolation and restricted diet” -- MacDougall’s treatments followed “psychotherapeutic lines,” emphasizing recalling the traumatic experience and discovering its individual meaning to the patient (Howorth 226). Not only should the experiences of the soldier get him “back into the fight,” but they should also help society figure out if the war is worth its experiential and psychological toll. But that toll -- for instance, the years of depression, anxiety, and nightmares that my grandfather suffered after serving in WWII -- would mean nothing in the behaviorist’s schema of morale, since my grandfather was able to “get back into the fight” shortly following several injuries and a Purple Heart.

But Yerkes’ behaviorist influence on the American military’s discussion of morale would be seen a month later when the group met a second time. The title of the group changed from “Conference on Morale” to “Conference on Control of Morale” [emphasis mine]. Yerkes’ report on the “Scope of the

\[1\] I should note the apparent strangeness of Latour’s choice of the word “trick;” it might seem to imply that there is something trivial or deceptive about his suggestions for making objects speak. But I understand this choice in two ways. First, Latour writes with a wicked sense of humor -- for example, in 1997, he announced that the name Actor-Network Theory should be recalled on the basis of four things about it that don’t work -- the word “actor,” the word “network,” the hyphen, and the word, “theory” (“On Recalling ANT” 2). Several years later, he changed course, asserting that Actor-Network Theory is “a term so awkward, so confusing, so meaningless that it deserves to be kept” (Reassembling 9) In the context of his playfulness, the word “trick” comes as no surprise. Second, I understand the word in the context of an inventive way of solving a problem not obvious to those generally confronted with the problem: The trick to getting the ketchup out of the bottle is to tap the neck gently on your wrist as you hold it at a 45 degree angle above your fries.

\[2\] Alfie Kohn calls this dynamic, in which a practice or policy with no pedagogical value is applied to younger and younger students, BGUTI thinking: “Better get used to it!” (“Getting Hit on the Head...”). By pointing out the difficulties encountered by high-performing schools in meeting AYP, I don’t mean to minimize the effect of AYP calculations on low-performing schools. Low performing schools -- which most often have the least resources to serve the most needy populations -- are devastated by SINI designations. The “rehabilitation” demanded of SINI schools is as bad for low-performing schools as I suggest that it was for this high-performing NH middle school. In an important sense, high-performing schools are sometimes better prepared to handle the fallout of such designations, since they can claim (though they don’t always) that the designation isn’t a true reflection of their performance. And they can make use of community outrage that their “excellent school” has been labeled a failure. Low-performing schools, on the other hand, are often viewed by everyone as getting what is coming to them—creating an impossible environment in which to educate children.

Yerkes’ practical applications of “the new psychology” to the war effort—through his development of large scale standardized testing and his plan to “control” military morale—likely substantiated MacDougall’s fear that behaviorism would not be used to not to further human interests, but to control individual humans in the interests of political and economic power.
Problem," frames the problem in behaviorist terms, citing a “great variety and complications of conditions affecting morale” [emphasis mine], asserting that “the problems are in the main those of human behavior” [emphasis mine] and so the appropriate person to study such problems is one “who has the ability alike to predict reactions and to properly relate methods of control to military requirements and needs” [emphasis mine]. In other words, Yerkes framed the problem of morale as a behavioral one, offering the behaviorist psychologist as its solution.

The group consciously drew on the German system of propaganda as a model for their recommendations, viewing morale as a lifelong process of patriotic conditioning. Approvingly reporting the Germans’ use of school as a tool of propaganda and their use of “furloughs and rewards” (positive reinforcements) with soldiers, the group began to plot a comprehensive system to control of morale from the ground up. Yerkes credited his work with this group, and the multiple-choice test he devised to sort and reward recruits with promotion, with helping to win the war (Gould 224).

Yerkes’ application of behaviorism to the military may have helped to win the war, but MacDougall implied that “human effort” would involve the effort to end war. Even if the majority agrees that the war is a good cause, the experiences of those actually participating in the war cannot be dismissed unless, as Watson’s behaviorists held, their behavior is all that matters. If it can actually be attributed to him, Yerkes’ success in controlling soldiers’ morale by focusing solely on their behavior and the conditions shaping that behavior likely confirmed MacDougall’s fear: behaviorism at the expense of mind and consciousness is paralyzing to human effort.

Watson would leave the imprint of behavioral psychology on advertising (and generations of consumers) through his work for the J. Walter Thompson Company, first as a door-to-door salesman, and, later, as a vice-president of the company. In Mechanical Man: John Broadus Watson and the Beginnings of Behaviorism, Kerry Buckley argues that before 1910, advertisements emphasized rational appeals to consumers (138). Watson used his behaviorist techniques to condition consumers to associate products with emotions:

Advertisers, [Watson] cautioned, must always keep in mind that they are selling “more than a product.” There are “idea[s] to sell — prestige to sell — economy to sell...It is never so much as dry, solid, or liquid matter”...in one carefully controlled experiment funded by the J. Walter Thompson agency, Watson found that smokers with definite brand preferences could not distinguish one brand of cigarettes from another. This reinforced Watson’s conviction that the marketing of goods depended not on an appeal to reason but upon the stimulation of desire. (139-41)

Watson’s large-scale experimental test on consumers (i.e. — advertising campaigns) was an extension of his test on eight-month old “Little Albert.” Just as he conditioned Albert to associate the rabbit with a loud noise, Watson encouraged advertisers to condition consumers to associate the product with prestige and love, or to associate the lack of a product with fear and rage. While Watson claimed to have been capable of reconditioning Little Albert back to a state of fearlessness (his mother removed him from the study before he could do so), there would be no attempt to “recondition” the public back to their senses. The application of behaviorism to advertising, which took place after MacDougall’s debate with Watson, would confirm MacDougall’s fears that the application of behaviorism would be useless in terms of serving “human purposes.”

Watson has been called the first pop psychologist, spreading his message through the mass media: radio broadcasts and articles in newspapers and popular magazines (Buckley, “Behaviorism”).

Of course, writers and artists continued taking up this question, even when American psychology was unwilling to ask it (and unfit to answer it).

These human computers would soon be replaced with mechanized computers. In fact, Feynman staged a competition in Los Alamos that symbolized the tipping point between the efficiency of human and mechanized computers; during Feynman’s competition, the mechanical computers outstripped the human computers within two days, who tired after the first burst of enthusiasm (Skinner 103). But mechanizing this work wouldn’t bring a sense of meaning and purpose back to all workers’ lives. Consider the computers we now celebrate for doing our grunt work. Someone still does the grunt work — in the case of Apple computers, it’s a worker in a factory in China with a horrible safety record and a string of work-related suicides (See Duhigg and Barboza).

This isn’t to suggest that, on the whole, human experience was the foremost consideration in the skilled artisan model, which included the labor of plenty of slaves and indentured servants. It is to suggest,
The particular ways in which the industrial included its own large-scale dismissal of the experience of its workers.

My portrayal of Eliot's conception of efficiency here doesn't necessarily match the reputation that Eliot would earn as a key player in the efficiency movement's transformation of American education, particularly through his implementation of an elective credit system at Harvard and his toubles over composition class sizes with Barrett Wendell (Newkirk, "Politics"). But the concept of efficiency developed in his 1908 book is very different from what we typically associate with the efficiency movement, demonstrating that different interpretations were possible before Taylor's narrow definition that would remain unchallenged by behaviorist psychology.

Thorndike used the term "impulse" reluctantly, noting sadly that there was no better. By "impulse," he meant the "direct feeling of the doing," not the thought of oneself doing something.

John Dewey's learning theory put the experience of the learner at its center (See Experience and Education). But Thorndike was a significant figure in both the pedagogical and educational measurement worlds. While Dewey's progressive version of education, with its explicit focus on "experience" as its key term, had a real impact on American education and has never completely gone away (and we'll see Dewey's impact on composition pedagogy later in Chapter 7), Dewey's philosophy of experience in education never had any impact on educational measurement. As Linda Rief's story illustrates, the power vested in educational measurement by the accountability movement has the effect of remaking pedagogy in its own theoretical image. In other words, Dewey's attention to experience doesn't stand much of a chance when Watson's behaviorist exclusion of experience is embedded in the accountability movement and given political and even legal power.

Dewey was not, however, opposed to bringing industrial work into the classrooms, as long as it was used as a point of entry for students into the history of the entire process of production, and the social uses to which the product was put. In Schools and Society, Dewey provides an extended example of a classroom study organized around the production of wool and cotton. On entering the classroom, a visitor might see children, sitting at looms and weaving. But, as Dewey pointed out emphatically, the purpose of such a moment was not to prepare children to work in a textile factory, or even to teach children to make something. After describing the study that preceded and followed the moment at the looms, Dewey concludes:

The aim is not the economic value of the products, but the development of social power and insight...you can concentrate the history of all mankind into the evolution of the flax, cotton, and wool fibers into clothing...(36)

In fact, this example gives us a clue to how Dewey resolved a conundrum: his concern for the meaningless work that he saw as enslaving workers in the factory on the one hand, and the industrial reality into which most students would enter upon leaving school on the other. Dewey didn't necessarily envision the work itself changing. Instead, he hoped that well-educated workers would bring their knowledge of the social and historical significance of that work to the job, creating meaning that would have been unavailable to them without their schooling. In effect, he was trying to create intellectual artisans, even if the work they did was fractured and industrialized.

Five years before Carl stormed out of her classroom, Amy worked in a district where "program fidelity" was monitored and enforced. Amy describes being handed boxes of overheads and worksheets at the beginning of the year, all of them designed to maintain the school's excellent test scores. One day, sensing her students were tired of what she considered to be a terribly tedious curriculum, Amy substituted a short story she thought would interest students. Walking by her classroom and noticing no overhead on her screen, the principal pulled Amy out of class, asking her what she'd done with the materials she'd been given to teach. The message was loud and clear: deviation from the test-prep program was not tolerated. Amy moved back to the school I taught at shortly after, where she was given more autonomy—except during MEAP week.

Most proponents of introspection were far more nuanced about its possibilities than Watson. Even Wilhelm Wundt, Titchener's teacher, whom Watson retrospectively named the founding father of introspection, believed that introspection had limited value in cultural psychology since a single, introspecting person cannot know the development of human thought that preceded the moment of his thinking (Costall 645). Similarly, social psychologists such as Margaret Mead, Baldwin, and John Dewey—all of whom valued individual reports of experience—had a nuanced and complex understanding of the issues involved in "private" experience. In addition to his claim that introspection changes the
feeling observed, Dewey reminded the Observer that his introspection was not a retreat to a wholly private place. Instead, as Dewey put it, the Observer listened to his own “soliloquy,” which was, itself, the “product and reflex of converse with others” (Dewey, qtd. in Costall 642).

Only later, because of common misunderstandings of these terms, would Freud use the familiar German terms (I, Over-I, and For-Itself) that would be translated into the English as the mysterious id, ego, and superego.

Surely, as Tom Newkirk points out in “The Politics of Intimacy: The Defeat of Barrett Wendell at Harvard,” we can’t draw a solid line in pedagogical history, with the Dark Ages on one side and the Enlightenment (including the process movement) on the other. While James Berlin had characterized Wendell’s teaching practice as current-traditional, with a positivistic emphasis on correctness (116), Newkirk found that Wendell’s teaching practices bore a strong resemblance to those advocated by Murray et al: daily writing, writing about experience, conferencing, and regular discussion of student texts.

While Wendell’s practices may have been similar to those advocated by writing process practitioners, my reading of Wendell’s composition textbook, published in 1891, indicates that his theory of composition pedagogy—and his method for generating this theory—differed significantly from Elbow or Murray’s. Like Murray or Elbow, Wendell introspected. But, unlike Murray or Elbow, Wendell did not introspect in order to build a theory of writing for himself—a process that students could then replicate in order to build a theory of writing for themselves. Instead, Wendell introspected as he read great works of literature in order to develop “general elastic principles” of good writing that are “observed by thoroughly effective writers” and should be followed by students (Wendell 2-3). For example, to illustrate the principle of good style, Wendell movingly describes how Robert Browning’s writing style in “Grammarian’s Funeral” (8-11) affects him. But his introspection served the purpose not of gaining insight into how writers work, but to extract textual principles—in this case, the principle of coherence—that students should observe.

In the first few years’ worth of “Notes” from NTNW, we can see the dismay and confusion involved as compositionists tried deal with educational measurement’s intrusion on their writing programs in the form of standardized writing competency, placement, entrance, and exit exams. From 1982-1985, entries in the “Notes” show attempts both at resistance and reconciliation: concern for the effects of writing tests on curriculum and students; resolutions to resist the worst aspects of the tests; attempts to use the concepts of educational measurement against itself; calls for compositionists to design the tests that would be used to sort, rank, certify, and track; calls for compositionists to insist on limiting the uses of tests; and calls for tests that can be used for both pedagogical and institutional purposes (Metalene, 1982; Herrington, 1982; Lloyd-Jones, 1982; Weaver, 1982; White, 1982; Broderick, 1983; Chew, 1983; Suhor, 1984; Quellalz and Stiggins, 1985).

Similarly, in 1987, Barbara Weaver’s review of three influential books about writing assessment returns—as the books themselves do—to this tension between resistance and reconciliation. Illustrating the urge towards resistance:

In Bruffee’s words, ‘A test is a bludgeon in a profession that advances, if it advances at all, only by strokes and pats’ (Writing Assessment 97) (qtd. in Weaver 44)

And,

‘most testing of writing is poorly done, destructive to the goals of teaching, and improperly used (2) (White, qtd. in Weaver 40-41)

Illustrating the urge towards reconciliation:

Writing assessment demands cooperation between the humanists who teach writing and the scientists who measure outcomes (44).

And,

While they are concretely aware of ‘pitfalls’ in writing assessment, contributors to this volume support its practice and urge refinements in methods (40).

Several years later, Guba and Lincoln would refer to four “generations” of evaluation rather than two paradigms, labeling “fourth generation evaluation” as “constructivist” rather than “naturalistic:” they worried that “naturalistic” implied a realist ontology that they specifically rejected.

I do not mean to imply that “simply” being fully alive and responsive within a rhetorical context (or any context) is a simple matter, more’s the pity.
One of the virtues of the slow and local food movements is its insistence that we know where our food comes from -- that the agricultural and labor practices that led to it are, in other words, accessible.

Many WPA’s would claim that program assessment serves their goal of program improvement. I don’t dispute that they find ways to do this. But program assessment, as it is currently constituted, did not arise from WPA’s expressed needs: it follows the behaviorist infrastructure’s use of assessment to control individuals and institutions by withholding and awarding resources.

To read the happy ending to this momentarily depressing story, see chapter 5 of Teacher Leadership in Assessment, by Chris Gallagher and Eric Turley.

Notably, Huot himself has argued for the use of machine essay scoring in some situations (“RE: Machine Scoring”) and declined to sign “Professionals Against Machine Scoring of Student Essay in High-Stakes Assessment” -- a petition signed by writing assessment scholars such as Edward White, Bob Broad, Bill Condon Richard Haswell, Les Perlman, Peggy O’Neill, Chris Anson, Chris Gallagher, etc. His basic argument is: if you’re going to score essays, why not use machines? I accept this as an ironic point, but he’s not arguing ironically.