Institutional Cunning: Writing Assessment as Social Reproduction

A dissertation submitted to
Kent State University in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

by
Jeffrey W. Perry
December 2008
Dissertation written by
Jeffrey W. Perry

B.A., University of Louisville 1994
M.A., University of Louisville 2000
Ph.D., Kent State University 2008

Approved by
Dr. Brian Huot, Chair, Doctoral Dissertation Committee
Dr. Pam Takayoshi, Member, Doctoral Dissertation Committee
Dr. Sara Newman, Member, Doctoral Dissertation Committee
Dr. Rebecca Chism, Member, Doctoral Dissertation Committee
Dr. Steven Brown, Member, Doctoral Dissertation Committee

Accepted by
Dr. Ronald Corthell, Chair, Department of English
Dr. John R. Stalvey, Dean, School of Arts and Sciences
Table of Contents

List of Tables ........................................................................................................ iv
Acknowledgements ............................................................................................... v
Chapter One: An Introduction to the Problem and the Study ................................ 1
Chapter Two: Review of Literature ...................................................................... 25
Chapter Three: Directed Self-Placement and the Experiences of Educational Assessment … 61
Chapter Four: The ACT Writing Exam at CUNY: A Technology for Reproduction .... 96
Chapter Five: Towards an Understanding of Critical Validity Inquiry ................. 132
Appendix 1 ........................................................................................................... 165
Appendix 2 ........................................................................................................... 173
References ............................................................................................................ 174
List of Tables

Table 4.1 SAT Scores and Family Income .............................................. p. 108
Table 4.2 SAT Scores and Ethnicity ..................................................... p. 111
Table 5.1 Critical Validity Inquiry Model .............................................. p. 152
Table 5.2 CUNY Enrollment/Ethnicity ................................................. p. 157
Acknowledgements

I would like to thank the many kind people who entered into my life at Kent State University and sacrificed their own time and energy so that I could realize my professional goals. First, to Brian Huot, who, as friend, professor, director, and mentor, shaped my intellectual world in a way that allowed me to filter my political views into the world of academic scholarship. Your introduction to the powerful technology of assessment not only awakened the scholar within but rekindled the fire of the writer that had faded with the passage of time. To Pam, you made me feel smart when the world was telling me something altogether different, and you promoted my sense of professionalism and allowed me to become a member of the academic community to which I always felt removed. To Sara, our early morning talks about cunning intelligence and Gould’s genius are missed each and every morning. My friends, a motley collection that could only be gathered under the umbrella of graduate school, Chad, Christa, Dayna, Dirk, Hazel, and Perry, thanks for all the memories. To Dawn Lashua and Virginia “Genny” Dixon, they call you secretaries but they should call you saviors. Thanks for all the little things you did that allowed me, and others like me, to concentrate on our own selfish endeavors. And to Erin, my best friend and the mother of my precious little boys, who has moved with me from one end of the country to the other, leaving behind her parents and friends. I hope that in the end we can say it was all worth the trouble. I love you now and always.

This dissertation is dedicated to my father and my sons. To Dad, thanks for giving me the opportunity that you never had, I am grateful. To Satchel and to Cassius, may the achievement that this piece of scholarship represents help enrich your lives in the way my father’s work enriched my own.
Chapter One

Scope of the Study

The field of college writing assessment has traditionally had two major roles that present their own unique challenges. One of those roles resides in the need for colleges and universities to place large numbers of students in the proper writing class, while the second challenge resides in the institutional requirement that teachers evaluate individuals at the end of a particular course. Although these two challenges, one about placement, the other about grading, seem quite different, they are both linked by the universal demand that writing assessments achieve reliability while being cost efficient. The answer to the dilemma of reliability and efficiency has invariably rested in standardized forms of assessment.

The history of writing assessment is linked with the history of standardized testing (Camp, 1993; Elliot, 2001; Huot, 2002; White, 1993; Williamson, 1993; Yancey, 1999; Huot & Neal, 2006) and this connection has created a lasting tension. The tension is between those dedicated to achieving high levels of validity and those institutions committed to placing large numbers of students efficiently. Like standardized tests, large scale writing assessments are institutional devices that perpetuate socio-economic divisions of power. Standardized forms of assessment, I argue, are a form of technology employed by institutions to recreate a social order that is most congenial to that institution’s survival. Taken as a whole, a capitalist nation such as the United States of
America uses standardized testing, and, more specifically, standard writing assessments to reproduce the means of production\(^1\).

When viewed as a technology (Madaus, 1993; Huot & Neal, 2006) the role of writing assessment in social reproduction becomes apparent. As Madaus explains:

much of present technology is specialized arcane knowledge, hidden algorithms, and technical art; it is a complex standardized means of attaining a predetermined end in social, economic, administrative and educational institutions (Ellul, 1990; Lowrance, 1986). Testing, embedded in our system of education with its arcane psychometric underpinnings, clearly fits this definition of technology. Testing also fits some very simple definitions of technology—the simplest being something put together for a purpose to satisfy a pressing or immediate need, or to solve a problem. (pp. 12-13)

\(^1\) The reproduction of the means of production refers to Marx’s statement that “no production is possible which does not allow for the reproduction of the material conditions of production: the reproduction of the means of production” (Althusser 86). Louis Althusser, however, draws a distinction between “the productive forces” (materials) and “the existing relations of production” (socio-economic or class relations). The socio-economic relationships that make production possible (class relations) must also be reproduced.
Approaching writing assessment as a technology begs the question, “What problem is the technology of assessment trying to solve?” Martin Nystrand, in “The Social and Historical Context for Writing Research,” reveals that “problem.” Nystrand cites Time’s “Bonehead English” article of 1974 that decried the “sharp increase in unprepared students,” and Newsweek’s 1975 cover story “Why Johnny Can’t Write,” which “blamed the public schools for neglecting ‘the basics’” (p. 15), as examples of a series of successive crises that traces back to the 18th century. The real problem, I argue, is that the changing population of students who inhabit the schools of America has always created problems for those interested in maintaining their socio-economic dominance, and the technology of writing assessment has been crafted to solve these problems.

The use of large scale assessments (Elliot, 2005; Gould, 1996), in general, has been a successful technology for maintaining this dominance: literacy requirements for voting staved off the threat posed to white landowners by freed slaves; intelligence quotient (IQ) tests bolstered Eurocentric claims of racial superiority and served to explain a social hierarchy as a natural state; ACT/SAT entrance examinations continually punish minorities and lower class students; and, the NCAA’s Proposition 16 targets and punishes African-American male athletes at an alarmingly high rate\(^2\). A historical look at \(^2\) TestFair.org, citing NCAA research, found that NCAA data on student-athletes’ academic performance prior to 1986 implementation of Prop 48 reveal the discriminatory impact of these rules. The data, reanalyzed by the McIntosh Commission on Fair Play in Student-Athlete Admissions, show that had Prop. 48 been in effect in 1984 and 1985, it would have denied full eligibility to 47% of the African American student-athletes who
standardized testing reveals the insidious nature of an assessment practice that sacrifices a high level of validity for reliability and efficiency. By employing critical theories of reproduction to both standardized writing assessments and other placement methods that sacrifice high levels of validity for efficiency and reliability, like directed self-placement (DSP), the consequences of such assessment practices can be better understood.

A common definition and standard for validity has been elusive for writing assessment theorists, but the educational measurement community and other associations have been in agreement for sometime. The *Standards for educational and psychological testing* (1999) explains that statements about validity should refer to particular interpretations and uses. It is incorrect to use the unqualified phrase “the validity of the test.” No test is valid for all purposes or in all situations. Each recommended went on to graduate, but just 8% of the white student-athletes. More recent NCAA research shows that the test score requirements disqualify African American student-athletes at a rate 9-10 times the rate for white students.

3 Directed Self-Placement is a method of placing first year students into writing courses. DSP places the decision in the hands of the student. The original DSP program provided students with written and oral descriptions and expectations for a remedial course and general first-year writing course. Students were then asked to make a decision based on a self-assessment of their writing skills. There have been several modifications to this method, but, as a whole, DSP centers on student decision-making.
use or interpretation requires validation and should specify in clear language the population for which the test is intended, the construct it is intended to measure, and the manner and context in which test scores are to be employed. (p. 18)

In the field of writing assessment, specifically, scholars (Williamson, 1994; Huot, 2002; Gallagher, 2007) have also argued for a unified theory of validity to encourage assessment practices that include the social and individual consequences of decisions based on test scores. These scholars often cite the work of the Samuel Messick (1989) of the Educational Testing Service who defines validity as “an integrative evaluative judgments as to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p. 5). Operating within the parameters of validity theory, it becomes clear that decisions based upon standardized writing assessments have a very low degree of validity for a variety of reasons. Chief among these reasons is the fact that the social nature of writing is altogether ignored. As Stephen Witte and Roger Cherry (1998) note, “reflecting on the extent to which and the ways in which acts of writing are embedded in larger social practices… problematizes some underlying assumptions that guide assessment practices” (p. 73). Witte and Cherry argue that the basic construct of writing assessment, “writing ability,” is problematic because direct forms of writing assessment fail to provide an opportunity for writers to display their true writing ability. The fact that writers write for different reasons in different ways is neglected by standardized writing assessments. As Witte and Cherry argue, the idea that writing
prompts, timed writings, and controlled writing environments adhere to contemporary theories of language and writing is clearly wrong because

the ‘contexts’ specified in writing assessment prompts are both ahistorical and acontextual. Rather than providing information that facilitates text production, ‘contextual variables’ named in writing assessment prompts actually deprive writers of the opportunity to shape contexts of production and use. (p. 76)

Witte and Cherry discuss the impact that the measurement community has had on writing assessment, and the tension between the need for standardization for measurement purposes and the attempt to achieve validity in accord with theoretical views of writing scholars. This tension is clearly explained by Pamela Moss. She writes:

--assessments need to be standardized to some degree. Standardization refers to the extent to which tasks, working conditions, and scoring criteria are similar for all students. Emerging views of literacy, however, suggest the need for less standardized forms of assessment to support and document purposeful, collaborative work by students and teachers. This result is the tension between competing validity criteria that simultaneously advocate standardization and purposeful, collaborative activity. (p. 110)

Moss’s (1994) article reveals a historical moment when educational testing and writing assessment scholars were working together. The wave of test-based educational reform typified by No Child Left Behind, the ACT writing exam, and a vast array of other
standardized assessments are harbingers of a deteriorating collaborative effort. The need, as Brian Huot explains in (Re) Articulating Writing Assessment (2002), to “begin a reflective inquiry to examine the problem with the practices we now use in assessment and to suggest practices that are more consonant with our theories” (p. 63) is as relevant now as when he wrote those words, but the current trend in which external forms of assessment ignore teachers and writing scholars is evident. The real problem for scholars and teachers lies in the cultural dominance of testing and its ability as an external assessment to hold power over our educational practices via curricular control.

Historically, standardized tests have been used (Gould, 1996; Elliot, 2005; Hanson, 1993; Shohamy, 2001; Sacks, 1999) to marginalize, compartmentalize, and organize individuals. Allen Hanson (1993) writes, “regardless of people’s feelings about them, qualifying tests are a key factor for living successfully in contemporary society” (p. 186). He notes that an individual might “turn off on tests,” but he stresses that “but then the system turns off on them. They are excluded from educational opportunities and good jobs and they never are able to accomplish much” (p. 186). This exclusion reveals the power of testing as a force for social reproduction. Those who opt out of standardized testing, those who participate and succeed, and those that participate and fail are all tied to those scores in very important ways. Is it conspiratorial to believe that all

4 One of the particular cases that this dissertation will analyze is evidence of this statement. The writing program at CUNY of Kingsborough had the ACT writing examine imposed on its students and teachers by city and state policymakers who ignored scholars working inside the writing program like Barbara Gleason.
of this money is spent on testing our students in public education only to recreate the social inequality that marks members of our society through difference rather than a shared humanity? Harold Berlak (1992) writes:

Assessment procedures are inherently political, not only because whoever controls the assessment process shapes the curriculum pedagogy and ultimately the students’ life chances, but also because particular forms of assessment promote particular forms of social control within the organization, while suppressing others. (p. 18)

This dissertation brings together scholarship from educational testing, critical theory, validity theory, and writing assessment to better understand the ways in which external assessments and pressures to achieve cost efficiency and reliability undermine writing programs’ and writing instructors’ ability to educate students. To achieve this goal this investigation asks the following questions and approaches them through a métis-based institutional critique that is informed by critical theories of reproduction (Althusser and Bourdieu). Why are standardized writing exams flourishing in the face of research that shows decisions made from such assessments to be invalid or that achieve low levels of validity? How do specific cases of large scale writing assessments work to reproduce current socio-economic power relations? How can a métis-based institutional critique provide researchers and teachers with an avenue for influencing policy issues regarding assessment, placement, and admissions?

---

5 A detailed explanation of a métis-based institutional critique’s methodology is provided in chapter three.
The Problem

To better understand the “why” behind institutional decisions to employ large scale writing assessments that, at best, deliver low levels of validity, I propose a form of institutional critique that seeks to unite research on writing assessment with critical theories of social reproduction. The genesis for my ideas on a métis-based institutional (see methodology section) critique rise from a convergence of interests in both large scale writing assessments (Hillocks; Berliner and Nichols; Elliot; Gould; Bracey) and an overriding concern with ideological reproduction (Althusser; Bourdieu; Giroux; de Certeau; Foucault). I feel the two subjects are intimately connected. One need look no further than the ritual-like rite of passage known as the SAT and ACT tests that serve to filter graduating students into the proper social channels (trade skills, community college, four-year institutions, armed-forces) to see the powerful effect that testing has on reproducing power relations within America. Gee et. al. (1996) are blunt, “Basically, our claim is this: under the guise of the larger story we tell here there is a growing alignment between the business world in the new capitalism and various non-business spheres of interest, including schools and academic disciplines promoting school reform efforts” (p. 49). My concern as a scholar of writing assessment and a future writing program administrator is the monolithic appearance of the institutions that have aligned themselves with business interests rather than the interests of their students. As Porter et. al. (2000) point out, it is important to look for space in which to operate, to find gaps in the larger structure of an institution through which change appears possible. To this end,
they write that institutional critique “insists that institutions, as unchangeable as they may seem (and, indeed, often are), do contain spaces for reflection, resistance, revision, and productive action” (p. 613). The need for a positive attitude for the power of individuals to affect change is crucial. Otherwise, the entire agenda of change will be stymied by the seemingly impossible task of confronting such massive institutions.

Porter et. al.’s notion of institutional critique recognizes and addresses many of the basic problems with institutional critique. It is a known tendency in academia to critique for the sake of critique. Scholars point out the problem, publish, and move on to a new subject for critique. An understanding of space and the notion that institutions are “physically situated” is an important element of institutional critique. Porter et. al. explain by saying, “we use some of the ways that they (postmodern cultural geographers) deploy visual analysis to question and destabilize institutions, to provide an alternative route to interrogating how power circulates in particular institutions, and to complicate our construction of institutions” (p. 620). My spatial analysis of the contemporary university keys in on the institutional use of testing as a means of power. Many tests are given in most universities, but Porter et. al. emphasize the need to localize institutional critique. They write, “by conceiving of institutions as also operating locally, we better situate ourselves in visible contexts within which we conduct our lives and, again, have out lives conducted for us. We can begin to locate agency more so in the micro conception” (p. 621). To this end, my institutional critique of the CUNY writing exam at Kingsborough writing program locates itself within the local issues surrounding the program. The battle over remediation, ESL students, pressure from state and city policy-
makers to increase graduation rates, and many other factors play a role in the decision to install an external form of assessment in place of a portfolio designed for teaching and assessing. This localized targeting of institutional critique is in keeping with Porter et al.’s methods. They explain that “institutional critique examines particular institutional formations that are a local manifestation of more general social relations, nodal points, in the rhetorical relationships between general social (if not sociological) processes and local practices” (p. 621). The fusion of institutional critique and critical theories of reproduction provides a lens through which institutional powers are visible and their consequences, social reproduction, clear. Institutions “are rhetorically constructed human designs (whose power is reinforced by buildings, laws, traditions, and knowledge-making practices)”, according to Porter et al., are involved in the duplicitous behavior that Gee et al. describe. For this reason I have also chosen to appropriate the Classical Greek notion of cunning intelligence as a way of approaching my analysis of institutional power.

My rationale for reappropriating métis is that it fits well with the realpolitik of fast capitalism. Gee et al identify fast capitalism, which they also refer to as the new capitalism, as possessing four primary traits:

1. More stressful and more responsible work for those who have ‘good’ jobs, a proliferation of low-paying and temporary jobs, and many people with no jobs.

2. A widening gap between the poor and the rich.

3. A world where national borders—and the concerns of our fellow citizens in need—are considered to matter less and less.
4. A world in which there are often a small number of big winners and a large
number of losers in the race for jobs, wealth, and markets. (pp. 36-41)

In addition to these four traits, fast capitalism is also characterized by the necessity of
corporations to deal in duplicity. Apropos to this duplicity, a métis-based critique
approaches the world as a duplicitous and ever-shifting landscape where things are not
what they seem. Detienne and Vernant (1978) describe métis’ temporal nature:

   It (métis) operates on a shifting terrain, in uncertain and ambiguous
   situations. Two antagonistic forces confront each other. Over this fraught
   and unstable time of the agon, métis gives one a hold without which one
   would be at a loss. During the struggle, the man of métis—compared with
   his opponent—displays at the same time a greater grip of the present
   where nothing escapes him, more awareness of the future, several aspects
   of which he has already manipulated, and richer experience accumulated
   from the past. (p. 14)

The terrain of new capitalism as defined by Gee et. al. is riddled with duplicity. This
duplicitous nature is not only evident in corporate scandals like Enron and Tyco but also
in the very nature of the corporate model of fast capitalism. As Gee et. al. note, “and yet
the language that encapsulates the ends, goals, and vision of the organization and its
culture must be translated into a language of social intervention or manipulation
(depending on the view that you take of the matter) so as not to insult the
workers/partners and so as to motivate them” (p. XV). The new ideological hurdle
confronting modern corporations of the new capitalism, then, is effecting a change in the
way Americans view their relationship to work and their employer. The traditional view of loyalty to a business was directly connected to a reciprocal relationship where the employer would provide benefits and pension plans for the employee when time and health interfered with an employee’s ability to contribute. The new capitalism, based as it is on distributed systems and loyalty, can no longer reciprocate loyalty in this way, though it demands loyalty more than ever. Gee et. al. note that:

new capitalist businesses want such ‘deep’ learning, with it concomitant identity and value formations. These businesses are well aware too, of the conflict between traditional “American individualism” and the ‘teamwork’ of the new work order, or the conflict between the traditional ‘job’ and the ‘portfolio career’ of new capitalism. (p. 15)

This notion of “portfolio career” is a rather nice way of saying that once a particular “job” is completed, workers are often “down-sized” another nice way of saying “fired” by the company due to its need to stay “lean and mean” within a global competitive economy. For Gee et al, it is this competitive economic reality that has changed the face of capitalism and spawned a new business model. They write:

in fact, it is the responsibility of the new capitalism to push down control and responsibility to the lowest possible level, closest to the actual products, services, and customers of the business. This, however, requires workers now who can learn, adapt quickly, think for themselves, take responsibility, make decisions, and communicate what they need and know to leaders who coach, supply, and inspire them. (p. 19)
To create the kinds of workers needed in this business model, a certain retraining/education must happen. On one level, schools prepare students by teaching skills (reading, writing, computers). On another level, however, this training must be ideological—a retooling of the American work ethic—an overhaul of the socio-historical view of the relationship between the employee and employer. As Althusser (2001) and other critical theorists have noted, this reproduction of the means of production is an enculturation process that develops over the course of an individual’s life. The church, the school, and the family must all be working from the same script. For as Gee et. al. note, the new capitalism’s greatest challenge is

How to gain the fullest loyalty and trust of newly empowered workers so that they will throw themselves heart and soul into the work of the company in very risky times. In a hypercompetitive, fast-changing environment workers must be ‘eager to stay’, but also ‘ready to leave’ if the business is failing or even if it must innovate new projects that no longer require the core competencies of the current workers. This is hard situation, indeed, within which to motivate people. (p. 19)

In terms of reproducing the means of production, having control over schools holds for fast capitalism one of the answers to overcoming the daunting challenge of motivating workers to “throw themselves heart and soul into the work of a company” that they very well know will sever ties when the work is complete. Althusser stresses that the Repressive State Apparatus (police, military, justice/penal system) functions by exerting
force and violence, while the Ideological State Apparatuses (ISAs) (i.e. churches, schools, the family, the communication industry) function through ideology.

If schools⁶ are the primary ISA operating within the United States then the use of standardized writing assessments is an important site for investigation to understand how external assessments such as the ACT writing exam, created and administered by for profit corporations, work to advance the agenda of fast capitalism. In terms of confronting this external control, then, a form of institutional critique and change must take into account the duplicity and cunning of its adversary. We can use métis as a way to critique institutions and to provide scholars and administrators with a notion of how to advance their own agenda. Métis supplies a cunning intelligence that allows the awareness of shifting circumstances, and compliments another aspect of métis that is akin to kairos. Detienne and Vernant write:

the man of métis is always ready to pounce. He acts faster than lightning.

This is not to say that he gives way to a sudden impulse, as do most Homeric heroes. On the contrary his métis knows how to wait patiently for the calculated moment to arrive. The operation of métis is diametrically opposed to that of impulsiveness. Métis is swift, as prompt

---

⁶ Many would argue that the mass media apparatus is the dominant ideological apparatus of the 21st century. There is neither the time nor the inclination on the part of the author to take up this dispute. One could argue that the corporate owned media (Robert McChesney’s Corporate Media and the Threat to Democracy) and schools are becoming closely linked through technology and advertising in schools.
as the opportunity that it must seize on the wing, not allowing it to pass.

(p. 15)

A métis-based critique is fluid and able to respond quickly to social exigencies, whereas institutions are notable for their slow, bureaucratic pace. The ability of the scholar/activist to define and frame an argument is a matter of timing and timely action. That said, the ability of institutions to resist change is also closely related to this bureaucracy. Finding a space where change is possible, at a time when change is possible, is one of the most important aspects of institutional critique. The resistance to the federal No Child Left Behind act has brought standardized testing back into the national debate on educational reform. Therefore, while in terms of reproducing the means of production, the use of standardized forms of assessments in schools holds for fast capitalism one of the answers to overcoming the challenge of convincing future workers that being “eager to stay” but “ready to leave” is the natural order. The time is also ripe for scholars and activists to engage these institutions in the name of change.

Corporate interests realize that the shift in the way Americans view work and their relations to their employers must happen on a socio-cognitive level rather than a physical one. To that end, Althusser stresses that whereas the Repressive State Apparatus (police, military, justice/penal system) functions by exerting force and violence, the Ideological State Apparatuses (i.e. churches, schools, the family, the communication industry) function through ideology. Althusser writes,

if the ISAs ‘function’ massively and predominantly by ideology, what unifies their diversity is precisely this functioning, insofar as the ideology
by which they function is always in fact unified, despite its diversity and its contradictions, beneath the ruling ideology, which is the ideology of ‘the ruling class.’ Given the fact that the ‘ruling class’ in principle holds State power (openly or more often by means of alliances between classes or class fractions), and therefore has at its disposal the (Repressive) State Apparatus, we can accept the fact that this same ruling class is active in the Ideological State Apparatuses insofar as it is ultimately the ruling ideology which is realized in the Ideological Sate Apparatuses, precisely in its contradictions. (p. 98)

The unified class ideology that informs ISA’s is clearly a corporate ideology that looks at the bottom line, the interests of shareholders, and organizes its institutions around the business model. The manner in which the corporate ideology is functioning, ideologically, on a political and economic level with the goal of standardizing public education as an attempt to enact its socio-economic agenda should now be clear. As we have seen, the cooptation of higher education’s lifetime of learning strategy behooves fast-capitalism’s notion of the “portfolio” career in which workers are able to move from one job to another, one company to another, and take their lives into their own hands (by lives I mean pensions and healthcare plans), while also being “on call” for the next occasion where their skills are needed by an emerging business. If, at this historical moment, one views the standardization of public schools as a failed campaign, then they need only to look more closely at the inroads made by these corporate interests into education.
A métis-based critique approaches the standardization of schools through test-based educational reform as a corporate assault on public education, and holds an inconvertible belief that schools (K-12-4 year colleges) are the battleground for an ideological struggle that has nothing less than the future as its prize. Although efforts to privatize public schooling over the last twenty years have been, for the most part, unsuccessful, the use of standardized assessments as an external means of controlling curriculum has been very successful.

A tenet of a métis-based critique is that this reproduction of the means of production is at the heart of for profit institutions. Bill Readings, writing in the *University in Ruins*, argues that Althusser’s Ideological State Apparatuses (ISA’s) are no longer useful means by which to discuss the university because he believes the university no longer plays a role in the ideological reproduction of the means of production. The University for Readings is just another corporation and thus worries only about itself. I disagree with him on two counts. One, the university still plays an ideological role because the university has become a corporation with the agenda of profit at its core, and, two, corporations have concerns outside their own institutions (stability, supply and demand, access, exigency). Corporations (and this includes universities run on a corporate for profit model) rely on the reproduction of the means of production for their own success and survival. Simply put, universities are in the business of ideological and skills training for the new capitalism. For these reasons, Althusser’s ISA’s are very much part and parcel to my methods of institutional critique because ISA’s (the church, schools, the corporate media, and politico-judicial system) are the very institutions that teach
individuals how to act and how to believe in a manner that is conducive with the dominant ideology.

Methodology

A métis-based critique of two separate cases works to demonstrate how institutions of higher education reproduce socio-economic power relations by employing particular forms of assessments for particular purposes. In keeping with Porter et. al.’s notion of institutional critique, my investigation analyzes both a local, or micro-level, institutional case of assessment, the ACT writing exam at CUNY at Kingsborough, and, on a micro-level, the relatively new Directed Self-Placement (DSP). A métis-based institutional critique methodology is based on a theoretical analysis of assessment practices, social and individual consequences as a result of low levels of validity as a result of these assessment practices, the institutional decision making process that informs the decision to implement these assessment procedures, and, ultimately, the ways in which these decisions, procedures, and consequences result in the reproduction of socio-economic power relations.

Case One: CUNY at Kingsborough and the ACT Writing Exam

The analysis of this case of writing assessment as socio-economic reproduction begins by locating the writing program of CUNY at Kingsborough within the larger structure of the university. The intent is to understand how institutional forces, in this case writing assessment, do the work of social reproduction. To this end, this investigation focuses on four points of interests that consistently appear in the scholarship
of writing assessment and educational measurement. The first point of interest is the policy and the decision-making process that took place at CUNY at Kingsborough. The investigation analyzes the relationship between the teachers and scholars within the writing program (experts) and the city and state policy-makers (non-experts). This aspect of the investigation draws upon the scholarship of Barbara Gleason and Bruce Chadwick, teachers and scholars within the program, who provide a detailed account of their relationship with policy-makers. For balance, the investigation provides public testimony of policy-makers that explains the rationale behind their decisions.

The second area of interest for this investigation is the demographic breakdown of the students involved in the writing assessment. This analysis charts the gender, racial, and ethnic breakdown of CUNY at Kingsborough, and asks questions about the consequences of the decisions based on this assessment. What kinds of students test out of the ACT writing exam based on other standardized test scores? What kinds of students are forced to take this exam to be admitted to larger opportunities? What kinds of students fail? How often do they fail? What happens to students who are ultimately unable to pass the test? Finally, what do the answers to these questions reveal about the ability of external forms of assessment to reproduce socio-economic relations?

The final area of interest is to ascertain the effect that the writing examination has on students’ opinions, views, and understanding of writing. This aspect of the investigation draws primarily on specific scholarship about the writing exam at CUNY.  

Case Two: Directed Self-Placement
Part two of this investigation into the ability of institutions to reproduce socio-economic power relations through assessment practices focuses on Directed Self-Placement (DSP). This is an analysis for the purpose of understanding the theoretical foundations on which DSP is built. The analysis of DSP revolves around several key issues. The first issue is whether or not DSP avoids the pitfalls of standardized assessments, as it claims to do, or whether DSP simply shifts the burden on to the individual student. The second issue stems from the administrative decisions that inform the decision to adopt DSP. What are the reasons and rationale behind implementing DSP on an administrative level? The third issue is whether or not DSP achieves even low levels of validity based on the theoretical premises on which it is espoused? The final issue, and perhaps most important, is whether or not DSP’s theoretical foundation is built upon existing theories of literacy and learning or whether it simply recapitulates the notion of education and knowledge as static, telementational, and independent of socio-economic factors.

These social, economic, and theoretical issues are approached from the perspective of a métis-based institutional critique that asks certain questions of DSP and its proponents. How does DSP measure up against a unified notion of validity? Isn’t DSP simply a cunning ploy to use standardized assessments without openly relying on scores produced by those same standardized assessments? Are the theoretical premises on which DSP is founded, Dewey’s instrumentalism, in keeping with the theoretical works from which they are drawn or are they simply cherry picked, ignoring Dewey’s experiential continuum, and taken out of context from the larger theoretical body of work from which they come?
This métis-based critique begins with the flagship of scholarly work on DSP published in the CCC’s on September 1998. Daniel Royer’s and Roger Gilles’ “Directed Self-Placement: An Attitude of Orientation” served three functions. Foremost it introduced writing scholars to the theoretical foundations of DSP. Second, it provided two particular cases of students who had participated in this direct self-placement procedure and, according to the authors’ opinions, serve as examples of why DSP is successful. Finally, the article reported the overall “happiness” of all involved (students, teachers, administrators). This element of the investigation focuses on three key areas. A critique of the Deweyan theoretical foundation that is shaky at best, the authors’ blatant refusal to address the effects of educational assessment on students’ self-image and self-assessment, and, finally, the ambiguous and equivocal nature of the two student profiles of success as presented by the authors. In this vein the critique also investigates Royer’s and Gilles’ “Pragmatist Foundations of Directed Self-Placement.”

The goal of critiquing DSP is to show the institutional desire, often abetted by writing scholars and teachers, to find the path of least resistance (Williamson, 1994). This trend, often supported by the best of intentions, highlights the dangerous threat of ignoring validity theory and validity theory’s tenet that all practice must be informed by theory. If writing scholars venture into assessment practices they must do so with an understanding of educational assessment’s power to create both an individual’s social image (GPA, ACT/SAT), and an individual’s self-image (Hanson, 1993; Shohamy, 2001).

Significance of the Problem
Issues of assessment have always been critical to both the composition classroom as well as to the discipline as a whole. What role does writing assessment play in the struggle of citizens in a material society? How can education provide the citizens of a democracy with the ability to fully participate in the political and social life of the nation? The cultural capital of testing and assessment provides external control over the educational process from the earliest levels of education through graduate school and into the classroom curriculum itself. This study provides both a method of critique and a means of making institutional change by illuminating the power of assessment in writing and as a whole. This study approaches the powerful technology of writing assessment through both a critical theory of reproduction, but also through a perspective of cunning, métis, that is acutely aware of the social consequences of high-stakes decisions. Although plenty of fruitful scholarship has been done on the impact of assessment in the classroom, on curriculum and on its ability to socially define the individual, no single work has approached the power of writing assessment as an institutional means of social reproduction. This work is important for the discipline of Rhetoric and Composition because the power of institutions to reproduce socio-economic relations lies at the heart of our field’s effort to prepare students to be critical members of a society in which corporate, for profit institutions look at the bottom line as the ultimate goal of their enterprise. Furthermore, our role as administrators of English Departments and writing programs make it essential that we understand how things work, and, more importantly, why decisions are made.
Furthermore, the institutional critique will provide a lens for approaching policy battles that to this point have shunned the experts (writing scholars) in favor of external powers and non-experts (policy-makers and administrators). At the core of this investigation is the importance for writing scholars to be instrumental in designing and administering writing assessments that determine entrance, placement, evaluation, and matriculation rather than having non-expert outliers making such crucial decisions. For, as this project makes clear, the ability of external forms of assessment to control curricular decisions, pedagogical methods, and thus undermine scholarship and theory is crucial to the success of rhetoric and composition programs nationwide.
Chapter Two: Review of Literature

To contextualize this investigation into the ways in which institutions employ writing assessment practices that work to reinscribe socio-economic power relations, this review of literature addresses three areas of scholarship: writing assessment, its history and its contemporary state of affairs; the evolution of a unified theory of validity and its impact and role within writing assessment; and, critical theories of reproduction.

An understanding of writing assessment’s origins in the positivist tradition of science and educational measurement is a necessary facet for contextualizing writing assessment’s role in the reproduction of socio-economic power relations. It is common knowledge that standardized tests are biased along racial and gender lines, but the fact that the original scholastic aptitude test was designed by C.C. Brigham, an outspoken supporter of the eugenics movement (Elliot, p. 76), is a lesser known fact. This review offers a look into standardized testing’s origins within positivism, its outgrowth from a racial ideology of superiority, and writing assessment’s trajectory from these positivistic beginning to the current struggle to introduce assessment practices in league with a socio-cognitive model of literacy.

Educational testing was, for the most part, atheoretical prior to the last fifty years (Angoff, p. 19). In part, this lack of theorizing can be attributed to a priori notions held by testing advocates arising from a Eurocentric racial ideology of superiority. This review approaches the theorizing of testing, specifically the notion of a unified theory of
validity, to provide an understanding of the impact that validity theory might ultimately play in the future of educational assessment, but, more importantly, in institutional writing assessment practices and policies. The relatively new concern with the social and individual consequences that decisions based on test scores have for individuals resuscitates the meritocratic origins of testing and provides an avenue for affecting change in the nature of how assessment is understood. Rather than using assessment for its traditional purposes, for surveillance and punishment, taking into consideration the social consequences of decisions based on test scores provides a theoretical justification for designing and implementing assessment practices that are equitable and theorized as instruments that provide insight into improving teaching and learning.

The final area, critical theories of reproduction, provides a theoretical lens through which to view writing assessment practices on an institutional level to understand why institutions decide to undertake particular assessment practices. Although the reality that all test scores correlate perfectly with household income is no secret, the fact that tests are used to rank, order, and mark students as gifted or challenged is a neglected effect of an educational career rife with testing. The role of writing assessment within social reproduction illuminates the power of institutional decision-making and reveals the “cunning,” métis, of institutional power.

The History of Writing Assessment

The history of writing assessment is enmeshed with the history of intelligence, aptitude, and qualifying tests. To understand how standardized writing examinations are currently being employed as a means for reproducing socioeconomic relationships in
contemporary settings of higher education, it helps to understand the political and social atmosphere from which testing emerged. As Hanson (1993) and Foucault (1979) explain, testing, in one form or another, “trial by battle,” “trial by ordeal,” has been an element of human, social structures throughout history. The “science” of testing, what we refer to today as psychometrics or test theory, however, has a unique history with an overt ideological agenda at the core of its conception.

Gould (1984) might appear an unlikely place to begin a conversation about testing, but in truth it is the perfect foundation for a conversation about testing’s social and historical origins. The stated goal of this investigation is to refute “the argument that intelligence can be meaningfully abstracted as a single number capable of ranking all people on a linear scale of intrinsic and unalterable mental worth” (p. 20). For Gould, the myth of innate intelligence allows for an environment in which “others will be thereby demeaned, and their lower socioeconomic status validated as a scientific consequence of their innate ineptitude rather than society’s unfair choices” (p. 27). We will see this same argument reiterated throughout the scholarship of testing and standardized writing assessment.

Gould begins his investigation with the 19th century debate surrounding heredity versus environment (nature vs. nurture) to explain the positivist argument of biological determinism. Gould points out that “biological determinism-based arguments arise when the sociopolitical climate is looking to cut government spending on social programs” (p. 28), a point that has been remarked upon in contemporary conversations concerning the latest wave of test-basted reforms. His narrative ultimately leads to five scientists (Binet,
Goddard, Terman, Yerkes, Brigham) who played important roles in legitimizing the institutional and cultural acceptance of testing which allowed testing to be incorporated into the very fiber of America’s educational system. Gould illustrates the manner in which these men (some consciously, others taken out of context) created a science of testing that allowed advocates of innate intelligence to use this socially accepted “science” to restrict access to jobs, higher education, and equal rights, but, moreover, this “science” allowed those in power to argue that society was ordered in a natural hierarchy based on innate intelligence. Apropos of a conversation about educational testing, C.C. Brigham, as the secretary of the College Examination Board, would design the Scholastic Aptitude Test (SAT), a test that continues (with very little difference in its design and administration) to perform a powerful role as the gatekeeper of higher education.

The power that Brigham’s SAT would exert and continues to exert over the social organization of American society is one of the central subjects addressed by Lehman’s *The Big Test* (1999). Lehmann outlines the original plan of the Educational Testing Service (ETS), as envisioned by Harvard President James Bryant Conant, to install in America a meritocracy that would supplant the traditional dominance of the European white male through testing.

Lehmann presents the four competing ideas concerned with the future of education during the early 20th century to contextualize the atmosphere in which standardized testing entered the enterprise of America’s educational system. One, those in favor of a progressive education believed in “the individual blossoming of each student as a creative thinker and an active, skeptical citizen” (p. 22), while the second view
believed that “tough, uniform standards should be imposed on American schools” (p. 23). The third view believed IQ test “should be given to all American children, so that the high scorers could be plucked out and given the best schooling and the average and low scorers cosigned to a briefer, more limited education” (p. 24), while the fourth view called for expanding the number of students who attended college after high school, and “condemned the use of tests to select gifted students or to impose uniform standards on the schools, because that would be ‘excellent and beneficial for one type of boy and girl and young man and young woman, namely, those of superior bookish ability…those same standards, because of their uniformity and rigidity, are thwarting and damaging to all other kinds of degrees of capacity’” (p. 25). Lehmann’s book captures how the arguments of the 20th century surrounding the use of standardized testing continue unabated into the 21st century.

The competing ideas concerning education in the middle of the 20th century included one of the watershed controversies of writing assessment, direct versus indirect assessment practices. Elliot (2005) explains that the desire for reliability and efficiency caused the College Entrance Examination Board (CEEB) to drop the traditional entrance examination, a written essay, because the assessment of such essays was seen as unreliable. A failure to achieve “inter-reader reliability was an enduring problem” (p. 84) and the demand for numeric representations of writing ability was a necessity of the psychometric school of “objectivity.” The almost fetishistic demand for reliability and efficiency paved the way for indirect methods of assessing writing ability. The use of “objective” tests, comprised of multiple-choice, vocabulary, and grammar problems, to
measure literacy aptitude was founded on the intelligence testing of the late 19th and early 20th century. With indirect assessment practices replacing direct assessment practices the ability to achieve high levels of reliability was attained. Many, however, were not pleased with the objective test as an indicator of writing ability. John Stalnaker wrote that “in a field like English composition, the objective test is a feeble instrument. One can but urge that more effort and attention be given to the many problems of improving the preparation and evaluation of the discursive type of test which now seems so essential in a field like English” (Elliot, p. 119). In an effort to develop such a discursive test, the Educational Testing System (ETS), in the persons of Godshalk, Swineford, and Coffman, devised a direct form of writing assessment that we now know as Holistic Scoring.

The Contemporary State of Writing Assessment

A closer look at the contemporary state of writing assessment provides an illustration of the ways in which the tension between the traditional psychometric notions of reliability and efficiency and the more nuanced notion of validity informed by theories of writing has played out. The two most popular models for assessing writing are currently holistic scoring and portfolios. Holistic scoring represents a compromise between psychometric demands for numeric representations on the one hand and the belief held by writing theorists that real writing (direct) must be the focus of a writing assessment practice. In truth, holistic scoring, while changing the sample, it looks directly at a real piece of student writing, continues the psychometric tradition of addressing writing “objective” traits that can be separated and measured as independent pieces of a larger construct, i.e. writing ability. The revolutionary aspect of portfolio
assessment is that it approaches writing as a complex social act that is contextual and purpose driven. Portfolios, when used properly, allow writing to be viewed as a process. When misused, portfolios can simply be another way of collecting a writing sample.

Camp (1993) addresses the theoretical rationale for creating new approaches to writing assessment that are modeled on writing as a complex “meaning-making activity” (p. xx).

Camp’s “Changing the Model for the Direct Assessment of Writing” approaches writing assessment from a historian’s vantage point to better understand the theoretical forces guiding writing assessment practices. Camp illustrates the theoretical paradox that has puzzled writing assessment theorists. On the one hand, as Camp writes,

We now know, too, that writers switch among processes and strategies as they generate text, depending on their perception of goals for writing and their plans for addressing them, and that writers’ awareness of the strategies they use is important to their performance in writing. (p. 46)

While on the other hand, as Camp notes, “this complex view of writing is not easily reconciled with traditional approaches to assessment” (p. 46). The first success for writing assessment, on a historical scale, was the recognition that our assessment practices were not in harmony with our theories of language and writing. Although the theoretical underpinnings for writing assessment were in place, the practice of assessment was lagging behind (Gere, 1980; Faigley et al, 1985; Huot, 1996). The shift to holistically scored essays appeared to be the answer. There was now a method by which actual writing could be assessed, and, because of writing assessments heritage within
psychometrics, the compromise between reliability and efficiency was always looming over the assessment practice. As Camp notes,

we attempt to minimize or compensate for those factors most likely to create variation in writing performance from one occasion to another. We try to present prompts that are immediately accessible to all test takers—whatever their individual interests, knowledge, or cultural experience—and that can be dealt with in the limited time available for writing. (p. 50)

Camp lists a score of attempts by assessment designers to control and make equitable the assessment experience. As she writes, however, “ironically, many of the efforts we made to enhance the reliability of the writing sample—especially those aimed at streamlining and universalizing the writing experience—appear now to limit the value of the assessment” (p. 52). In the end, Camp comes to the realization that “very likely we are also seeing the signs of growing incompatibility between our views of writing and the constraints necessary to satisfy the requirements of traditional psychometrics—in particular, of reliability and validity narrowly defined” (p. 52). The fact that assessment strategies within writing assessment have attempted to appease the psychometric community is at the heart of this investigation into the role of writing assessment in reproducing the means of production. Writing scholars and scholars in fields such as cognitive psychology, education, language acquisition, and other researchers who understand the socio-cognitive nature of learning are continually shut out from the design and administration of our cultures most powerful wing of education—assessment/testing. Michael Williamson’s “The Worship of Efficiency: Untangling Theoretical and Practical
Considerations in Writing Assessment” addresses the historical role that efficiency has played in the educational assessment practices of our nation, and his explanation provides a concrete motive for psychometrics being accepted as the dominant paradigm of educational assessment, its cost efficiency coupled with its reliability.

Williamson’s work traces the changing role of education in America, from the role of education “to prepare the social elite for college admission, with a college education assuring passage through the gateway to participation in leadership roles in American society,” through the shifting role of Twentieth century education “to prepare workers for their appropriate place in the work force” (pp. 147-148). Williamson illustrates the ways in which the Meritocracy envisioned by Americans gave way to the need for efficiency and reliability:

the growing need of a system that has had to educate increasing numbers of students and two important technological innovations intervened to foster the development of systemic assessment oriented toward sorting and selection of students. At the same time, standardized interview procedures were subsumed by the psychometric theory they helped to develop, a body of theory that enabled the development of large scale assessment, due to the technological innovations of multiple choice and computer scoring. (p. 151)

Prior to theoretical views on education that posit learning (reading, writing, social skills) as social activity, education was viewed as telementational. The teacher, having the knowledge within his/her own mind, had merely to pass this information on to the
students. Unfortunately, this is still the primary educational method used. Williamson writes:

> the factory management model seems to have provided the initial impetus for the development of assessment of achievement and ability in education. Since the nature of the skills that students are expected to learn is straightforward and unproblematic to describe, assessment was not a complex undertaking. (p. 159)

Although our contemporary theories on education, especially writing, view learning as complex, to say the least, it is no wonder that writing assessment, while constrained by psychometric notions of reliability and validity, remain so problematic. Williamson’s view of the early Twentieth century is still apropos of the early Twenty-first century, “Since written tests were considered too subjective and, like standardized interviews, are labor intensive, they were seen as too costly and unfair” (p. 159).

The link between psychometrics and holistic scoring is further detailed in Williamson’s “An Introduction to Holistic Scoring: the Social, Historical, and Theoretical Context of Writing Assessment.” Williamson makes the connection early on when he writes that “the assessment theory used in direct assessment, holistic scoring in particular, is an alternate application of the same empirical methodologies of educational and psychological measurement used in developing machine-scored indirect writing assessment” (p. 1). Williamson reveals that in our rush to embrace direct assessment, a form of assessment that “is more valid than indirect assessment because it involves
performance of the behavior being assessed,” writing scholars “prevented a critical
examination of the limits of various approaches to direct assessment” (p. 3).

These limits are explored more closely by Edward White (1993) in “Holistic
Scoring: Past Triumphs, Future Challenges.” White argues that “the rapid and
widespread acceptance of such scoring, and the kind of testing that led to it, let us take a
relatively uncritical stance. We were promulgating a kind of religious belief in an
approach to writing as well as a testing device, and we were more interested in converting
nonbelievers than in questioning the faith” (p. 79). For White the battle between
psychometrics, in the form of the Educational Testing Service (ETS), and writing
scholars came to a pivotal moment when Holistic scoring entered the scene. He writes
that “the CCCC leaders talked of sociolinguistics and the importance of writing as a
central form of critical thinking on tests and in the schools; ETS staff talked of
administrative convenience and the practical needs of students” (p. 85), as with the
Twentieth century’s desire for efficiency, so too were the powers-that-be clinging to their
most prized possessions. Even within the framework of a holistically scored, direct
assessment of writing the “ETS had an internal conflict with the requirements of
community during holistic readings; with an eye to the bottom line, it tends to see too
much debate about scores (or anything else) as time taken away from production of
scores” (p. 88). What is telling about White’s reminiscence on Holistic scoring’s rise to
power is the role that psychometric concerns about reliability and efficiency, also the
concerns of corporate/business interests, play in repressing the most critical element of
holistic scoring, community. To simplify, holistic scoring was reconfigured by ETS and
other private, psychometric modeled testing companies to look very much like the computer scored assessments outlined by Williamson in “Worship of Efficiency” (1994). It is out of the constant pressure to assess that portfolios emerged as an alternative model that allowed assessment to mesh with classroom pedagogy and practice.

Peter Elbow and Pat Belanoff (1997) explain that portfolios were alive and in the classroom long before the portfolio explosion of the late 1980’s and early 1990’s (Brunham, 1981; Ford and Larkin, 1978). The “urgent and growing pressure for assessment, assessment, assessment; test everything and everyone again and again; give everything and everyone a score; don’t trust teachers” (pp. 22-23), recall Elbow and Belanoff, “was the matrix for a greater than usual hunger for an alternative way to assess student writing and learning” (p. 23). The ability of holistically scored essay exams to provide a numeric representation of an individual’s writing ability was a compromise between writing scholars dedicated to direct assessment and an educational culture which demanded a numeric representation. Portfolios, as a response to holistic scoring, argued that a direct assessment was not enough. As Elbow and Belanoff write, “For teachers who already knew how problematic such assessment (holistically scored) was, the pressure for more of it drove them to seek assessment that was more compatible with their classroom practices” (p. 23). As with indirect forms of assessment and direct forms of assessment based on a single piece of writing, the ultimate problem for writing scholars was to find a form of assessment that matched the theoretical models that informed pedagogical practices. Elbow and Belanoff argue that “this process started when portfolios helped testers face up to a problem they had been ignoring (probably
because the problem was so intractable till portfolios came along): any writing exam is inherently untrustworthy if it calls for only one piece of writing. That is, we cannot get a trustworthy picture of writing ability unless we look at various kinds of writing done on various occasions” (p. 25). In this way, although not a silver bullet for the inherent troubles of assessing the complex construct of writing ability, portfolios offered, and still do, a new approach to thinking, responding, teaching, and assessing student writing. For Elbow and Belanoff, “portfolios kick back.” They kick back by frustrating the typical psychometric notion of numerical representation. They write that “no complex performance can be accurately summed up in a single number because it almost always has stronger and weaker aspects or dimensions” (p. 28).7

On another level, portfolios also kick back at traditional notions of the teachers as one part instructor and one part evaluator. As other writing scholars cited in this work have mentioned, and as Elbow and Belanoff also point out, “portfolios help teachers stay longer and more productively in the supportive role, but then in turn, help them move more cleanly but less frequently into the critical role” (p. 29). It is the two pronged ability of portfolios to, at one level, resist psychometric reduction of writing ability to a numeric representation while still performing the realities of writing assessment (placement, grading), and, at another level, to practice an assessment that is in harmony with theoretical notions of writing ability and learning and conducive to better writing.  

7 Elbow and Belanoff fail to take into account Analytic Scoring (Diedrich, 1976) and Primary Trait Scoring (Lloyd-Jones, 1977). Two scoring methods that provide a numeric score while also judging complex performances.
instruction. Elbow and Belanoff sum up the major theoretical points that portfolios support and foster:

- Grades undermine improvement in writing because they restrict and pervert students’ naturally developing sense of audience awareness.
- Writing is its own heuristic; it doesn’t have to be graded to lead to learning.
- Portfolios lead to a decentralization of responsibility which empowers everyone involved.
- Teacher authority needs to be shared if writers are to have genuine authority.
- All evaluation exists within a context of literacy defined by a multitude of factors, not all of which are products of the classroom.
- Knowledge, whether of grades or of teaching strategies or of theoretical underpinnings, is a product of discussion among community members.
- Evaluation, judging, liking, and scoring are inextricably bound up together and need to be thoughtfully examined.

The ability of portfolios to operate within the theoretical and practical realities of the writing classroom, educational institution, and larger society also carry certain dangers that must be guarded against. On the one hand, portfolios allow for a natural blend of evaluation and instruction because they focus on writing as a process of revision, peer review, and teacher feedback (both written and oral). The nature of the portfolio process allows for the classroom conversation to remain on writing rather than on grades (Huot and Perry, in press). Although portfolios allow for a more natural blend of evaluation
and instruction they are also capable of being standardized and abused in the same ways that holistically scored writing samples have been in the past.

The unique issues that portfolios bring to writing assessment are the focus of Huot’s and Williamson’s (1994) “Rethinking Portfolios for Evaluating Writing: Issues of Assessment and Power.” Huot and Williamson, remark that the field of writing assessment has more often than not been misled by the testing community’s intractable stance on validity, reliability, and efficiency. With the widespread use of portfolios to assess writing in the classroom, these same issues have come into play. They write:

to control testing is to control education, to control what will be valued and taught within the schools. Crucial decisions concerning assessment are often made by regulatory agencies and political and educational policymakers based on practical and political concerns of cost, efficiency, and public opinion…Examining the various theoretical and political pressures which influence what measurements are chosen and how they are implemented allows us to conceive of assessment procedures as instruments of power and control, revealing so-called theoretical concerns as practical and political. (p. 44)

Huot and Williamson explain how departments and institutions need to guard against a “knee-jerk” reaction that compels them to standardize their portfolio assessments. Their

---

8 The Kentucky Educational Reform Act mandated portfolio use throughout the state in the 1990’s, but the standardization of portfolios resulted in nullifying the pedagogical principles on which portfolios are theorized and practiced when properly used.
explanation reveals that the psychometric tradition demands standardization in order to achieve reliability. This desire to achieve reliability, as we have seen, is enmeshed in the psychometric tradition because this tradition has grown out of a need and/or desire to order, rank, and organize its students and citizens.

The inherent problem in this move towards standardizing portfolios is that the standardization, much like the training of holistic raters, eradicates the positive effects of the portfolio assessment. Huot and Williamson explain that the portfolio, if viewed as a test, will lose its ability to foster those elements of writing and writing instruction that its proponents praise. Where portfolios are mandated by city, state, or federal policymakers, Huot and Williamson warn that

we have no assurance that portfolios can encourage a learning environment in which the teachers and students have no say in how they are used, compiled, and scored. In these instances, it appears that the use of portfolios in high stakes assessment scenarios are predicated on political rather than educational rationale. (p. 51)

Instead, Huot and Williamson argue that “the simple truth of educational assessment is that what we choose to evaluate in our students’ performances will determine what they attend to in their approach to learning” (p. 48). Therefore, if teachers and writing programs are not in control of the assessment, then the reality behind this “simple truth” is that standardized portfolios are just as susceptible to external control of curriculum as the traditional standardized test has proven to be. Thus, for Huot and Williamson, portfolios are faced with the same problems that have hampered other efforts to free
teachers and students from the “accountability agenda” within the United States. Huot and Williamson illustrate the power of testing within our culture with an array of testing’s influence from property values in a specific school district to testing’s as a form of surveillance that governs the actions of administrators, teachers, and students. As Huot and Williamson conclude, the real power being exerted on portfolios as a form of assessment, namely, the need to standardize the portfolio process in order to make it reliable, “results from the pressure to locate power in a central regulatory agency such as the state education department rather than in the schools and the school districts themselves” (p. 54). This pressure, I argue, is and at the same time explains a larger cultural pressure that seeks to reproduce the means of production through testing.

The field of writing assessment is a unique place to inquire into this means of social reproduction because as the tireless work of writing scholars over the past five decades has finally produced a socio-cognitive model of literacy and writing pedagogy, scholars in testing and in writing assessment have also produced methods of assessment that take the socio-cognitive model into account. Unfortunately, these methods of assessment (performance, portfolios, direct) resists the traditional attempts of psychometricians to standardized writing as a stable, static, and testable construct. In the face of almost insurmountable evidence, however, the non-experts, i.e. the testing industry, institutional administrations, and federal, state, and city policy-makers have still managed to design and administer writing assessments that have little or no claim to validity (by any current definition of the word). One of the goals of this investigation is to reveal the political, social, and economic forces that buoy up the culture of testing,
while scholarship and theories explaining complex human performances demand that new modes of assessment be not only created but administered.

A Unified Theory of Validity

As our short historical review has shown, the evolving history of validity within the scholarship of educational measurement is inextricably linked with the history of testing that we have already reviewed. Granted, the early stages of educational measurement and intelligence testing were largely atheoretical. One effect of this under-theorization was an agreement on the definition of validity. Validity was generally defined by the idea that if a test measured what it purported to measure then it was a valid instrument. As Angoff (1988) notes, “it was the use of validity in its predictive sense that dominated the scene” (p. 20), referring to test-theory’s conception of validity during these times. The confusion and lack of concern or consensus over validity is a telling sign. The a priori notions that informed much of the work that testing undertook allowed reliability to take center stage because the notion that “a test is valid for anything with which it correlates” (Guilford in Angoff, p. 20) allowed researchers to see what they wanted to see or to find what they expected to find. For instance, the Army Alpha and Beta tests administered in WWI found that blacks and eastern Europeans were mostly illiterate and severely “lacking” in mental aptitude. These tests, we now know, tested cultural knowledge that was alien to both blacks that were segregated from white culture while simultaneously being denied an education and to newly arrived immigrants that spoke little or no English. The tests were reliable, they correlated precisely with the
racial and ethnic stereotypes that the test designers believed, and they were therefore valid.

The unification of theoretical notions of validity is a watershed event for educational measurement and an even more promising event for those concerned with creating and administering writing assessment practices that are in harmony with socio-cognitive models of literacy. This agreement obviously took some time to reach. Cronbach (1971) is the first clear articulation of a unified theory, but most people cite Messick (1989) as the pivotal moment. He defined validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p. 13). The importance of a unified theory of validity for writing assessment is located in the simple fact that this definition of validity is a clean break from the positivist roots of psychometrics that defined validity in multiple ways and ignored the social and individual consequences that have always been an unavoidable effect of testing. Messick’s notion that “validity always refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores” (p. 14) calls for validity inquiry, testing of rival hypotheses, and the consideration of value assumptions that “shape how research questions are framed, what data are gathered, and how results are interpreted” (Shepard, 1993, p. 424). This unified theory of validity recognizes that tests are designed, administered, and used to make decisions by people with their own values and assumptions, and that the most important element of an assessment is not the scores
produced but the validity of the decisions made and inferences drawn based off of those test scores. For writing assessment, it is not, therefore, whether the assessment practices are reliable, but whether the decisions based on those practices have a reasonable degree of validity.

Camp (1993) and Huot (1996) are the first writing assessment scholars to investigate and theorize a unified theory of validity’s role in relation to a socio-cognitive model of literacy. Camp (1993) begins “Changing the model for the Direct Assessment of Writing” with an introduction explaining that, as a field, “we understand more thoroughly than we once did that the tasks of writing vary with purpose, audience, and context” (p. 102). This understanding of writing is important to Camp’s purpose because a “complex view of writing is not easily reconciled with traditional approaches to assessment” (p. 103) and her goal is to begin a conversation about “the development of new models for writing assessment” (p. 103). Camp’s most important contribution to the conversation about new definitions of validity and writing assessment is her understanding that “if assessments are to be created that truly reflect our new understanding of writing and writing processes, we will need, in addition to new methodology, nothing less than a redefinition of the very purposes and roles of assessment within our education institutions” (p. 115). The call to redefine the purpose and to “(re)articulate” the purpose and role of writing assessment would be answered by Huot (2002).

The role of validity theory in writing assessment is an important area to understand for an institutional critique of large scale writing assessments as reproductive
forces. Huot, like Camp (1993) is largely responsible for (1996; 2002) introducing validity theory to the composition audience as a justification and rationale for rethinking the field’s assessment practices. His work begins by tracing the historical relationship between writing assessment and test theory to make clear that writing assessment is not atheoretical, and that understanding the theory informing one’s practice is of the utmost importance to improving that pedagogy. By linking assessment to both classroom practice (teaching students to assess their own writing and the writing of their peers as part of the process of being a writer) and to the institutional demands placed on writing assessment, Huot frames writing assessment within what Phelps (1989) called the PTP (practice, theory, practice) Arc.

For Huot, the need to understand writing assessment as a theoretically informed practice is of paramount importance if we expect to create assessment practices that match our theories about the social nature and complex performances of writing and the teaching of writing. Huot reflects that “many writing teachers and scholars feel frustrated by, cut off from, or otherwise uninterested in the subject of writing assessment” (p. 81) because “writing assessment has been developed, constructed, and privatized by the measurement community as a technological apparatus whose inner workings are known only to those with specialized knowledge” (p. 81).

In Huot’s view, the shift in educational measurement’s definition of validity has provided an opportunity for writing teachers and scholars to take an active role in designing and administering assessment practices that have the cultural capital of test theory as support. He writes that “conventional writing assessment’s emphasis on
uniformity and test-type conditions are a product of testing theory that assumes that individual matters of context and rhetoric are factors to be overcome” and that “most of the procedures and improvements in writing assessment have had as their goals either the reliability of the scoring or the instrument itself” (p. 85). Huot’s call for a new theory of writing assessment seeks to take advantage of recent developments in testing theory like the unification of validity:

It is necessary for those of us who teach writing and work in writing assessment to examine some of the radical shifts in testing theory which have been emerging, because these shifts have been influenced by the same philosophical and theoretical movements in the construction of knowledge that have influenced writing pedagogy. (p. 90)

Huot unravels the cultural myth of objectivity in testing and shows that standardized tests are just as subjective as portfolio assessments: “the point is that no matter how we go about educational evaluation, it involves interpretation. Human symbol systems are involved, and thus there is no ‘objective’ measurement” (Johnston, 1989, 510 in Huot, 2002, 91).

Just as Huot dispels the myth of objectivity, he also dispels the notion that reliability is synonymous with fairness. He writes that “translating ‘reliability’ into ‘fairness’ is not only inaccurate, it is dangerous, because it equates the statistical consistency of the judgments being made with their value” (p. 88). The goal behind addressing the myth of objectivity and reliability as fairness is to address the epistemological bankruptcy of test theory promulgated within positivist views of reality.
Huot addresses the needs and concerns of writing assessment through a social constructivist’s lens of literacy.

In the place of traditional writing assessment practices that stress “generalizability, technical rigor and large scale measures that minimize context an aim for standardization of writing quality” (p. 104), Huot proposes new writing assessment practices that concentrate on “locally-controlled” procedures that are “site-based,” therefore responding to the “specific need that occurs at a specific site” (p. 105). These assessment practices are “context-sensitive” which allows the assessors to “honor the instructional goals and objectives as well as the cultural and social environment of the institution or agency and its students, teachers and other stakeholders” (p. 105). They are also “rhetorically-based” and “accessible” in that the “procedures” and “rationales” behind the assessment procedures are “available to those whose work is being evaluated” (p. 105).

Moss, working within the field of educational measurement (1992; 1994a; 1994b; 1996), informs Huot’s (re)articulation of writing assessment’s future course. Moss (1992) explores performance assessments and the challenges they place on conceptions of validity. Moss explains that the “growing prominence of performance assessment in educational measurement—of assessment resulting in extended discourse, work exhibits, portfolios, or other products of performances—the demands placed on validity researchers have changed substantially over the past few years” (p. 229). She argues, ultimately, that “given the well-documented power of assessment to influence what students learn and what teachers teach, it becomes crucial to discover or develop
approaches to assessment that reflect the full range of skills, knowledge, and interests that educators want to nurture” (p. 252). The convergence of a new definition of validity (Messick, 1989), Moss’ work on writing assessment from the perspective of educational measurement, and Huot and Camp’s work on writing assessment informed by validity theory is a powerful nexus.

When Moss (1994a) addresses the “concerns about the quality of information and consequences from these standardized writing assessments—typically first drafts completed in brief amounts of time—taken with new understandings about the cognitive and social aspects of learning” (p. 109) and argues for authentic writing assessments that allow students “to explore more of their own purpose, to rethink and revise their work over extended periods of time, drawing on existing resources and responses from readers, and to reflect on the process and quality of their writing” (p. 109), she is reflecting the wishes of many compositionists. Moss notes that “assessments need to be standardized to some degree” (p. 110) and that this standardization “results in the tension between competing validity criteria that simultaneously advocate for standardization and purposeful, collaborative activity” (p. 110). Moss traces the tension back to traditional psychometric concerns with reliability coming into contact with contemporary socio-cognitive models of literacy and learning, and she explains the fact that “choices made in designing assessment systems not only impact the nature of teaching and learning (both in intended and unintended ways) but also the nature of the discourse about the purposes and process of education” (p. 115). At the core of Moss’ scholarship is a unitary
concept of validity that demands that assessment practices concern themselves with individual and social consequences as well as complex views of human performances.

To design writing assessments founded on a unitary concept of validity, Moss (1994b) explores whether or not there can be validity without reliability, and, as in other works (1992; 1994a), situates her scholarship as an answer to the call of educators “for alternative approaches to assessment that support collaborative inquiry and foreground the development of purpose and meaning over skills and content in the intellectual work of students and teachers” (p. 6). Here Moss argues for a hermeneutic approach to assessment (that) would involve holistic, integrative interpretations of collected performances that seek to understand the whole in light of its parts, that privilege readers who are most knowledgeable about the context in which the assessment occurs, and that ground those interpretations not only in the textual and contextual evidence available, but also in a rational debate among the community of interpreters. (p. 7)

Moss uses the assessment procedures undertaken in making a faculty to hire at a traditional university as an example of a hermeneutic approach to assessment. Moss, like Huot, believes that the myopic concern with reliability is problematic and limits the ability of assessment to deal with complex performances. She writes that “if reliability is put on the table for discussion, if it become an option rather than a requirement, then the possibilities for designing assessment and accountability systems that reflect a full range of valued educational goals become greatly expanded” (p. 10). To create these types of
interpretive assessment practices, Moss (1996) argues “that the theory and practice of educational measurement would benefit by expanding the dialogue among measurement professionals to include voices from research traditions different from ours and from the communities we study and serve” (p. 20).

Moss’ (1996) attempt to “enlarge the dialogue in educational measurement” (p. 20) explores what she calls the “dialectic between naturalist and interpretive conceptions of social science” (p. 21) in an attempt to “support a wider range of sound assessment practices, including those less standardized forms of assessment that honor the purposes teachers and students bring to their work; to theorize more productively about the complex and subtle ways that assessments work within the local contexts in which they are developed and used” (p. 20). Like Huot (1996; 2002) and Camp (1993), Moss cites Messick (1989) and Cronbach (1971) as individuals who have attempted to integrate other “voices from research traditions” outside the educational measurement community into their conversation about assessing complex human performances like writing, and provides hope that a united front against the reductive psychometric tradition might be formed. Any hopes of replacing the positivist tradition of psychometric, standardized assessment practices that reduce complex human performances to numeric representations rests on uniting the social, political, and professional authority of educational measurement with the socio-cognitive approach to teaching and learning.

Until this united front is established, the powerful industry in control of the technology of assessment will continue to reproduce socio-economic relations through the educational apparatus. These forms of assessment will be designed using the basic tenets of
traditional test theory, ensuring that the racial, ethnic, and gender biases from which this tradition originated will be perpetuated ad infinitum.

Critical Theories of Reproduction

An important approach to institutional critique is through the lens of critical theory. Critical theory provides a view of assessment practices as an instrument of ideological control within institutional settings. The use of drug testing within professional sports, the use of intelligence testing within civil service, and the use of large scale standardized tests within institutions of higher education all work as instruments of control. These instruments shape the populations of a given institutions by allowing or denying entrance (gatekeepers). These instruments are also used to rank and organize those populations. Finally, these instruments are employed to surveill and punish these institutional populations. On a larger scale, that of an entire society such as the United States of America, the combined use of testing across institutions, works to reproduce socioeconomic relationships.

The notion of cultural and social reproduction finds its origins in Marx, and many neo-Marxists (Althusser, 2001; Foucault, 1979; Bourdieu 1990; De Certeau, 1984) have continued and expanded his notion of the reproduction of the means of production. This review of literature will address several of these scholars as well as several critical evaluations of the institutional use of testing (Hanson, 1993; Spellmeyer, 1996).

Chapter one provides a methodological explanation of this investigation’s approach to critiquing the institutional use of large-scale assessments: a métis-based institutional critique. In that explanation a review of Althusser’s notion of ideological
state apparatuses (ISA’s) is provided as well as a rationale based on Gee et al’s construction of fast capitalism. For now, it will suffice to know that Althusser expands Marx’s tenet that all capitalist systems have an inherent need to reproduce the conditions of production. This must occur on both a material level as well as on a social/ideological level. The adage that the “world needs ditch-diggers, too” is a euphemism for this socio-historical pattern. In short, just as a company that produces clothing must restock its fabric, thread, and other materials, so too must this company replace sewing machines, needles, and other necessary instruments of production. As Marx explains, the clothing manufacturer must also have a means to replace its seamstresses, salesmen, and managers. All of whom must, on a social and ideological level, be prepared to take their place within the matrix of social and power relations inherent in a class society that exchanges the labor power of the individual for a wage. The example of the clothing store is meant to simplify the larger superstructural relationships that arise from the base of a capitalist society. The role of ideological state apparatuses, as Althusser explains, is to prepare or “train” different groups or “classes” of people for their roles within the productive system of the society. Althusser lists the following as ISA’s: the educational, the religious, the family, the legal, the political system, trade unions, communications (press, radio, television), the cultural (literature, arts, sports, etc.).

Our concern is with the educational apparatus’s role in ideological training, a topic Foucault explores in *Discipline and Punish* (1979). For Foucault, the organization of schools are the product of a dominant, organizing power (the ruling class) being exerted over individuals. He writes:
The Normal is established as a principle of coercion in teaching with the introduction of a standardized education and the establishment of teachers’ training colleges…it is established in the standardization of industrial processes and products…For the marks that once indicated status, privilege and affiliation were increasingly replaced by a whole range of degrees of normality indicating membership of a homogenous social body but also playing part in classification, heirarchization, and the distribution of rank. (p. 184)

In Foucault’s explanation of the ideological and organizing function of schools, the examination (testing for our purposes) is the primary instrument through which power is exerted. As an institutional instrument, the examination is a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish. It establishes over individuals a visibility through which one differentiates them and judges them. That is why, in all the mechanisms of discipline, the examination is highly ritualized. In it are combined the ceremony of power and the form of the experiment, the deployment of force and the establishment of truth. At the heart of the procedures of discipline, it manifests the subjection of those who are perceived as objects and the objectification of those who are subjected. (pp. 184-185)
Foucault’s analysis of the examination, or institutionalized testing, provides a connection between the educational apparatus and the economic power base from which institutions arise, and whose ideological interests they perpetuate through the use of discipline. Just as Foucault’s historical analysis of the role of the examination highlights the connection between the economic base and power, so too does Bourdieu’s work on education and reproduction (1990) illuminate the role of the educational apparatus in reproducing the means of production. Bourdieu employs his notion of “symbolic violence” to understand how institutions that have legitimacy, the school, the church, the legal system, exert power by imposing meaning. He writes, “Every power to exert symbolic violence, i.e. every power which manages to impose meanings and to impose them as legitimate by concealing the power relations which are the basis of its force, adds its own specifically symbolic force to those power relations” (p. 4). This legitimization process explains the “ritualized” nature of testing in the institutions of education that Foucault mentions. The ritual invokes tradition, history, and therefore legitimacy, “it has always been this way!”

Bourdieu refers to “pedagogic action” as “symbolic violence insofar as it is the imposition of a cultural arbitrary by an arbitrary power” (p.5). Bourdieu outlines the relationship between Pedagogic Action and Pedagogic Authority to explain the relationship of teachers, who have pedagogic authority, to their students, to whom they transmit cultural and social material through pedagogic actions. This material has cultural capital, and through testing, and, ultimately, credentialing, students accumulate cultural capital. This process of transmission, education, is, however, skewed by power
relations. These power relations are reproduced through the educational process, and are legitimated by the cultural capital of both the educational system as well as the testing industry. Bourdieu on the role of examinations writes that

> it is clear why, in order to carry out in full this function of social conservation, the school system must present the ‘moment of truth’ of the examination as its own objective reality: the elimination, subject solely to the norms of educational equity, which it undertakes and conducts with formal irreproachability, conceals the performance of the function of the school system by masking, behind the opposition of the passed and the failed, the relations between candidates…the school system and the structure of class relations. (p. 159)

Both Foucault and Bourdieu stress the importance of legitimacy, power relations, and the examination as an invisible means through which to commit, in Bourdieu’s terms, “symbolic violence” on individuals in order to reinscribe socio-economic power relations. More contemporary works on the use of testing and large scale assessments for entrance, placement, and exit decisions build on these foundational works.

One such work (Hanson, 1993) is found in the field of cultural anthropology. Whereas Foucault and Bourdieu approach testing from their positions as critical theorists, Hanson’s *Testing, Testing: The social consequences of the examined life* (1993) aims “to uncover in the everyday operation of testing a series of well-concealed and mostly unintended consequences that exercise far deeper and more pervasive influence in social life than is commonly recognized” (p.2). Although Hanson quotes Foucault in the
epigraph of the novel, he appears to believe these consequences to be unintended whereas Foucault and Bourdieu would argue that these so called “unintended” consequences are indeed the disciplinarian powers of the dominant ideology working to reproduce the social order. This aside, Hanson argues that tests “are mechanisms for defining and producing the concept of the person in contemporary society and that they maintain the person under surveillance and domination” (p.3). Hanson argues that “test have invented us all” (p.3) to explain the role that testing plays in contemporary society in building up or tearing down a person’s sense of self worth. Tests, Hanson reminds us, also play a role in a person’s social status.

On an institutional level, Hanson argues that “all tests contain the condition of intent: they are planned, arranged, given, and conducted by someone with some purpose in mind” (p.16). In Hanson’s view, the intent is to gather information “as a mechanism for fabricating and dominating the person” (p.20). Hanson looks at authenticity tests, such as lie detection, and qualifying tests, such as a civil service examination, as examples of the social use of tests. Most important for an investigation into the institutional use of testing is Hanson’s understanding that “test givers are nearly always organizations, while test takers are individuals” (p.304). This seemingly obvious fact is essential to understanding the educational apparatuses ability to produce what Hanson calls “the priority of potential over performance because decisions are made and actions are taken on the basis of test results rather than on more direct knowledge of the target
information that they supposedly signify” (p.298). The test predicts a person’s potential rather than allowing their performance to be the basis for future possibilities.

The idea of potential over performance that Hanson articulates explains the process of symbolic violence espoused by Bourdieu. The uses of culturally biased tests that correlate precisely with household family incomes (see table 4.1, p. 96) perform the transmission of cultural capital along socio-economic lines through test scores. This transmission has the legitimacy of tradition, because both schools and tests have their own cultural capital, and because tests scores have “the security factor involved. Resting decisions on explicit and readily comparable tests scores makes it possible to claim that the best selection was made on the basis of the information available, should decisions work out badly or there be troublesome protests or lawsuits” (p.300). Tests, in the form of legitimized and socially accepted gatekeepers, decide who will pass through the doors of educational opportunity.

This review concludes with Spellmeyer’s (1996) response to the use of testing as surveillance. Spellmeyer argues that we should understand those legal issues” surrounding the laws that enact state and federally mandated educational testing, but “unless we can play an active role in the fashioning of policy, there is little chance that we will retain control over what we do as teachers or as scholars. (p. 179)

---

9 See Chapter Three, “A Case of Social Reproduction” for an example of potential over performance in action. The NCAA’s Proposition 48 and Proposition 16.
Spellmeyer’s main point of contention\textsuperscript{10} is that a perceived crisis in the system of public education, a crisis that is not apparent in reality, has been manufactured to ensure the continuation of traditional standardized tests. Spellmeyer’s frustration and argument that teachers must be involved on a policy-making level, reflects the power of testing to control both the curriculum as well as the social perception of our educational system. He writes that

standardized testing is one of the most dangerous, invasive, and reactionary developments in the last hundred years of educational policy. We should not forget that for much of these hundred years, without the benefit of state and national testing programs, our society somehow managed to produce the world’s most broadly educated citizenry (pp.177-178).

Spellmeyer provides statistical evidence that America has continuously expanded its number of educated citizens since the 1930’s and argues that the idea of a crisis is often a misnomer for the privileged class’s unhappiness at sharing the fruits of literacy and educational opportunities with the lower classes. His analysis makes clear the reproductive forces of test-based educational reform when he pronounces that

testing will not raise up the lowly but will help to naturalize economic disparities that grow wider every year. With test results in hand, authorities are in a position to retract even the vestigial appearance of

\textsuperscript{10} Martin Nystrand argues the same point. He is cited in Chapter One of this investigation on Page 1.
equality that a high school diploma still confers on the marginal. Armed now with solid evidence, the best and the brightest can rest assured that the poor are stupid and undeserving. (pp. 178-179)

Spellmeyer’s understanding that testing is mandated through what Althusser calls the politico-legal apparatus (through federally and state mandated legal acts) in order to control curricular content and pedagogical action within the educational apparatus strikes at the unified ideology (the dominant ideology) that informs each of the ideological state apparatuses. These apparatuses, working in harmony and reading from the same script, work to reinscribe socio-economic power relations.

At the end of this review of literature it becomes possible to understand how the powerful engine of educational measurement is capable of exerting such an enormous influence on both individuals who fall under the gaze of educational assessment as well as the larger social organization of society itself. The history of test theory is rife with examples of tests being used to control educational opportunities, employment opportunities, and civil opportunities. The failure of test designers to create, administer, and employ assessment practices in congress with a socially responsible conception of validity is directly tied to this corrupted history. Even with educational measurement’s new conception of validity, large scale standardized tests continue to be designed according to the positivist tradition of test theory (ACT, SAT, NCLB, ACT Writing Exam). The test scores produced by these tests continue as well to be the basis for
decision making that informs entrance and placement decisions across the country.\textsuperscript{11} Finally, the lens of critical theory and the insight of cultural anthropologist like Gould\textsuperscript{12} and Hanson reiterate the power exerted by tests in general and educational assessment in particular and make clearer our desperate need to replace the traditional notion of testing with more nuanced and equitable assessment practices.

\textsuperscript{11} As we will see in Chapter Three, even when the ACT or SAT scores are not directly used for placement, they are still capable of exerting their influence over placement decisions.

\textsuperscript{12} Officially Steven J. Gould is an evolutionary paleontologist but many of his published works are centered on cultural anthropology.
Chapter Three:
Directed Self-Placement and the Experiences of Educational Assessment

As a technology, writing assessment plays an integral role in a person’s educational and economic opportunities, and, this powerful role makes writing assessment a place of contention. This view of writing assessment as a technology allows for a conversation about the role that writing assessment plays in reproducing the means of production. Brian Huot, writing in (Re)Articulating Writing Assessment for Teaching and Learning (2002), reminds us that “for the most part, writing assessment has been developed, constructed, and privatized by the measurement community as a technological apparatus whose inner workings are known only to those with specialized knowledge” (p. 81). This specialized knowledge has had the effect of limiting the conversation concerning issues of writing assessment by limiting the number of people who feel qualified to enter into such a conversation, and because of this, as a point of contention, writing programs and writing instructors have historically been underrepresented in the battle over assessment practices.

One reason that issues of writing assessment are so important is located in the role that reading and writing play in educational opportunities. Huot argues that “instead of envisioning assessment as a way to enforce certain culturally positioned standards and refuse entrance to certain people and groups of people, we need to use our assessments to aid the learning environment for both teachers and students” (p. 8). However, writing
assessment practices that occur outside of the classroom often take on a far more sinister role. For the purposes of entrance, exit, and placement, writing assessment, specifically, and educational assessment, more generally speaking, are more often than not employed as gatekeepers. It is this use of assessment practices as a form of technology designed to solve a problem that Carl Madaus (1993) describes when he writes that testing, embedded in our system of education with its arcane psychometric underpinnings, clearly fits this definition of technology. Testing also fits some very simple definitions of technology—the simplest being something put together for a purpose to satisfy a pressing or immediate need, or to solve a problem. (pp. 12-13)

If we posit the “problem” facing institutions as a need to reproduce the social conditions of reproduction necessary for a global economy, then the role of writing assessment in higher education becomes clearer, and the battle over writing assessment practices become all the more important.

The idea that the mission of the university has changed with the emergence of a global economy is the focus of Bill Readings’ The University in Ruins (1996). Readings claims that “since the nation-state is no longer the primary instance of the reproduction of global capitals, “culture”—as the symbolic and political counterpart to the project of integration pursued by the nation-state—has lost its purchase” (p. 12). The decline of the nation-state and the abandonment of “culture” have been replaced by “an increasingly transnational global economy” (p. 12) in which universities deal in what Readings describes as the discourse of “excellence”. This shift in the university’s core mission is
accompanied by other important changes, one notable change being that “the central figure of the University is no longer the professor who is both scholar and teacher but the provost…the challenge of the contemporary University is a challenge addressed to him as administrator” (p. 8). The paradigmatic shift that Readings identifies will receive further treatment as this chapter explores the manner in which writing assessments are employed as administrative tools for the purposes of organizing and placing students.

The use of writing assessment practices for the purpose of organizing, evaluating, and placing students is, of course, not new. However, within the field of composition studies there has been an emerging agenda for assessment that recognizes it as an important element for social action allows us the ability to guard against over-privileging the values, gestures and customs of certain groups and provides assessment with the potential to become an agent for progressive social change that includes and highlights the improvement of educational environments and opportunities for all students. (Huot, 2002, p.9) This agenda has continually been undermined by institutional demands for accountability through traditional psychometric testing procedures that fail to represent more complex notions of writing. This chapter investigates a relatively new first-year writing placement procedure known as directed self-placement in the context of writing assessment as a form of technology used for the purpose of reproducing the means of production. This
chapter focuses specifically on two pieces of scholarship\textsuperscript{13} offered to the writing assessment community by the founders of directed self-placement (DSP), Dan Royer and Roger Gilles. DSP has been received with much enthusiasm and has been implemented on at least twenty five college and university campuses across the country. A \textit{métis}-based critique is employed to understand how DSP has been theoretically fashioned and institutionally molded to “please everyone”, as the authors claim, while, at the same time, this placement technique has largely avoided the validity research necessary to argue for its implementation and continued use.

By approaching Royer’s and Gilles’ presentation of DSP to the composition community through the lens of institutional cunning this investigation attempts to understand how the cultural capital of test theory as developed within the discipline of educational measurement is able to perform its central role of socioeconomic reproduction while appearing to be absent from the procedural process of DSP. This ability to act and influence while remaining unseen is pernicious and crucial, and the manner in which DSP is presented, as good for students and free from the pitfalls of institutional assessment (Elbow, 2003, p. 15), reveals the cunning and compromise at the heart of DSP.

The compromise can be found in the very duplicity in which DSP deals. On the surface it argues for its ability to liberate students from the traditional forms of

institutional assessment, while in reality DSP relies on the effects of a lifetime of experiences with educational assessment and places the onus on the individual students. In the same way, while espousing the progressive educational ideals of John Dewey, DSP, in reality, relies on the past work of the most traditional, positivistic notions of intelligence, ranking, and categorizing that we reviewed in Chapter 2 to do the work of placing students in the “appropriate” writing course. To understand how DSP deals in duplicity and represents a compromise between the goals of writing programs seeking to develop equitable assessment practices and institutions that demand efficiency and reliability, this chapter investigates three key foci of interest: the exigency of DSP; the presentation of DSP as a feeling or intuition of “rightness” that allows for the casual dismissal of educational assessment’s influence on individual’s self-perception; and, the theoretical foundation of DSP as it has been presented to the composition community.

A basic outline of DSP as a described by Royer and Gilles at Grand Valley State University begins at orientation where students are given a tri-fold brochure. On the front cover is the question, “English 098 or 150: Which Course is Right for you?” The writing program administrator informs two hundred or so students that most placement methods are not valid (Royer and Gilles, 1998, p. 55), and that teachers and administrators normally place students but that at their university the decision is made by the individual student. The brochure also contains a list of “characteristics that we faculty look for in solid writing students” (p. 56). The brochure also contains a list of “general characteristics that may indicate that English 098 is best for you” (p. 57).
Students are asked to make a decision, and it is important to note that advisors are available to aid students who need help in making this decision.

It is also important to note that DSP, while generally defined by certain core attributes, changes according to the local context in which it is being employed. This critique of DSP is directed at the particular presentation provided by Royer and Gilles in their CCC’s article (1998) and their book chapter “Pragmatist Foundations of DSP” published in 2003. To introduce the theoretical and practical aspects of this placement procedure, Royer and Gilles presented their rationale to the writing assessment community, and, as time has shown, many people found something in DSP that they liked. The idea behind DSP is hard to argue against: that a student should play an integral role in their own educational decisions. In fact, one of Royer’s and Gilles chief claims is that “directed self-placement pleases everyone involved” (1998, p. 65) (students, teachers, admission officers and the administration). The critics (Schendel and O’Neill, 1999; Huot and Neal, 2003), however, have restrained their applause and approval until validity inquiry has been undertaken to validate this placement method. Unfortunately, with the publication of Directed Self-Placement: Principles and Practices (2003), seven years later, very little validation has been produced, and several areas of concern have been completely ignored.

Part I: The Exigency for Directed-Self Placement at Grand Valley State

Royer and Gilles introduced DSP at Grand Valley State University (GVSU) in 1996 because of “wide-spread frustration over our traditional placement method” (1998, p. 59). They describe the frustration and general unhappiness of all parties involved
(teachers, administrators, admissions people) with the traditional placement method of a
holistically scored timed writing sample in conjunction with students’ ACT-English
score. The exigency for employing DSP is the first of several problematic rationales
presented by Royer and Gilles. Our review of Williamson’s “The Worship of Efficiency”
provides insight into the “why” behind many institutional decisions. In this case, why
employ DSP? The need to implement a new procedure is predicated on the fact that
something is wrong with the current procedure.

The Writing Program Administration

In terms of the writing program administration (WPA), Royer and Gilles inform
us that

we first considered trying to improve traditional placement-test
procedures… but we realized that no matter how site-specific and
contextualized we made our reading of placement essays, we might side-
step some reliability concerns and finesse our notion of validity, but we
would inevitably wind up making decisions based on the inadequate data
of a single writing sample. (p. 59)

There are several troubling elements to the authors’ statements. The most obvious is the
unfounded assumptions about the inevitability of their decisions. If the traditional
placement methods of GVSU had been researched and validated then a mountain of data
should have been available about the validity of their traditional placement methods.
Data, we should add that are not presented in either of the major documents Royer and
Gilles produced about DSP. Further, William L. Smith’s University of Pittsburgh and
Richard Haswell’s Washington State studies on contextualized placement procedures are mentioned and dismissed in a single sentence. The reality is that Smith produced two lengthy book chapters documenting his placement procedures’ validity and his placement procedures have been replicated and proven, through validity inquiry, to achieve both a high degree of reliability as well as valid and equitable placement decisions (Smith 1993; Haswell, 1994; Lowe and Huot, 1997; Hester et al, 2000). An early example of the cunning and duplicitous nature of Royer’s and Gilles’ scholarship is found in their description of the orientation talk given at Grand Valley State. Although their talk calls into question the traditional methods of placement (such as Smith’s University of Pittsburgh model), their method, as understood through their own descriptions and the brochure that they provide incoming students, obviously relies on the same traditional ideas they say they are working to overcome. Their opening day talk says:

Before I get to the specifics, let me explain why it is we want you to make this decision and why we aren’t going to make it for you. At many schools, in fact at this school until very recently, people like me ‘place’ you into a writing course by looking at your ACT or SAT score, your high school GPA, and perhaps by having you step into another room and return to us two hour late with a ‘sample’ of your writing. But it turns out that this is not a very valid or reliable way to find out which first-year writing course is best for you. Writing ability, at least as we conceive of it, is far too complex to measure so quickly and easily. (p. 55)
The duplicity, as noted, of saying one thing while doing another is precisely the territory for which a métis-based critique was designed. While Royer and Gilles say that ACT/SAT scores are not valid, their brochure provides students with a verbal score cut off of 20 as a marker that they are prepared or not prepared for English 150. Like wise, the time writing sample that “is not a very valid or reliable” way to place students is indeed part of the process. Page four of the brochure, under the section “What to Expect on the First Day,” informs students that

In both WRT 098 and WRT 150, you will be asked to write a brief essay on the first day of class. Your teacher will read the essay as an indication of your writing abilities and let you know whether he or she feels you are beginning at the appropriate level. On the advice of your teacher, you will have the opportunity to switch WRT 098 to WRT 150, or vice versa, during the first week of classes. The final decision is yours, but your teacher will be able to help assess your writing in relation to course goals and the abilities of other entering freshmen.

This use of a timed sample of writing that is being read by an expert reader is modeled on William Smith’s University of Pittsburgh placement study. Royer and Gilles discuss Smith’s University of Pittsburgh study briefly as an example of spending too much time and resources on placement upfront rather than spending those same resources on teaching, grading, and program evaluation. It begs the question, whose decision is this, the student’s or the teacher’s? It is hard to believe that a teacher’s “advice” would be ignored by a student in this situation. Again, the cunning and compromise of DSP comes
to the surface as with one hand they dismiss the Smith model as an expensive, misguided use of resources and with the other they applaud the efficiency of a procedure that implements Smith’s expert reader model.

Add to these facts the statement by Royer and Gilles that we toyed with the idea of entrance portfolios—which would move us beyond the single piece of writing—but the Admissions directors balked. ‘This isn’t Stanford,’ they told us. “If we make students put together an entrance portfolio and the next school doesn’t, the students will simply pick the next school.’ Besides, we knew that asking for entrance portfolios would place quite a burden on our already overburdened summer faculty,

and it appears that the motivation for implementing DSP has something to do with efficiency rather than promoting the power of students to be involved in their own educational decision-making processes. Now it would be naïve to pretend that writing programs operate in a vacuum and are not subject to other interests within the university, but this particular representation of DSP was presented in terms of writing pedagogy, writing assessment principles and practices, and as a socially responsible, student-centered method of placement. If the real answer behind “Why implement DSP at your institution” is that it doesn’t overburden the summer faculty, it doesn’t upset the admissions people’s orientation schedule, etc., then let’s say so. The fact that this placement procedure claims to be founded on Dewey’s philosophy of instrumentalism

14 Why DSP? is a section title within the 1998 CCC’s article.
and the inherent failure of more traditional placement methods appears to be an act of rhetorical cunning for the benefit of institutional efficiency. Even more troubling is Royer and Gilles use of academic scholarship that was created with the goal to create more equitable forms of assessment (Huot; Moss; Yancey) to help ensure for-profit institutions of higher education their market share. The notion that, “this isn’t Stanford” is a clear statement of product value, brand identity, and the need to attract as many “seats”\(^\text{15}\) as possible for the sake of generating revenue.

One of the cornerstones of validity theory is the notion of gathering data prior to making a change. How effective were the traditional placement procedures? To what degree were the placement decisions made using the traditional methods at GVSU valid? What elements of the traditional placement methods used at Grand Valley State needed to be changed? Because no data are provided we are left with the anecdotal evidence provided by the authors. Royer and Gilles inform us that there was “wide spread frustration,” that “we’d never liked using ACT-English scores”, that “the admissions people who ran orientation didn’t’ like our method much, either; they had to schedule an hour for writing, then wait for the results before they could help the students register,” that “our English 098 students weren’t very fond of the system either,” and finally that “the teachers themselves were frustrated” (p. 59). If done systemically and with rigor, each of these areas of “frustration” could have been the focus of research. Here is the research Royer and Gilles provide their audience, and it, too, is a message of duplicity.

\(^{15}\) In the opening of the 1998 CCC’s article, the students are referred to as seats. The correlative can been found in referring to students as per semester hours.
They preface this scarcity of data with more talk about feelings: “in retrospect, we believe that our discomfort with traditional placement methods arose from a uneasy feeling of impropriety” (p. 63). More troubling is Royer’s and Gilles’ use of grade point average (GPA) and ACT-English scores as evidence that DSP is at least as reliable as the traditional methods used at their institution even though they admit several key facts about GPA and ACT-English scores: 1. that these two statistics are “problematic measures of writing ability per se,” (p. 63) and 2. “though as we have begun to discover, there may be good reasons to dismiss these general indicators of academic ability as unable to predict success in writing courses” (p. 64). Nevertheless, GPA and ACT-English scores are the only comparative data provided or gathered by Royer and Gilles. They write concerning ENG 150 students of 1995 as compared to ENG 150 1996:

> We compared two of the most easily measured characteristics of the two populations and found that the groups shared very similar high-school GPAs and ACT-English scores. This suggests that the students took their high-school GPAs and ACT-English scores into account: we didn’t need to do it for them. So as a “replacement” of the old system, directed self-placement worked… (p. 64)

Although Royer and Gilles dismiss the usefulness of GPA and ACT-English scores for placement purpose in one breath, they then use the comparison of these statistics as evidence that DSP is a valid replacement for the “old” system. One might then ask, why not just place students according to GPA and ACT-English score? The answer, these scores have proven to have little predictive quality for college performance and no
correlation to writing acumen. The use of this data is rife with duplicity, and the cunning is found in the fact that “students took their high-school GPA and ACT-English score into account: we didn’t need to do it for them” (p. 5). This statement nullifies the statement that we will see from Royer and Gilles when they confront whether or not test scores define students’ self-image as writers.

The Students

This cunning approach of cloaking DSP within the scholarship of writing assessment and student-centered theories of education continues as Royer and Gilles present the “frustrations” of the students. To this end they present a single survey of English 098 students from the prior semester (Fall 1995). The actual questions or survey is not presented but they present their findings in the following manner: “we surveyed our students in the Fall of 1995 and found that only 38% of the ENG 098 students felt they were properly placed in the course. There were quite a few negative comments about both the placement procedure and the course itself” (p. 59). The correlative to this, though no data is provided, is that using DSP would bring this number up considerably because the students make the placement decision themselves. The only data that may provide some insight into this seemingly arbitrary data set is that “our ENG 098 ‘placement rate’ has dropped from 33% to 22%, but for the first time we feel that the right students are taking our developmental writing class” (p. 61). The idea of “feeling” and “rightness” is an ongoing theme within Royer and Gilles presentations of DSP, and on the level of validity inquiry or any evidence-based argument for important educational
policy change is a very problematic notion. How do we reflect, research, or validate feelings of rightness? The cunning, once again, is that we can’t.

Part II: It Feels Right

To couch institutional decisions and procedures in qualitative terms such as their degree of “rightness” (p. 62) is not necessarily a new gimmick within higher education. The current rhetorical trend embraced by a majority colleges and universities, the striving for “excellence” in higher education, has been theorized and interpreted by Readings (1996) and is similar to Royer and Gilles use of “rightness”. Readings notes that “as an integrating principle, excellence has the singular advantage of being entirely meaningless, or to put it more precisely, non-referential” (p. 22). Of course, my argument is that the same can be said of the feeling of “rightness.” This is not simply a positivist’s wish for quantifiable categories, but rather the demand for an arguable basis for making important assessment decisions and educational decisions about students. For example, how can one argue with the degree of “excellence” or someone or some group’s feeling of “rightness.” To make institutional and programmatic decisions there needs to be accessible paths along which validity can be traced. In the case of DSP as with more

16 Directed Self-Placement Feels Right is another section heading from the 1998 CCC’s article.
traditional placement procedures there are multiple paths on which validity inquiry can be undertaken\textsuperscript{17}—paths that Royer and Gilles neglect. For instance:

1. The success of students within the particular course can be researched.
2. The professional and highly contextualized professional opinion of writing instructors on whether students were properly placed within their particular writing course(s).
3. The individual opinions of the students on whether they were placed or placed their selves correctly.

To argue for particular procedures or decisions based on these qualitative categories of rightness or excellence purposefully creates obstacles to validity inquiry. How can we measure or quantify “rightness”? How does one know if they are fulfilling a university’s mission of “excellence”? The cunning can be found in the answer to these questions.

You cannot inquire into these qualitative measures.

These are rhetorical choices, and, like all rhetorical choices, they are important choices. To argue “for” DSP because it “feels right” is an abdication of a responsibility to inquire into the validity of the decision to implement DSP. These rhetorical choices mask the institutional duplicity in order to explain “why” a certain decision has been made to a group of people who will not accept “efficiency” as the sole and overarching motive behind their policies and procedures. Readings’ remarks on “measures of excellence” can be transposed over Royer’s and Gilles’ notion of “rightness”. He writes

\textsuperscript{17} These paths were researched and written about by William L. Smith long before Royer and Gilles created, implemented, and published on directed self-placement.
that “a number of things are obvious about this exercise, most immediately the arbitrary quality of the weighting of factors and the dubiousness of such quantitative indicators of quality” (p. 25). The use of such qualitative and immeasurable categories by Royer and Gilles is but another example of many of the duplicitous rhetorical moves involved in DSP’s presentation to the composition community.

However, the rhetorical framing of DSP in qualitative measures has more insidious roots and is merely one symptom in a much larger and systemic transformation of the contemporary university. Readings identifies the need for “qualitative” and “non-referential” measures with a paradigmatic shift in the role that American universities now play in a global economy: “this seems directly symptomatic of the reconception of the university as a corporation, one of whose functions (products?) is the granting of degrees with a cultural cachet, but whose overall nature is corporate rather than cultural” (p. 11). The corporatization of the university brings with it a shift in the role of the student to the university and the university’s relationship to the student. Readings points to the fact that “students’ frequent perception of themselves and/or their parents as consumers is not merely wrongheaded, since the contemporary university is busily transforming itself from an ideological arm of the state into a bureaucratically organized and relatively autonomous consumer-oriented corporation” (p. 11).

Royer and Gilles (1998) provide their own “measure of excellence” by submitting as evidence of DSP’s “rightness” the following survey of 230 students who voluntarily placed themselves in the developmental, non-credit earning writing course (ENG098):

1. 24% said “Generally I don’t read when I don’t have to.”
2. 23% said “I don’t think of myself as a strong writer.”

3. 15% said “My ACT-English score was below 20.”

4. 12% said “My high school GPA was about average.”

5. 12% said “I’m unsure about the rules of writing.”

6. 9% said “In high school, I did not do much writing.”

7. 6% said “I’ve used computers, but not often for writing and revising.”

Royer and Gilles argue that this survey data reveals that only 27% of the 230 students surveyed cited test scores and grades as the reason for placing themselves in ENG 098. It is unclear, however, how this data makes this case especially when we look at the remaining answers. We are compelled to ask what three quarters of these students used as criteria to make their decision. In yet another example of rhetorical cunning, the authors arbitrarily, and quite wrongly, specify that “items 1, 2, 5” of the survey “reflect self-image and self-assessment, items 3 and 4 (27%) reflect external judgments, and items 6 and 7 (15%) reflect high school or other past educational experience” (p. 62). If questions 1, 2, and 5 reflect self-image and self-assessment, it becomes a question of how self-image, as a writer in this case, and self-assessment of one’s ability as a writer can be extricated from external judgments that have been made about one’s own writing ability and therefore one’s self-assessment of that ability based on one’s self-image. The language needed to discuss this web of internal and external assessments reveals the interrelatedness of the categories. If that is not confusing enough, questions 6 and 7 are categorized as “high-school or other past educational experience.” Is it possible to extricate a person’s educational experiences from their own self-image? Can I regard
myself as a strong writer if I have continually been told by authority figures (teachers and tests) that I am a poor writer? Inversely, if I have been told that I am a gifted writer, I have received A’s in English classes throughout my education, and I have a strong degree of self confidence because of these experiences, will I really label myself in need of a developmental writing course? Whatever chemistry of experience and self-worth we cook up, the notion that educational experience, experience with tests, and self-image can be separated out like elements of the periodic table flies in the face of a great deal of highly regarded scholarship on educational experience as well as on the power of assessment.

Although there is not a lot of scholarship on DSP, O’Neill and Schendel (1999) respond to self-assessment in general and to DSP specifically. They voice concerns about Royer’s and Gilles’ assumption that

Students come to college “hyper-aware” of their capabilities and with a “good sense” of how they compare with their peers. This may be an accurate statement about students at Grand Valley, but there is no evidence given to this claim. In fact, our experience as first-year writing teachers is that students have to be taught how to evaluate their writing and that many do not have an accurate sense of their own abilities. (p. 217)

To support their own “experience” with students as first-year writing teachers, O’Neill and Schendel cite psychological research, Kruger and Dunning (1999), that “demonstrate students consistently misevaluate their performance relative to their peers” (Schendel and
O’Neill, p. 218). O’Neill and Schendel also agree that it is problematic to separate self-image, self-assessment, and external assessments. They write that

students come to college experienced with the gaze of educational assessment—both large-scale and classroom-based—and that their self-assessments and self-images may be influenced by the internalizations of others’ evaluation of them. (p. 218)

The survey that Royer and Gilles rely on reveals several facts. One, that the use of the information gathered is suspect. Can we really separate these construct of self-assessment, self-image, external judgments, and educational experiences? And, if we could, why would we want to? The answer can be found in the need of Royer and Gilles to provide the academic community with some form of authentic data on the degree to which these decisions have validity. If the designers of the survey wanted to provide useful information the design should have integrated these different constructs. The survey as it was presented to the students asked them to choose which of the “seven potential indicators most strongly influenced them to take ENG 098.” If the designers of the survey wanted to be more thorough it might have framed the question in order of influence, one being the highest and seven being the lowest, “how did these potential indicators influence your decision?” In this manner the designers of the survey could more fully realize the complex social and psychological web of influences that inform student’s decision-making, not to mention a demonstrated knowledge of the ways in which surveys can and should be used.
The irony of this casual dismissal of educational assessment’s impact on students’ self-perception is that a student’s sense of self is at the very core of DSP’s theoretical foundation. Citing the work of educational philosopher John Dewey, Royer and Gilles argue that DSP is supported by Dewey’s pragmatist philosophy of self-determination. This foundation, however, is also very problematic. It is not surprising either, that the problems with DSP’s theoretical foundation arise from a duplicitous presentation of the philosophy of James Dewey. Royer and Gilles support their rationale for student-centered educational decisions by espousing the pragmatist philosophy of Dewey’s instrumentalism, but they fail to mention one of the most prominent aspects of Dewey’s educational philosophy as outlined in many of his works: educational experience.


Royer and Gilles (2003) begin their explanation of directed-self placement’s pragmatist foundations by providing an overview of the leading scholars on writing assessment (Camp; Huot; Yancey) and distill from these scholars three important questions that all placement methods should ask:

1. How does our placement method connect with the curriculum it seeks to serve?
2. Which self does our placement procedure permit or construct?
3. What are the personal and educational consequences of our placement method?

The authors then engage their audience with the specific manner in which DSP approaches each questions. According to Royer and Gilles, DSP connects with the curriculum it seeks to serve by saving all of its energy and resources that are traditionally expended by placement methods “we call good” (2003, p. 55) by streamlining the
placement process. In short, there is little or no time spent on designing the assessment and its validity inquiry. Instead Royer and Gilles “prefer to spend it directly on teaching, grading, and program evaluation” (p. 55). One question that arises from this early statement is whether or not the validity of a program’s placement method is an important element to take into account when program evaluation occurs, but as with so many aspects of DSP this too is left unanswered. Instead, Royer and Gilles move to the question of “self” and how DSP is more equitable than traditional methods because it permits students to define themselves. This idea, that directed self-placement removes the institutional gaze of assessment from the placement procedure, is perhaps the most glaring error of all.

The authors answer the second question, “Which self does our placement procedure permit or construct?” by saying that “a unique feature of DSP is that is permits the construction of a self within a community because it fosters agency, choice, and self-determination” (2003, p. 57). The degree to which this statement concerning DSP is meaningful is explained through a theoretical explanation that links DSP with pragmatism rather than through another survey or through anecdotes (as was the case in the 1998 article). In other words, in addition to the empirical evidence offered in support of DSP, Royer and Gilles also attempt to link DSP theoretically to the prestige available in pragmatist philosophy as espoused by John Dewey, long an important authority in education.

To review the history of pragmatism is beyond the scope of this investigation, but to believe that the single decision, English 098 or English 150, that students are invited to
make under DSP has the ability to foster in students a sense of belonging to a community (the learning community of the university) seems to be a stretch. Royer and Gilles observe that “our talk with students about choices in first-year composition is just a small step toward this kind of communication and community building” (2003, p. 61). This placement decision, made after a 10 minute orientation speech, and perhaps some one-on-one advising is the extent of DSP’s involvement in community building. This placement decision, made by students, is, however, also the extent to which community building is coordinated by DSP. That is not to say that a given university might not extend this “agency, choice, and self-determination” that Royer and Gilles see as the goal of a university, but DSP, as a method, provides no further guidance for this project.

This oversimplification of pragmatism and a student centered education is evident throughout the theoretical explanation of DSP. A serious scholar must extend the questioning of this method by asking how this single choice is able to create such a change in student’s perception of education when we consider the environment from which the majority of American students have been educated as well as the environment that waits for them at so many contemporary universities. Readings explains the relatively new role of the university: “the University of Excellence serves nothing other than itself, another corporation in a world of transnationally exchanged capital” (p. 43). In this new university, the students, like all consumers, are always right. The real coup for DSP proponents is that to challenge DSP is to somehow challenge the right of the student (consumer) to play a central role in their own education.
This adversarial role of anyone challenging DSP is clearly articulated by Royer and Gilles, and, though the passage is long, the entire passage is needed for clarity and transparency:

we have been saying that agency, choice, and self-determination are the kinds of consequences DSP wants to foster in our placement procedures. Such consequences may sound naïve to a Bourdieuan intent on the ways that educational practices replicate culture and privilege in a particular way, but if such is the view we hold, then we really have no “placement problem” to begin with. What we would have in this case is the theoretical difficulty or ethical paradox of trying to extricate ourselves from the infinite regress of participating in and fostering vicious educational structures that replicate a prevailing practice that, presumably, we believe needs changing. That is why it is so important to stay oriented by the original problems that prompt inquiry, for we can easily spend all our time pondering problems that our theories created, instead of solving the problem that our students confront. (p. 60)

The ability to present one theoretical model, Dewey’s instrumentalism, while at the same time dismissing all other theoretical models as being overly concerned with “theories” rather than with the problems of our students is in itself a contradiction. More importantly, Royer and Gilles tend to push away the very notion of a rival hypothesis. Perhaps DSP will create the community of learners we in education all long for. Perhaps it will provide students with a perception of their own abilities that will in turn foster a
love of learning. However, to dismiss theories as overly theoretical while positing a theory of self-determination will not do that. Instead, Royer and Gilles exhibit one more instance of cunning in which they highlight the importance of one theoretical lens while dismissing the use of any other.

One area where DSP needs serious validation research is the question of how this pragmatist theory of self-determination is extended into the students’ four-year educational experience. Do students also get to choose the coursework they will undertake, or, as in many universities, are there required courses? Does the university have a policy on liberal education requirements (LERs) as many universities do? Are students required to attend class on a regular basis? Are students required to declare a major before their junior year? Are students required to write on certain topics within a writing course? Are students required to take a certain section, with a certain reading list, with a certain instructor whose pedagogy is not informed by a pragmatist model of education and self-determination? These questions are not meant to be trivial or sarcastic, but they are meant to call into question the reality of a single decision (one about placement in a single course) having such a lasting impact on how students not only view themselves but their own role within an institution. The DSP Brochure used at Grand Valley State provides even more evidence that students have a very limited role to play in the decisions about curriculum that will follow this initial choice. As with many consumer choices, once the initial commodity is purchased, the terms of agreement tend to be overwhelming.

As early as page one, paragraph two, students are informed that
at least three of your required classes will include a significant amount of writing instruction 150 (freshman composition) and two Supplemental Writing Skills (SWS) classes, usually taken after the freshman year. All students are also required to fulfill the junior-level writing requirement either by writing a passing two-hour essay or by taking Writing 305, a course introducing students to writing in specific academic disciplines.

If students are paying attention on that first day of orientation, it will become clear that their place in the university is not one of freedom, choice, and self-determination. As with most schools, institutional and departmental requirements abound. The very point of an institutional critique is to understand how institutions replicate the necessary conditions to assure their survival. In the case of many institutions of higher education, the ability to control the student body as well as the faculty is built into the very organization and procedures of the institutions. Faculty hiring guidelines, tenure requirements, committee and service requirements, teaching and publication requirements serve as quality control measures as well as what Foucault would call the discipline of power. Is it no wonder we call fields of study disciplines? On the same level, students must meet certain entrance requirements, maintain certain grade point averages, acquire credits, and write theses and dissertations. Royer and Gilles (2003) admit that our talk with students about choices in first-year composition is just a small step toward this kind of communication and community building…the ideal we would strive for is agency, choice, and self-
determination coupled with association, connection, and dependence—in short, the ideal of a democratic community, (p. 61)

but DSP is prefaced on the notion that this “self-determination” is a transformative experience. In light of Readings’ view of the university as a corporation and students’ role as that of consumers, the entire presentation of DSP takes on the feeling of a long and drawn out infomercial designed to “please everyone” while remaining at its core empty like “measures of excellence” that cannot be tested, measured, or validated.

Ultimately, the question of what kind of self does this placement procedure permit, like all placement methods, is enmeshed in other aspects of educational assessment. A student’s response to DSP will not and cannot happen in a vacuum. Proponents of DSP need to take into account the manner in which DSP relies, unknowingly or perhaps cunningly, on students’ past experiences with educational assessment. Royer and Gilles actually provide evidence of this reliance in the survey we discussed on page 76 where the authors state that “students took their high-school GPA and ACT-English score into account: we didn’t need to do it for them” (p. 5), and the scholarship of cultural anthropologist like Gould and Hanson that we reviewed in Chapter Two, who illustrate the ability of assessment to “invent us all” only enforce the power of these experiences.

In many ways it is the very notion of experience itself that needs to be looked at more closely. Royer and Gilles cite Dewey in terms of pragmatism, but little is said about Dewey’s theorizing on the role of experience in education. Under the lens of instrumentalism, as they note, experience for Dewey is about experiencing “an-
‘indeterminate situation’ that will promote inquiry” (p. 59). Dewey, however, discusses experience and education in more than one manner. Dewey’s aptly titled *Experience and Education*, written in 1938, approaches experience in terms of past experiences. When he writes that “everything depends upon the quality of the experience which is had. The quality of the experience has two aspects. There is an immediate aspect of agreeableness or disagreeableness, and there is its influence upon later experiences” (p. 27) we can see both experience as present and experience as past.

More specifically, Dewey’s notion of the “experiential continuum” helps to achieve an understanding of how students’ entire lives, especially in the current atmosphere of test-based reform, have been powerfully influenced by educational assessment. One could argue the influence of educational assessment on students’ self perception far out weighs the ability of a one time choice concerning “English 098 or 150, which is right for you?”

To understand how important Dewey believed past educational experiences were to an individual’s view of education and their role within education we need only to revisit his writing’s on the subject. Dewey writes:

>How many students, for example, were rendered callous to ideas, and how many lost the impetus to learn because of the way in which learning was experienced by them? How many acquired special skills by means of automatic drill so that their power of judgment and capacity to act intelligently in new situations was limited? How many came to associate the learning process with ennui and boredom? How many found what
they did learn so foreign to the situations of life outside the school as to give them no power of control over the latter? (pp. 26-27)

This passage alone raises several questions. One, what effect has standardized testing and other prevalent forms of assessment had on students’ views on education in general and on themselves as students specifically. Two, when faced with important decisions about new environments, like “English 098 or 150, which is right for you?”, how prepared are students to understand the new discourse community that college writing demands in comparison to their high school discourse community, remembering Schendel and O’Neill’s comment that “our experience as first-year teachers is that students have to be taught how to evaluate their writing and that many do not have an accurate sense of their own abilities,” (p. 217) and Kruger’s and Dunning’s work (1999) that demonstrates “that undergraduate students consistently misevaluate their performance relative to their peers” (Schendel and O’Neill, p. 218). Three, how many students, because of past experience, would forego the idea of a course that provides no credit toward a degree, while providing all of the work and responsibility of a credit earning course? This is a question that Royer and Gilles leave unanswered. In fact, Royer’s and Gilles’ own statistics show a drop of 11% in students that were enrolled in ENG 098 between 1995 (33%) and 1996 (22% the year DSP was implemented). Their response to this drop is that “for the first time we feel that the right students are taking our developmental writing class” (1996, p. 61). Although this 11% drop in ENG 098 enrollment is passed over, Royer and Gilles address the fact that
We do know that about 66% of our 1995 ENG098 students went on to earn credit for ENG 150 by the end of the next semester, while just 55% of the 1996 group did the same. The difference seemed to be that while in the 1995 about 87% of the ENG 098 students went on to take ENG 150 the next semester, in 1996 only 75% of our ENG 098 students took ENG 150 the next semester. (p. 64)

The authors have no answers for this trend and decide that “even with questions like these unanswered, we are convinced that directed self-placement is working at our school” (p. 64). It is almost mysterious, dare I say it just doesn’t feel “right” that scholars are convinced when there is no research to enforce their confidence. The cunning dismissal of research comes on the following page as the first line of the section entitled “Directed Self-Placement Pleases Everyone Involved.” They write that “to analyze numbers is, to some extent, to fall back into the thinking that what’s most important about placing students in developmental or regular first-year writing courses is a quantifiable assessment of writing ability” (p. 65), another statement that is somewhat mysterious. How does data that reveal the successful completion of a course by a certain number of students quantify the construct of “writing ability”? The research that is needed, the research that Royer and Gilles are trying to produce, is data about the degree to which the decisions made by students about the most appropriate writing course for their own personal situation is valid. If students earn credit for ENG 098 or ENG 150 it seems fair to argue that they placed themselves properly. These data do not, however, quantify writing ability. The grade students received for passing the course, an A, B, or
C, would quantify writing ability to a degree. Though as writing teachers are aware, a grade for a course takes into account many factors other than writing ability: participation, attendance, etc.

More importantly, Dewey’s view of educational experience makes clear the fact that placement procedures cannot operate outside the larger experiential-continuum of the students that are being placed—or under DSP are placing themselves. DSP, like any other placement procedure, is part of a larger web of educational assessment. The reality that students’ lifelong experiences with educational assessment influence their decision-making process is not an issue that Royer and Gilles address in either of their theoretical discussions regarding DSP (1998; 2003). Instead, Royer and Gilles disregard, as noted previously, the powerful influence of past experiences by considering reproduction in education as overly theoretical. However, Dewey, whose philosophy informs DSP, writes that “failure to take the moving force of an experience into account so as to judge and direct it on the ground of what it is moving means a disloyalty to the principle of experience itself… He is also unfaithful to the fact that all human experience is ultimately social: that it involves contact and communication” (p. 38). One could argue that DSP removes the burden (pressure, responsibility, cost, time) from the teachers (they don’t have to read essays or review ACT-English scores) and administration (they don’t have to design lengthy placement procedures or disrupt their orientation plans or pay for papers to be read) and places it squarely on the shoulders of the students. One could also argue that, contrary to the motives outlined by Royer and Gilles, this relocation of the placement burden is in fact a form of institutional cunning that relies on the past
experiences of students to do the work of sorting and placing. In other words and in a Bourdieuan sense, DSP relies on the past experiences of students with educational assessment to reproduce socio-economic relations while providing the institution with an efficient (time and money saved) placement procedure that “pleases everyone involved”.

The third and final question that Royer and Gilles present in their 2003 piece asks, “What are the personal and educational consequences of our placement method?” In their 1998 article they answer this question with the very optimistic fact that “the peculiar feature of directed self-placement that, on one sense, it can’t really fail,” because “if they fail, they will, we hope, learn that a college education is a serious endeavor and that success often begins with a proper estimation of one’s abilities” (p. 70). Talk about tough love.

By 2003 the authors have a much more lengthy examination of the consequences of DSP. The length, however, seems to recapitulate the same critical errors. Under the heading “Consequence 1—Summer Orientation” the authors argue that students who choose to begin with the basic writing course signal to us a desire to take things slowly, to test the waters of college writing before jumping into the standard course. Students who choose to begin with the standard course—including those who might previously have been labeled basic writers—signal to us a desire to dive right in. (pp. 64-65)

The decision to sink or swim, however, is still based on the authors’ assumption that “just because they may struggle with certain aspects of writing, we need not assume that they cannot be thoughtful about their own experiences and abilities and act responsibly in their
own best interest” (p. 64). This view, seven years later, does not address or even recognize the research on and published warnings of writing assessment scholars, specifically Schendel and O’Neill, concerning DSP’s lack of validity inquiry. Schendel’s and O’Neill’s (1999) “Theories and Consequences of Self-Assessment,” as we have seen, question many of the assumptions made by the 1996 Royer and Gilles presentation, several of which we have already noted. Schendel and O’Neill further warned that in the original scholarship on DSP “left out are how race, class, gender, or disabilities might affect students’ self-assessments—information that we think warrants research and discussion because such information is linked both to validity of the assessment and the ethics of the assessment practice” (p. 219). The 2003 edited collection does provide an in-depth account of “The Role of Self-Efficacy” in DSP. Again, the flaw in this research is the assumption that “self-efficacy” defined as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” (Bandura, 1977, p. 391 in Reynolds, p. 74), can be transposed over the complex social performance of writing. Reynolds provides the caveat that “although self-efficacy has been studied in relation to many areas of academic performance, unfortunately there have been relatively few studies analyzing its specific role in writing outcome expectancy” (p. 74). At the same time, Reynolds writes that “research focusing on self-efficacy’s role in relation to student writing suggests that confidence in one’s ability to write well and actual writing ability have been shown to possess a high degree of correlation” (p. 74). Again, the central question concerning the consequences of DSP revolves around experience: self-perception and self-confidence as Dewey explains is
united with educational experiences. Reynolds as well as Royer and Gilles find some way to separate a student’s previous experience from the student’s self-perception. In the same manner, self-efficacy is somehow separated from self-confidence when Reynolds argues that “directed self-placement must be considered a justifiable agency for placement because there are very definite theoretical bases for its utilization—students can reliably assess their writing skill and, one would presume, accurately place themselves accordingly” (pp. 73-74). Can high-school students assess their writing ability in terms of university level writing without know the expectations and conventions? Perhaps, but proponents of DSP neither produce nor attempt to produce this evidence. In the place of this evidence we instead get “feelings rightness” and assumptions about student ability to understand a discourse that they have yet to encounter.

Conclusion

The idea that students have control over educational decisions that affect their current and future opportunities is a positive idea. To criticize a procedure that posits a notion of self-determination and a community of learning is not a pleasant task. The consequences of scholarship and procedures presented in the manner of directed self-placement, however, must be understood in their complexity and must be critiqued with the kind of scholarly rigor that educational policies deserve. As noted, DSP, for all of its good intentions, is supported by a cherry picked theoretical foundation and presented with the duplicity of a commercial commodity: two characteristics that undermine the very principles of a humanist tradition of education within the United States of America.
One consequence that DSP white washes is the financial and emotional fallout connected with making the wrong choice. On one hand, those students who aim too high fall back to earth with the shock and disappointment of failure. On the other, those who find themselves unchallenged by a developmental writing course waste both time and money, not to mention the ennui that comes with an unchallenging academic experience. These effects, perhaps, are not that dire and could be better understood if retention data were collected. As Royer and Gilles end their 1998 piece, and as I have already quoted, “if they fail, they will, we hope, learn that a college education is a serious endeavor and that success often begins with a proper estimation of one’s abilities” (p. 70). The more dangerous effect of scholarship and placement procedures that fail to adhere to a unified theory of validity and the rigorous validation processes that this entails, returns the field of composition studies to a time when our principles and practices were ignored and deemed the work of rote drilling and repetition.

It has been the struggle of compositionists of the last fifty years to theorize the social and cognitive nature of writing as a complex performance and to enact pedagogies that implement these theoretical principles that have increased the stature and funding of the discipline. This stature and funding is evident in the increased graduate degree granting institutions nationally and internationally, and it is this stature that allows issues of progressive pedagogies and theories to push back against the reductive forces of test-based reform. In the end, DSP compromises the theoretical and humanistic principles of writing scholarship by compromising what we have learned about the social and cultural influences on students’ views on themselves and writers. The very notion of a process
oriented approach to writing was a reaction to the product fixated theories of knowledge and learning that privileged certain ways of knowing. It looks for all intents and purposes that DSP returns composition studies to a time when those that are fortunate enough to experience the benefits of an unequal system move forward, while those that are less fortunate are left behind. Under the guise of self-determination, however, no one is to blame but the individual. After all, they had a choice.
Chapter Four:

The ACT Writing Exam at CUNY: A Technology for Reproduction

Literacy is part and parcel a relationship that involves the vertical and horizontal exchanges of the means of livelihood in a literate society. Thus, it is the relationships of literacy, which a society bent on unequal distribution of wealth and power dominates completely, that literacy educators must understand in order to proceed in ways that do not implicate them in the domination.

J. Elspeth Stuckey in *The Violence of Literacy*

The previous chapter investigated the manner in which the placement procedure known as directed self-placement was presented to the writing assessment community through duplicitous rhetoric and cherry picked philosophical and theoretical principles. As noted in chapter three’s conclusion, the type of scholarship presented by Royer and Gilles is dangerous on several levels. Their scholarship compromises the principles and practices of a socio-cognitive model of literacy, misrepresents the work of several leading writing assessment scholars, and dismisses the role that the accumulated effects of educational assessment play in influencing student decisions while claiming that this influence is absent from the decision-making process. When all is said and done, however, the case of directed self-placement is ultimately a case of low-stakes or perhaps lower-stakes assessment. Although students might waste money and/or time taking an
unnecessary course or a non-credit earning course, they will, in the end, be allowed to continue with their educations. Chapter Four focuses on a case of high-stakes assessment practices implemented by political leaders and policy-makers at The City University of New York college system that have the capacity to deny students the educational opportunities of higher education.

The battle over remediation and open admissions at The City University of New York (CUNY), a large system (17 separate campuses) of individually administered four-year senior colleges and two-year community colleges, has seen internecine outbreaks since the Civil Rights movement won the battle for open admissions in 1970. The most recent outbreak of hostility towards open admissions came, unsurprisingly, during the late 1990’s as the tide of test-based reform that would usher in No Child Left Behind began to gain momentum. At the behest of the then Mayor of New York City, Rudolph Giuliani, and the then Governor of the state of New York, George Pataki, both CUNY’s open admission’s policy and remedial education programs became the focus of their educational reform policies. Like the president would do in campaigning for test-based reform, the mayor and the governor argued for “higher standards,” “accountability”, and “objective” forms of assessment to achieve the goal of “excellence in education.”

This chapter’s investigation of standardized writing assessment practices introduced into the CUNY system as instruments of reform focuses on the manner in which these reforms serve to reinforce and reproduce the social conditions necessary for production. The primary artifacts through which this investigation is undertaken are the 109 page report produced and disseminated by the Mayor’s Advisory Task Force on The
City University of New York, June 7, 1999, entitled *An Institution Adrift*, and several articles of scholarship produced by scholars within the CUNY system\(^{18}\). Newspaper reports, predominantly from the *New York Times*, provide context as well as insight into the stated intentions, motivations, and reactions of proponents and opponents of the reforms.

The chapter comes in three parts. Part one contextualizes the problems identified within the mayor’s task force’s report as another event in a series of perceived literacy crises in American education. Part two illustrates a conservative agenda for privatizing public education and for implementing test-based reforms as the solution to perceived literacy crises. Part three examines the specific case of the ACT-writing exam at CUNY at Kingsborough community college to understand how standardized writing assessments affect writing program policies, pedagogical practices, and students’ attitudes towards writing. In aggregate, these three sections reveal the ability of writing assessment practices to reinscribe socioeconomic relationships by exploiting underprivileged literacy practices.

*Part One: the Mayor’s Task Force and CUNY*

To properly understand the enormity of the effect that the mayor’s task force’s report produced through the policies and procedures that were implemented at its behest,

\(^{18}\) Bruce Chadwick’s “Tilting at Windmills: The CUNY ACT Writing Exam” and Barbara Gleason’s “Evaluating Writing Programs in Real Time: The Politics of Remediation”.

we must first have a clear picture of the CUNY system for which the recommendations were made. As the report informs us, “CUNY’s more than 200,000 students are an extraordinarily diverse and ambitious group” (p. 12), and the “ethnic composition of New York City is undergoing a dramatic change. During the 1990’s the white population of New York City declined 19.3% while the black, Hispanic, and Asian populations have risen 5.2%, 19.3%, and 53.5%, respectively” (p. 13). The high rate of minority students, cites the report, directly accounts for the fact that “CUNY conducts remediation on a huge scale: in 1997, 87% of community college freshman and 72% of senior college freshman failed one or more of CUNY’s remediation placement tests, and 55% of CUNY freshman failed more than one” (p. 6). The importance that the task force places on this data, (1) the number of non-traditional students and (2) the level of remediation being undertaken at CUNY, is linked to the task force’s concern that CUNY approaches remediation in a flawed manner and that these remedial students pose a grave threat to both the CUNY system and the non-remedial students attending. The task force cites the following as costs of flawed remediation:

- students who confront life without basic skills;
- the waste of unprepared students’ time in classes that are incomprehensible;
- poor use of prepared students’ time, to the extent that courses are watered down;
- the distraction of professors from college-level teaching; and
- the erosion of standards that will result. (p. 7)
To repair this flawed remedial system, the task force recommends that:

- CUNY must replace its current student assessment program with one that is consistent with modern assessment science.
- CUNY must evaluate the effectiveness of its various remedial programs according to objective standards, and must hold administrators and instructors accountable for results.
- Students who require remediation should be given a range of remediation options funded by education and training vouchers from a mix of public providers…CUNY and the City should conduct a pilot project in outsourcing remediation services, in order to stimulate competition and generate performance data from the various providers.

(p. 7)

The “problem” that the task force outlines in the early stages of its report is a problem that America’s educational community has confronted time and again. As mentioned in chapter one, Martin Nystrand (2005), in “The Social and Historical Context for Writing Research,” reveals the recurrence of the “problem” throughout the history of public education in the United States. Nystrand cites Time’s “Bonehead English” article of 1974 that decried the “sharp increase in unprepared students,” and Newsweek’s 1975 cover story “Why Johnny Can’t Write,” which “blamed the public schools for neglecting ‘the basics’” (p. 15), as examples of a series of successive crises that traces back to the 18th century. Likewise, Neal Lerner charts the history of writing centers in postsecondary
education as a liberal response to changing populations of students at the postsecondary level. In fact, the interest in intelligence testing at the turn of the century that was the beginning of the measurement community was largely due to laws mandating universal education for all children. Feeling overwhelmed by the new population of students and dismayed at their failure, intelligence testing was seen as a way to classify these new students for specialized instruction. The crisis at CUNY is yet the newest addition to this series of events. These “problems” share the same characteristics. In this instance, the student body has changed from a homogenous group of white students to a diverse group of minority students. The learning and literacy habits of these students are as diverse as their racial and ethnic origins, and the pedagogical practices necessary to help these students succeed are just as diverse.

The Politics of Literacy

The “solution” for this educational problem has always been a major point of contention and the political left and the political right holds, not surprisingly, polarized views. Historically, the left has addressed the problem through remediation, writing centers, tutors, community outreach programs, after school programs, and progressive pedagogical practices. On the right, the historical “solution” has been to argue for higher standards and accountability through standardized (so called “objective”) testing. The core of the conservative argument being that those struggling to learn have either failed themselves or been let down by poor teachers and schools. The ability to learn and the
manner by which one teaches are static\textsuperscript{19}. The solution is, therefore, to hold students and teachers accountable to higher standards. Through standardized testing a system of merit (a meritocracy) will emerge, and, as we saw in chapter two, the notion that a test is objective has been proven time and again to be the material of myth (Hanson; Gould; Lehmann; Sacks; Shohamy; Bracey). In fact, scholarship on testing informs us that standardized testing privileges a particular way of knowing, and, through the work of scholars like Shirley Bryce Heath and Brian Street, we have learned that this particular way of knowing, learning, and testing is particular to a certain group of people—namely those that are white and middle to upper middle class.

In the field of literacy studies this debate between left and right, liberal and conservative, is encapsulating in the notion of the “great divide”. The political and social ideologies that inform the polarized battle over education and educational reform argue from different sides of this divide. Shirley Bryce Heaths’ \textit{Ways with Words} provides an ethnographic study of the different literacy and learning habits of a predominantly white community (Roadville) and a predominantly black community (Trackton). This groundbreaking study offers scholars interested in literacy, pedagogy and learning a more nuanced understanding of the role that family and community discourse habits play in the way individuals learn. This social constructivist understanding of literacy and learning

\textsuperscript{19} Also known as a telementational model. The teacher holds the information, lectures, and passes the information from his mind to the students. See Roy Harris’ \textit{The Language Myth}. 
(Vygotsky; Heath; Street; Brandt) recognizes the need to provide multiple approaches to teaching and learning when dealing with a diverse student population.

Brian Street’s contribution to literacy studies helped redefine academic notions of literacy. Prior to Street’s publication of *Literacy in theory and practice*, literacy was understood in terms of what Street refers to as the “autonomous model.” The autonomous model cleaves to the idea that literacy, defined as being able to read and write,

fosters or even enforces the development of logic, the distinction of myth from history, the elaboration of bureaucracy, the shift from little communities to complex cultures, the emergence of scientific thought and institutions and the even the growth of democratic institutions. (p. 44)

This notion, as Street points out, is tied to a host of socio-political issues that deems “literate” cultures as superior to “oral” cultures, and, like Heath, is more closely aligned with the kind of literate practices privileged in traditional “schooling”. The autonomous model is also tied directly to a social order that profits through hegemonic control via literacy training. For Street, scholars like Ong, Havelock, and Goody fail to take into account the political, social, and cultural realities that surround orality and literacy. To this end, Street argues that “in Goody’s descriptions of literacy practice, there is a peculiar lack of sociological imagination in his determination to attribute to literacy per se, characteristics which are clearly those of the social order in which it is found” (p. 45). Street’s “ideological model”, in contrast, is positioned to understand literacy in terms of its role as a social, cultural, and political technology. For Street, writing is an instrument
to be used by an agent. Thus he characterizes Goody’s notion of “the technology of the intellect” (p. 97) as misleading. Street argues that “no one material feature serves to define literacy itself. It is a social process, in which particular socially constructed technologies are used within particular institutional frameworks for specific social purposes” (p. 97). Street’s language makes his definition of literacy is functional. The idea that “particular” technologies (writing) are “constructed” for and “within particular” social frameworks for “specific social purposes” makes it clear that literacy refers to being able to operate within a specific and particular culture.

Through the work of scholars like Heath and Street the importance of understanding literacy in terms of “particular communities,” “actual forms and functions,” and in terms of “traditions” becomes evident. Heath’s findings support the reality of Street’s ideological model. She found that the residents of Trackton were literate, but that their “particular” community had a very different relationship to writing, reading, and orality than did the practices of their white counterparts in Roadville. The residents of Roadville used and learned their literate practices in the same way that the mainstream schools of these communities expected students to acquire literacy. It should be noted that mainstream schooling is still very much based on an autonomous notion of literacy. The students from Trackton, though literate, had a very different kind of relationship towards literacy and orality because of the “particular” practices of their “specific” community. Heath comes to the realization that a literacy continuum serves no point. It is the “particular” and “specific” aspects of a community that defines what “literate” means.
It is not surprising, therefore, to hear the language of the “autonomous” model of literacy represented in the task force’s “solution” to the perceived educational crisis occurring at CUNY. The defining of “what literate means,” in this case as in most educational policy cases, is in the hands of policy-makers and politicians, and their policies of standardized writing assessments have the ability to mandate a curriculum informed by the “autonomous model” of literacy.

In the task force’s report can be seen the plans for a corporate model of efficiency and accountability transposed over the organization of the CUNY system. These plans are couched in the language of educational test-based reform and together with the corporate model merge into a single conservative view of both the nature of literacy as well as the mission of a contemporary institution of higher education.

The Future of Education

The rhetoric of corporatization and test-based reform found in the task force’s report reveals a certain ideological understanding of the nature of literacy as well as a plan for the future of education. Gerald Bracey begins chapter four of Setting the Record Straight: Responses to Misconceptions about Public Education in the United States with a question and answer. The question: “What do I say when someone says, ‘SAT scores have plummeted’”? The answer: “You can say, ‘No, they haven’t.’” People who say the scores have plummeted are invariably critics of the schools who use emotion-laden words like plummet, plunge, and nose-dive to persuade rather than to convey facts” (p. 53). With this in mind, the mayor’s task force’s report’s title provides an unsettling beginning, An Institution Adrift, and continues as in the opening page of the “Executive
Summary” the audience is informed that “CUNY is adrift…CUNY is currently in a spiral of decline…accountability is largely ignored…academic standards are loose and confused…there is tragic and personal loss and institutional waste implicit in CUNY’s high dropout rate and low graduation rates” (p. 5). Obvious among the emotional language used to describe the drastic and dire situation is the non-referential, qualitative standard described by Bill Readings as the moniker for the corporatized University. CUNY, according to the report, is “potentially a model of excellence…CUNY should be that model of excellence…CUNY must reinvigorate it commitment to excellence…” (5-6). The fact that the Chairman of the CUNY Task Force, and retired President of Yale University, Benno C. Schmidt, Jr., was at the time of his appointment, and remains today, one of the top executive officers of the Edison Project, a corporation which defines its primary mission as the privatization of public education\(^{20}\) as a for-profit industry, is worthy of some concern.

The obvious concern that arises from this situation is the relationship between interested parties that believe in the privatization of public education, and the use of standardized tests to represent institutions of public education as failures. Gerald Bracey

\(^{20}\) The Edison Project, now The Edison Learning Institute, was founded by Chris Whittle. Whittle is famous for founding Channel One, a service that provides free televisions and DVD/VHS players to schools that broadcast Channel One throughout the school day. This brought advertising directly into the public schools. See Michael Apple’s Official Knowledge for a critique of Channel One as well as the neoconservative campaign to privatize public education.
(2003), writing about No Child Left Behind’s use of testing, argues that “the goal of NCLB is the destruction of public schools, not their salvation. NCLB sets the schools up to fail and be privatized” (p. 69). The comprehensive implementation at CUNY of test-based reforms in conjunction with the conservative interests of both the mayor and the chairman of the task force call the motives for implementing such reforms into question.

Part Two: The Technology of Reproduction

More disconcerting than the emotional language that paints an apocalyptic educational scene of CUNY’s eminent demise, more disconcerting than the conflict of interest represented by the Chairman’s professional position is the diagnosis the task force produces and the recommendations that follow. As we have seen, the task force is quick to point out the CUNY schools are primarily populated by a “diverse and ambitious” student body. They provide additional data that reveals that, of the 200,000 CUNY students, “just over one-third at the senior colleges and just under one-half at the community colleges” report a “household income under $20,000” (p. 25). This number is staggering when one takes into account the correlation that exists between household income and SAT scores.
The task force’s reliance on SAT scores becomes even more troubling when the task force employs these test scores for the purpose of comparing the quality of the students attending CUNY schools to other state college systems. One must remember that the task force has already pointed out how unique this student population is compared to the rest of the country. Comparing national tests scores leads the task force to reveal that CUNY does not have a single four-year college that is in the top tier of public institutions nationwide in the academic quality of its entering students. Similarly, based on the limited information that is available, not a single CUNY senior college has a graduating cohort that would rank above average among American college graduates” (p. 6).

The task force legitimizes these comparisons and the failure it implies of the CUNY system, when it claims that “research has shown the SAT to be a reliable, objective indicator of students’ academic readiness for college” (p. 23). A different view of the
SAT, provided by FairTest, the National Center for Fair and Open Testing\textsuperscript{21}, indicates a very different view of the SAT and cites study after study providing research data that shows the SAT to have a very low degree of validity for predicting freshman success in college.\textsuperscript{22} Regardless of the well published facts about the SAT as a poor predictive

\textsuperscript{21} Fairtest.org

\textsuperscript{22} The SAT I is designed to predict first-year college grades - it is not validated to predict grades beyond the freshman year, graduation rates, pursuit of a graduate degree, or for placement or advising purposes. However, according to research done by the tests' manufacturers, class rank and/or high school grades are still both better predictors of college performance than the SAT I. The College Board and ETS conduct periodic studies of the SAT I. This usually involves examining the relationship between test scores and first-year college grades, generally expressed as the correlation coefficient (or r value). The College Board's Handbook for the SAT Program 2000-2001 claims the SAT-V and SAT-M have a correlation of .47 and .48, respectively, with freshman GPA (FGPA). This number is deceptive, however. To determine how much of the difference in first-year grades between students the SAT I really predicts, the correlation coefficient must be multiplied by itself. The result, called r squared, describes the difference (or variation) among college freshman grades. Thus, the predictive ability (or r squared) of the SAT I is just .22, meaning the test explains only 22\% of the variation in freshman grades. With a correlation of .54, high school grades alone do a better job, explaining almost 30\% of the variance in first-year college performance.
instrument for college success and its correlation to household income, the task force informs its audience that

RAND found that the SAT scores of the one-third of CUNY freshman who took the SAT correlated reasonably well with the scores the students got on the reading and math tests (RAT and MAT) CUNY gives all freshmen to determine who needs remediation. Using the fall 1997 entering classes, RAND was able to extrapolate from the RAT and MAT test score taken by all freshmen to create a comprehensive academic picture of CUNY’s incoming freshman.²³

That the task force believes an extrapolated score presents a valid and “comprehensive academic picture of CUNY’s incoming freshman” reveals a very reductive outlook. That the complex intellectual and educational abilities of an entire freshman class can be reliably or validly extrapolated and ranked reveals the kind of positivistic bias at the origins of traditional test theory. The idea of ranking intelligence, as we saw in chapter two, is a traditional view of education that has historically argued that a student’s ability can be “objectively” measured and explained by a single, one time standardized test. This is what Alan Hanson refers to as the priority of potential over performance at work. He writes that

the fabricating process works according to what may be called the priority of potential over performance. Because tests act as gatekeepers to many

²³ See appendix 3
educational and training programs, occupations, and other sectors of social activity, the likelihood that someone will be able to do something, as determined by a test, becomes more important than one’s actually doing it.

(p. 288)

The pro-testing bias that we explored in chapter two and the positivist tradition that argues for a single, observable reality and an objective means of gauging that reality is found in the language, findings, and recommendations of the task force. The history of standardized testing, also explored in chapter two, tells the story of testing’s origins in a racial ideology of superiority that saw the European white male as the intellectual elite.

One final correlation of SAT scores:

<table>
<thead>
<tr>
<th>ETHNICITY</th>
<th>READING</th>
<th>MATH</th>
<th>WRITING</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amer. Indian or Alaskan Native</td>
<td>487 (0)</td>
<td>494 (0)</td>
<td>473 (-1)</td>
<td>1454 (-1)</td>
</tr>
<tr>
<td>Asian, Asian Amer. or Pacific Islander</td>
<td>514 (+4)</td>
<td>578 (0)</td>
<td>513 (+1)</td>
<td>1605 (+5)</td>
</tr>
<tr>
<td>African American or Black</td>
<td>433 (-1)</td>
<td>429 (0)</td>
<td>425 (-3)</td>
<td>1287 (-4)</td>
</tr>
<tr>
<td>Mexican or Mexican American</td>
<td>455 (+1)</td>
<td>466 (+1)</td>
<td>450 (-2)</td>
<td>1371 (0)</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>459 (0)</td>
<td>454 (-2)</td>
<td>447 (-1)</td>
<td>1360 (-3)</td>
</tr>
<tr>
<td>Other Hispanic or Latino</td>
<td>459 (+1)</td>
<td>463 (0)</td>
<td>450 (+1)</td>
<td>1373 (+2)</td>
</tr>
<tr>
<td>White</td>
<td>527 (0)</td>
<td>534 (-2)</td>
<td>518 (-1)</td>
<td>1579 (-3)</td>
</tr>
<tr>
<td>Other</td>
<td>497 (+3)</td>
<td>512 (-1)</td>
<td>493 (0)</td>
<td>1502 (+2)</td>
</tr>
<tr>
<td>No Response (9%)</td>
<td>480 (-7)</td>
<td>497 (-9)</td>
<td>474 (-8)</td>
<td>1451 (-24)</td>
</tr>
</tbody>
</table>

Table 4.2 Fairtest.org

The unquestioned and unexamined legitimacy of testing that informs the task force’s recommendations and is revealed in their rhetorical choices has consequences. One consequence of blindly ignoring the well published bias and limitations of standardized tests leads the task force to several predetermined conclusions, tautological
arguments, and unmasks the cunning necessary to fulfill the institutional demand that socioeconomic relationships be reproduced.

The tautological argument of the task force begins with the premise that CUNY is failing in its “historic mission—to provide broad access to a range of higher education opportunities of quality suited to New York’s population and to the City’s needs” (p. 5). CUNY is failing in part because “it is inundated by NYCPS graduates who lack basic academic skills” but is to blame because “it has not made a strong effort to get the public schools to raise their standards” (p. 5). Charged with providing open admissions and inundated with unprepared students from NYCPS, CUNY’s answer to the Catch-22 of this inundation and this unique mission has been to provide remediation in order to prepare the unprepared for future educational challenges. However, the task force goes on to say that “the Mayor and the Governor need to work together and with the State legislature to revamp financial aid policies to ensure that students can obtain remediation without depleting federal or state financial aid that is intended to support college-level work” (p. 7). The wheel of the argument turns in this way. CUNY is faced with a “diverse” population, unprepared due to a failed public school system that accounts for two-thirds of CUNY’s enrollment and must therefore offer remediation on a level more than double that of the national average (p. 19). This need for remediation and the State’s financial aid policy are in conflict. Students run out of financial aid because they are taking remedial education courses that do not earn credit, they dropout because their funding expires, they fail to graduate because they dropout and this in turn is cited as the failure of CUNY to meet its “historic mission.”
The task’s forces solution is to demand higher standards for entrance, less remediation, and more accountability by implementing standardized assessment practices (which, by the way, implies the end to open admissions). In short, the task force’s answer is to change the student body so that CUNY can claim that it is a successful educational system. In their words, “CUNY’s drift must me arrested by a comprehensive strategy of institutional renewal. Clear, rigorous standards of academic achievement must be the foundation of CUNY’s commitment to educational opportunity” (p. 20). The task force also notes that “we suspect that we have been more successful in identifying problems that need urgent attention than in fashioning particular solutions” (p. 20). Their solutions in all areas concerning the student body are directly linked to the use of standardized testing and privatization.

On addressing the problem of open admissions and remediation they recommend both testing and privatization. They argue that remediation “should reside in the community colleges, not in the senior colleges” (p. 36). They also argue that “one of the most constructive steps that CUNY can take to renew itself and to help NYCPS raise standards is to promulgate a spectrum of clear, objective admission requirements, keyed to the academic preparation required for different types of institutions and programs” (p. 37). These “objective requirements”, of course, would be measured by “SAT scores of 1200” for the “Regents Diploma,” and would regress downward in a tiered system to an “SAT score of 800 at the community college level” (pp. 37-38). Other recommendations shift paying for remediation to the NYCPS, revision of the Tuition Assistance Program (TAP), a “tapping” of other public funding educational projects, and, finally,
“competition and choice of remediation” in which “New York City implement a managed competition model of providing postsecondary remediation services: students who need remediation should be given educational vouchers and permitted to purchase remedial educational services from the provider of their choice” (p. 41). The notion that private providers in competition with one another will, due to market forces, drive down the price and improve the effectiveness of instruction is the core principle that informs all privatization practices, from military food services and laundry services to private security services and once publicly owned utility services. In fact, a New York Times article of March 6, 2000, reveals how quickly the task force’s recommendation was being taken advantage of by private interests to the detriment of the students. The Times informs us that “the mayor’s task force suggested the university pay for outside help, but the CUNY plan calls for students to pay their own fees…and that outside companies usually charge more than CUNY” (Arenson,1).

The conservative “solution” to what the task force perceived to be problems at CUNY was an expansive policy of test-based reform that mandated standardized assessments not only as entrance measures but as exits measures as well. The task force, in several areas of their report, argues for uniformity among the different colleges, both senior and community. The task force’s recommendations are analogous to a corporate restructuring in which the administration of several colleges (corporations) is streamlined under a single administration. This streamlining removes redundancies and makes the system more efficient. No where is this streamlining and policy of uniformity more
evident than in the task force’s recommendation for amending admissions standards and organizing the colleges on a tiered system:

1. To assure that New Yorkers have the opportunity for public higher education of the highest quality, CUNY should mandate that at least two senior colleges adopt rigorous standards of admissions: SAT scores in the vicinity of 1200 and high school class rank in the top 10 to 15 percent.

2. A second tier of senior colleges should have admissions standards that are somewhat more relaxed but that still ensure that students have the capacity for baccalaureate programs of clear and consistent quality: SAT scores in the vicinity of 1000 and high school class rank in the top 25 percent.

3. A third tier of senior colleges…SAT 800 and class ranking in the top half.

4. Community colleges should be open to all other high school graduates, either for associate degree programs or remediation. (pp. 37-38)

When you take this tiered system and compare it to household incomes and their correlation with SAT scores, the level at which the task force has devised a system that reproduces the means of production and reinscribe already existing socioeconomic relationships is stunning. The tiered system represents an end to open admissions for the majority of the CUNY system and creates an environment in which, according to the task force’s own data (see chart three), the vast majority of CUNY’s current student body would be denied access to all but the third tier of senior colleges and the community colleges.
The uniformity that the task force seeks among the colleges assumes that each college is providing for the same population. At one point they state that “CUNY’s haphazard evolution—characterized by rapid expansion, sudden change of academic direction, and frequent turnover at the administration level—has resulted not in a coherent university, but in a loose confederation of individual colleges” (p. 81). This “loose confederation,” however, is a natural product of institutions forming to meet the local needs of individuals and communities. In a rebuttal to the task force’s report, The Friends of CUNY, argue that “in pursuing this ‘grass must be greener someplace else’ vision, the commission again makes the mistake of assuming another model is right” (Arenson, Rebutting Task Force, 1). Again, the task force, because of an ideological tendency to value efficiency, accountability, and standards over a more nuanced understanding of student needs approaches educational decisions as if they were business decisions (and, for the task force, these are one and the same).

Part Three: The ACT Writing Exam: Writing Assessment as Social Reproduction

The task force’s argument for uniformity is applied by the task force to the teaching of remedial writing and serves as another example of literacy being used as a means by which to limit the opportunities of certain groups. To address the perceived problems in remedial writing programs across the CUNY system, the task force recommended that the schools replace their own placement and exit procedures with either the ACT or the SAT writing examinations. They write:

CUNY’s remediation instructors are largely free to design the courses they teach. There is nothing wrong with a system of
institutions such as CUNY approaching remediation in different ways. A healthy competition among approaches will provide innovation, efficiency, and choice. But this only happens if there is an effort to measure what works and useful information about outcomes is made available. (p. 31)

This policy of uniformity, like the task force’s unquestioning support of testing’s objectivity, assumes that each campus is populated by the same students and faces the same challenges. This view of education harkens back to the discussion of Michael M. Williamson’s “The Worship of Efficiency” in which the author explains the historical tendency of administrations to seek cost and time efficient measures of assessment that provide solid, seemingly objective data, while teachers struggle to implement assessment practices that adhere to theories of language and learning. The belief that assessment practices need to be anchored to local and contextual factors is widely shared. For instance, Brian Huot (2002), writing in (Re) Articulating Writing Assessment, explains that

there is a clear link between the judgments being made and the outcome of these judgments that is neither hidden nor shaded by reference to numerical scores, guidelines or statistical calculations of validity or reliability. These site-based, locally-driven procedures for evaluating student writing have their roots in the methods and beliefs held by the teachers who teach the courses that students are entering or exiting, or in the program under review. (p. 104)
The consequences of uniformity, standardization, and implementing assessment practices that deliver low levels of validity are evident in the environment created by the test-based reforms at CUNY. The historical maneuver to link the assessment of literacy skills of second language users, minorities, and the underprivileged to social, economic, and educational opportunities is also in evidence. In his article “Tilting at Windmills: The City of New York’s ACT Writing Exam,” Bruce Chadwick, an adjunct professor at CUNY at Kingsborough (a community college), investigates this environment more closely to understand the way writing assessment practices serve to reinscribe socioeconomic relationships. Chadwick situates his essay within the socio-historical context of CUNY as an open-admissions school serving an evolving urban population.

The historical proximity between the open admissions of CUNY in 1970 and the passage of the Voting Rights Act of 1965 provides a vivid backdrop to the standardized writing examination introduced at CUNY colleges in “the middle of the Fall 1999 semester” (Chadwick, p. 4). The use of the “literacy test” prior to the passage of the Voting Rights Act was a means through which white oppressors were able to deny non-whites the ability to participate in the democratic process. The eerie similarities between the use of the literacy test and the use of ACT writing exam at CUNY at Kingsborough reveal not only what Chadwick refers to as the “pernicious effect on the students and faculty at Kingsborough” (1), but also the insidious manner by which external forms of assessment are capable of controlling curriculum, pedagogical practice, and the manner in which students view writing and their selves as writers. The writing program of CUNY at Kingsborough designed their program around an “ideological model” of
literacy but these efforts were undermined by policies that forced a pedagogy informed by an “autonomous model” of literacy upon students and teachers through the implementation of an external form of assessment.

Chadwick presents the “pernicious effect” of the ACT writing exam by describing both the “pedagogical implications” as well as the “pedagogical consequences” that followed. Chadwick’s description of his English Language Learners (ELL) course is not surprising after reading the task force’s report. He tells us that his students have “arrived here either directly from their country of origin, or are second-generation children of parents who arrived here in Brooklyn to begin a new life” (p. 2). He goes on to inform his audience that “their command of spoken and written English varies, but for the most part they have great difficulty writing English because of the unfamiliarity with English syntax, grammar, vocabulary” (p. 2). As noted in the analysis of the task force’s report, the ACT writing exam is both an entrance and an exit exam. The students in Chadwick’s course have all passed their ESL (English as a second language) classes and have entered regular non-ESL remedial classes which they completed and passed. But they cannot move on to a regular, non-remedial freshman English course until they pass the CUNY-ACT. All remedial students, whether ESL or not, must take and pass Kingsborough’s remedial courses before they can take English W which is specifically for students who have passed everything but the CUNY-ACT test. (p. 2)
Thus, the English W course at CUNY is based on the concept of “teaching to the test,” and Chadwick then informs his audience that the failure rate among English W students is consistently between 40-50% (p. 2). This figure in the context of the task force’s examination of funding problems, i.e. students run out of funding because they spend too much time in remediation, paint a dire picture for the chances of many of these students to be successful in college.

If the high failure rate, and with it the added effect that many students dropout because they lose funding, isn’t bad enough, the pedagogical implications make matters worse. Chadwick writes, “before the arrival of CUNY-ACT test, the Kingsborough English department had implemented remedial courses based on portfolio assessment. The courses were designed to emphasize revision” (p. 3). The assessment practices designed and implemented by the writing program are examples of what Huot (2002) calls “site-based, locally-driven procedures for evaluating student writing have their roots in the methods and beliefs held by the teachers who teach the courses that students are entering or exiting” (p. 104), and adhere to the theoretical principles that view writing as a complex social activity. The impact of the ACT-writing exam on this theoretically informed localized assessment practice was, of course, undermined with the arrival of an external assessment. The CUNY-ACT, as Chadwick explains:

replaced the portfolios as the sole exit requirement for remedial students to pass into the first semester freshman composition course…it immediately cut short the painstaking work students did in preparing the portfolios, and, in fact, rendered them meaningless…the massive change in exit
policy served to erode the confidence of the students and the students’
trust in their instructors. (p. 4)

The ability of external assessments to undermine the theoretically informed practices and
procedures of a writing program is at the heart of this investigation into the use of testing
to reproduce the means of production. The writing program at CUNY at Kingsborough
had designed a writing sequence that allowed students to develop as writers according to
the most contemporary scholarship on writing and the teaching of writing. The focus on
process over product stressed the reality of writing to students: that writers plan, revise,
seek advice and peer review, research, and develop their skills slowly over time. In one
fell swoop, the task force was able to eradicate the work of experts (writing scholars and
instructors) and replace the efforts of their careers and daily work with a test that finds its
origins in oppression, racism, and classicism. Chadwick informs his audience that
“because of the test’s high stakes, preparation for it prevailed over preparation for the
portfolios despite the pleas of well-meaning Kingsborough English department
administrators urging instructors not to ‘teach to the test’” (p. 6). The triumph of the test
over the assessment practices that were designed based on a socio-cognitive model of
literacy have far reaching effects.

The most obvious, as I have stated, are that many students will simply never pass
the test and will lose funding and will proceed no further—they will not pass go, and they
will not collect $200. Others will, at least 50% will pass the course and enter credit
earning courses without the kind of writing instruction that the entire remedial apparatus
was designed to provide. The time and money spent on “teaching to the test” leaves no
one better off save the large testing companies who have been contracted to administer the test.

On a pedagogical level the move toward the ACT-writing exam is clearly a source of frustration for both the students and teachers. Whereas the portfolio system designed and implemented at CUNY at Kingsborough was informed by a widely accepted theoretical understanding of language and writing, the ACT-writing exam ignores the composition scholarship of the last fifty years. Stephen Witte’s and Roger Cherry’s “Direct Assessment of Writing: Substance and Romance” (1998) challenges several assumptions that many traditional writing assessments make, based upon outmoded notions of validity in particular and in a more general sense of measurement as a positivist science. One assumption, that that terms “writing prompt and writing task refer to the same phenomenon” (p. 72), speaks directly to design of the ACT writing exam. Witte and Cherry argue that “it is common for such research and assessment studies to assume that the same prompt given to different groups results in performances of the same task by writers in those different groups” (p. 72). The authors’ research leads them to believe that “it is problematic to assume that writers who are given the same prompt necessarily engage in the same writing task” (p. 73). Roberta Camp’s (1993) statement about the evolution of writing assessment: “we now see that this meaning-making activity occurs over time and involves processes that are recursive, that are used differently by different writers, and that vary with knowledge of the topic, the context for writing, and the personal and cultural history” (p. 102) is a widely shared and accepted understanding of the writing process. This process, however, is ignored by the very precepts of a
standardized, timed-writing examination. Timed-writing examinations attempt to create a context through the design of the prompt. Witte and Cherry argue that appearances to the contrary, the ‘contexts’ specified in writing assessment prompts are both ahistorical and acontextual. Rather than providing information that facilitates test production, ‘contextual variables’ named in writing assessment prompts actually deprive writers of the opportunity to shape contexts of production and use. (p. 76)

The writing instruction at CUNY, based initially on a socio-cognitive model of writing prevalent throughout various fields that address literacy education, has been hijacked by policy-makers through the implementation of a high-stakes test that forces students to prepare for the test and for teachers to help prepare them. The focus is no longer on writing, but on the task of taking an exam, which is not exactly the focus of the test – the ability to write. For Witte and Cherry one of the more vexing problems with this pseudo context is that it camouflages the variables that are operable in the assessment agenda itself (e.g. time constraints that impose limitations on composing processes, unknown readers who read principally to evaluate, writers directed to assume particular roles)…however, the variables that are at work in the assessment itself are wither hidden beneath or made to appear subordinate to ‘contextual variables’ specified in the writing prompt. (p. 76)

This camouflaging of the assessment agenda is at the heart of the reform movement undertaken by New York City policy-makers, and it is at the heart of the ACT writing
examination. As an instructional tool ACT writing exam is utterly useless. Chadwick informs us that

when the test are completed, they are collected and shipped off to another CUNY college where they are all read within a few days’ time…the test are then stored away into a vault. They are not sent back to the college from whence they came…If a student fails the exam, s/he can appeal…but cannot see the exam…I (the teacher) can work with the students in my course and try and help them, but I don’t know what they did wrong on the test because the results are simply noted by number. I do not have access to the tests that the students have failed. (pp. 9-10)

Although assessment in general and writing assessment in particular can have important, positive effects for the teaching and learning of writing, there has to be a certain set of conditions.

an agenda for assessment that recognizes it as an important element for social action allows us the ability to guard against over-privileging the values, gestures and customs of certain groups and provides assessment with the potential to become an agent for progressive social change that includes and highlights the improvement of educational environments and opportunities for all students. (Huot, 2002, p. 9)

The writing assessment implemented at CUNY for entrance, placement, and exit purposes have done just that, privileged a certain group and arrested progressive social change. Due to what Cherry and Witte see as a “perceived need to render writing—
which is fundamentally messy and complex in its production and use—neat, clean and simple in order to satisfy measurement experts’ notions of reliability and validity” (p. 80), the policy-makers of the mayor’s task force have privileged a certain way of knowing and writing by implementing the ACT writing exam. A way of knowing and writing that is represented by educational measurement’s positivist principles. Witte and Cherry close with alternative assessment procedures that have a possibility of more closing realizing the “messy and complex” nature of writing:

such procedures might involve having writers assemble portfolios of documents produced in such naturally occurring contexts and having them explain the history and rhetoric of each document, or such procedures might involve having writers, in more controlled settings, construct their own writing prompts by using variables they had derived from their own communication experiences. (p. 80)

With the irony befitting such a tragedy, the alternative assessment procedures that Witte and Cherry describe resemble the pre-ACT portfolio system that Chadwick and his colleagues had designed for assessing their remedial writing students. This system, of course, was made obsolete by the implementation of the ACT writing exam because passing the CUNY writing assessment became more important than learning how to write.

The ACT writing exam continues to be a source of frustration for both students and teachers at CUNY at Kingsborough, but perhaps the most frustrating incident in the entire CUNY reform is found in the reception of a certain study’s findings. Barbara
Gleason undertook a three year study of the effects of mainstreaming remedial writers into credit earning courses, regardless of their ability to pass a specific writing assessment. Barbara Gleason attempted to provide the task force with her data and was promptly ignored. When the task force argues that there was no institutional data on remedial outcomes they are making a startling omission. The task force published their report on June 7, 1999. The findings of Barbara Gleason’s (2000) study were published in CCC’s June of 2000. The abstract for the CCC’s article informs the reader that “this project’s complex evaluation report was virtually ignored by college administrators” (p. 560), and the Friends of CUNY might have had her findings in mind when they stated that the task force “confirmed the premature judgments put forward by the Mayor in his initial charge and omitted and ignored all countervailing data” (Arenson, Rebutting, 1).

Gleason’s pilot study sampled thirty-seven sections. Gleason writes that “to theorize the curriculum we drew heavily on Shirley Bryce Heath’s study of literacy” (p. 562) and other scholars’ work in sociolinguistics and curriculum development design. The pilot study developed and implemented a rigorous evaluative process for student writing, curriculum design, and program evaluation:

(1) formative evaluation of students’ and teachers’ experiences in the writing course, which included an “expert judgment” evaluation at the end of the project’s first year; (2) a statistical analysis of student progress and achievement both in writing courses and other undergraduate courses; (3) an expert judgment report of the project at the end of its third and final year (summative evaluation); and (4) an
evaluation of the writing of twenty-two randomly selected students and self-assessment of their learning. (p. 563)

The rigor and authenticity of this study was inexplicably ignored by the task force or other policy makers in charge of important decisions about the ways in which to make under prepared college students as successful as possible. The task force claims:

CUNY has made little effort to determine which approaches work well or badly for particular student populations. Neither we nor CUNY knows whether and how many remediation students are in fact mastering basic academic skill sufficient for college readiness. Moreover, there has been little analysis to determine which of CUNY’s various institutions and programs are best suited to provide which types of remediation…the information that does exist tends to be anecdotal or unreliable. (pp. 31-32)

The task force’s refusal to take the findings of Gleason’s study into account also highlights the tendency of policy-makers to dismiss the scholarship of compositionists as illegitimate or unreliable.

What Gleason’s study found was that “when we compared the grades of remedial-placed students with the grades of students placed directly into English 110, we discovered that our remedial-placed students were passing the core courses at a rate that was even higher than the rate for our pilot course student who had placed into English 110” (p. 568). And, although the task force argued that the success of remediation programs at CUNY rested on anecdotal evidence, they also argued that “CUNY has no systematic, objective way to measure student progress through remediation, or to
determine what students have actually accomplished in remediation‖ (p. 26). Gleason’s study, again, a three-year study that looked at thirty-seven sections, provides not only hard, “objective” evidence of students labeled remedial writers, but also provides hard data that suggests remedial writing courses may do more harm than good. At the very least, CUNY officials and task force members should have clamored for a large-scale replication of Gleason’s study. As Gleason notes, “organizing and reviewing up-to-date transcript files offer a powerful alternative to the sorts of anecdotal evidence that too often influence writing program policy decisions‖ (p. 568).

The fact that this study was ignored raises many questions and casts doubts as to the intentions of the reformers and their reforms at CUNY. The use of remediation to “weed out” students who are deemed unprepared reveals the institutional cunning that represents an amended conception of “open admissions” and remedial education as “serving the people and the City” when in fact these reforms reinsure that the existing socioeconomic relationships will be reproduced. The refusal to take Gleason’s study into account during the task force’s deliberations in conjunction with the mandated use of the ACT writing exam only serves to intensify the feeling that CUNY is reforming its school system by removing many of the people it was created to serve.

Perhaps the most frustrating aspect of the reform movement at CUNY is that Gleason provides very real evidence that argues that remediation need not be repaired or overhauled but removed all together; that the CUNY system need not have implemented a wide range of test-based reform, but, in its stead simply phased out remediation and allowed students to be judged according to their performance rather than their potential as
defined by a test. Instead, the test-based reforms of the task force created situations like we have seen at Kingsborough College. In the guise of rescuing students from a failed system, the reforms have further hampered the majority of students within the CUNY system in realizing their educational aspirations.

Conclusion

The importance of this investigation into the test-based writing assessment reforms at the CUNY college system can be located in the ability of non-experts (policy-makers and politicians) to implement policies and programs that undercut and undermine the pedagogies and policies designed by experts (writing scholars, writing instructors, and writing program administrators). The ability, power, and proclivity of policy-makers and politicians to produce environments such as those seen at Kingsborough community college reveal the institutional demand inherent in corporate, for-profit structures to reproduce the means of production.

The portfolio system designed at CUNY at Kingsborough community college had its foundation in the most widely accepted theoretical explanations for how people learn to write, how people can best teach writing, and how teachers, students, and writing programs can best assess writing. That the same unexamined and deeply flawed views of literate practices represented by the “autonomous” model of literacy, and the deeply flawed educational measurement techniques of standardized writing assessment practices can merge to undermine the efforts of teachers and writing program administrators in the name of efficiency, standards, and accountability are discouraging to say the least. The
latest wave of test-based reform, however, has been met with much resistance, and signs that it is receding are starting to appear.

Fairtest, the National Center for Fair and Open Testing, is an example of a coalition of different organizations that have, through their unity, been able to make a difference. Fairtest’s states mission “advances quality education and equal opportunity by promoting fair, open, valid and educationally beneficial evaluations of students, teachers and schools. FairTest also works to end the misuses and flaws of testing practices that impede those goals” (Fairtest.org). The coalition beneath the banner of Fairtest in an attempt to overhaul NCLB included, as of April, 2008, one hundred and forty-four associations and organizations from the American Association of School Administrators to the Young Men’s Empowerment Movement. As part of Fairtest’s “emphasis on eliminating the racial, class, gender, and cultural barriers to equal opportunity posed by standardized tests, and preventing their damage to the quality of education,” this organization provides organizing tools to aid individuals and communities to take action in their local areas, research information, links to publish data and scholarship on testing and the misuse of testing, and, most importantly, a unified voice for change.

Fairtest can serve as a model for writing programs nationwide. A coalition of the writing programs might make it possible that cases like CUNY at Kingsborough become horror stories of the past rather than examples of out-dated models of literacy and writing being perpetuated at the end of the first decade of the twenty-first century. The subversion of the scholarship produced within composition studies over the past fifty
years devalues both the field and its research, but, more importantly, leaves the students in our charge at the hands of corporate powers that view productivity, efficiency, and profit the as the primary purpose of public education. This corporate ideology, when adopted for the purposes of running an educational institution, inevitably reproduces the means of production by ignoring students who need our attention most. We have seen this reproductive technology at work through the use of culturally biased writing assessment practices designed on the tenets of traditional test theory, and the hue and cry of test-based reforms can still be heard through the words “accountability, higher standards, and measurement.”
Chapter Five:
Towards an Understanding of Critical Validity Inquiry

“If the misery of our poor be caused not by the laws of nature, but by our institutions, great is our sin.”

Charles Darwin

The investigations into the use of writing assessment for the purpose of reinscribing socioeconomic relationships has approached the duplicity embedded in the practices of corporatized institutions. The history of assessment, as reviewed in chapter two, is closely linked to social and intellectual movements, such as eugenics, that adhere to notions of racial superiority and innate intelligence that argued for their beliefs through a pseudo-science of measurement. This historical link to oppressive assessment also operates in contemporary educational settings through the institutional use of writing assessments. As we have seen, educational assessments in the form of standardized writing exams are well suited to meet the demands of reproduction through their ability to define the individual, to rank and organize members of a society, and to over privilege a particular group. This dissertation began by asking:

1. Why are standardized writing exams flourishing in the face of research that shows decisions made from such assessments to be invalid or that achieve low levels of validity?
2. How do specific cases of large scale writing assessments work to reproduce current socio-economic power relations?

3. How can a métis-based institutional critique provide researchers and teachers with an avenue for influencing policy issues regarding assessment, placement, and admissions?

In an attempt to answer these questions it has become clear that the opportunity to expand validity inquiry for future investigations into the use misuse of educational assessments needs to be taken.

The theorization of a unified concept of validity articulated first by Cronbach (1971) and later, to wider acceptance, by Messick (1989) and finally incorporated into the 1999 Standards adopted by the American research Association, the American Psychological Association and the National council on Measurement in Education expanded the requirements for validating an assessment practice and defined validity as a degree and validation as a process. Messick’s declaration that “validity always refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores” (p. 13) set a new standard for the use of test scores to make decisions and to draw inferences. The added concern for “the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes but the unintended side effects” (p. 16) placed a new burden on educational assessment. This new burden required validators to inquire into the validity of the use of test scores by examining the individual and social
consequences that were a result of inferences drawn and decisions made based on test scores.

It is somewhat ironic that the proponents of test based reform infuse their calls to action through a mantra of “higher standards,” “measurement through testing,” and “accountability,” when in fact, the people who need to be held most “accountable”, by their own standards, are the people employing these standardized assessments. The latest Standards for educational and psychological testing (1999), under Standard 13.1, states:

When educational testing programs are mandated by school, district, state, or other authorities, the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user. (p. 145)

When we inquire into the use of testing at schools like CUNY at Kingsborough it appears that the very guidelines that these educational assessors have set for their own practice are being violated.

We saw the language of test-based reform in the mayor’s task force report, and this same language is evident in President Bush’s speech to reauthorize No Child Left Behind where he uses “measure” ten times, “accountability” fifteen times and “standards” twenty-one times. As with the CUNY task force report, the real cunning or métis that is at the heart of this rhetoric can be found when validity inquiry is undertaken.
It appears that it might be appropriate for validity inquiry to apply a corollary to métis for its own research into the intended and unintended consequences of test use.

The role of métis in this investigation has been a rhetorical tool that provides a particular perspective from which to approach the rhetoric, policy, scholarship, and rationales put forth by institutional administrators, policy-makers, and scholars acting at the behest of such institutional participants. Métis contextualizes the actions, ideas, and rhetoric of institutional participants within the duplicity that Gee et al. identify as an integral element of fast capitalism and its institutions. At the heart of this duplicity is the institutional paradox that creates a conflict between the needs and aspirations of the for-profit institutions and the needs and aspirations of the individuals who make the institution possible.

Gee et al. illustrate this duplicity when they discuss the notion of “Six Sigma”. Six Sigma, Gee et al. inform their audience, “means ‘six standard deviations from a statistical performance average’” (p. xiii). The underlying goal of Six Sigma in a corporate setting is also known as quality control. The duplicity inherent in fast capitalism, as Gee et al. illustrate, can be found in the “dual coding” necessary for corporate structure’s to communicate on two very different levels. Gee et al. write, “As a formal tool the sigma talk is meant to set the goals of the organization as a whole and all its parts by giving a measure of productivity, where productivity is now defined as ‘total quality’” (p. xv). However, when this same language is “applied to individuals, and not at the level of the organization and its goals, it needs to be retranslated” (p. xv). It is within the process of translation that the cunning of institutions built on the corporate
model can be located. In the example of Six Sigma, Gee et al. argue that “as a tool it is no fiction at all, but as a sociotechnical device it needs to be made into more ‘human’ interaction commensurate with quality culture; presented ‘as such’ it would be insulting” (xv). The reality behind this “dual coding” is that the executives, policy-makers, and leaders of institutions that adopt the for-profit model, be they CEOs, presidents of universities, or factory owners, guarantee no place within their institution for their employees when the job is completed. This notion of a portfolio career is examined by Gee et al (1996):

> New capitalist businesses want such ‘deep’ learning, with it concomitant Identity and value formations. These businesses are well aware too, of the conflict between traditional “American individualism’ and the ‘teamwork’ of the new work order, or the conflict between the traditional ‘job’ and the ‘portfolio career’ of new capitalism. (p. 15)

As Gee et al. put it, the very language that encapsulates the ends, goals, and vision of the organization and its culture must be translated into language of social intervention or manipulation (depending on the view you take of the matter) so as not to insult the workers/partners and so as to motivate them” (p. xv). The perspective of a métis-based institutional critique searches for this duplicity, and, by locating this duplicitous language/behavior, is capable of revealing the misuse of power. In the case of CUNY, the implementation of test-based reforms is couched in language that seeks to replace a failed system of open admissions and remediation with a meritocracy. In reality, as we have seen through a métis-based critique, the reforms are meant to reorganize the schools
of the CUNY system along lines drawn by bias standardized tests. This reorganization will provide a new veneer of “excellence” to the elite schools within this new system.

Chapter three of this dissertation discusses the need for a unified front for the purpose of remedying the misuse of standardized testing:

Any hopes of replacing the positivist tradition of psychometric, standardized assessment practices that reduce complex human performances to numeric representations rests on uniting the social, political, and professional authority of educational measurement with the socio-cognitive approach to teaching and learning. Until this united front is established, the powerful industry in control of the technology of assessment will continue to reproduce socio-economic relations through the educational apparatus. These forms of assessment will be designed using the basic tenets of traditional test theory, ensuring that the racial, ethnic, and gender biases from which this tradition originated will be perpetuated ad infinitum. (p. 46)

To this end, the purpose of this chapter is to expand the concept of Critical Validity Inquiry (CVI) as a process that seeks to understand the misuse of power through educational assessment practices. By merging rhetorical and critical theory within the argument one can make for the use of educational measurement, CVI can provide an important lens for looking at the ways in which educational evaluation like writing assessment can be used to distribute opportunities, withhold access or other use or abuse of the power inherent in educational decision-making. The two cases that have been the
central focus of this métis-based critique, directed self-placement (DSP) and the City University of New York’s ACT writing exam, provide a rationale and the grounds for expanding the concept of CVI.

*Putting the Critical in Critical Validity Inquiry*

The expansion of a methodology into a critical methodology has several precedents: the most notable being Fairclough’s Critical Discourse Analysis (CDA). Rebecca Rogers describes CDA as “both a theory and a method. Researchers who are interested in the relationship between language and society use CDA to help them describe, interpret, and explain such relationships” (p. 2). What sets CDA apart, what puts the critical in critical discourse analysis, is the fact that CDA “includes not only a description and interpretation of discourse, but it also offers an explanation of why and how discourse works” (p. 2). CDA has at its foundation eight core principles:

- CDA addresses social problems
- Power relations are discursive
- Discourse constitutes society and culture
- Discourse does ideological work
- Discourse is historical
- A sociocognitive approach is needed to understand how relations between texts and society are mediated
- Discourse analysis is interpretive and explanatory and uses a systematic methodology
CDA is a socially committed scientific paradigm

The purpose and intentions of CDA reflect many of the ideas, concepts and work done through educational assessment. Recent unified notions of validity could certainly support and benefit from Critical Validity Inquiry (CVI).

CVI for the purpose of writing assessment is interdisciplinary in that it seeks to unite both the field of Educational Measurement’s concern for validity as defined by Messick with Rhetoric’s interpretation and explanation of language in context, and Composition’s concern with a sociocognitive approach to literacy and pedagogy. Just as the unitary concept of validity argues for the importance of guarding against the misuse of tests and the unintended consequences of decisions made based on test scores, so too does the field of writing assessment understand the need to guard against the exploitation of literacy. The social responsibility of both fields can be met by a critical approach to inquiry:

an agenda for assessment that recognizes it as an important element for social action allows us the ability to guard against over-privileging the values, gestures and customs of certain groups and provides assessment with the potential to become an agent for progressive social change that includes and highlights the improvement of educational environments and opportunities for all students. (Huot, 2002, p.9)

The case of the ACT writing exam at CUNY at Kingsborough investigated in chapter four provides a worthy example for a discussion of the foundational tenets of CVI. The rhetoric presented by the mayor’s task force’s report can appear as merely addressing
what are basically laudable concerns. It is hard to argue against the idea that students and
teachers should set high standards or that individuals should be held accountable in their
respective roles as teachers and students. The case at CUNY, however, can also be seen
as part of a move to restructure the public nature of education in America. Michael
Apple argues in *Official Knowledge* (1993) that a public relations campaign to cast public
education as a failed enterprise has been underway for thirty years and finds at its roots
the neoconservative movement. He argues that

One of the major causes of educational failure is seen to be the supposedly
nearly complete control over policy and practice by teachers and other
‗professional educators.’ Such autonomy “has left the educational system
both unaccountable to consumers (parents and communities) and open to
precisely the sort of curriculum initiatives associated with ‘liberal ideas.’”
(p. 3)

Apple also discusses the fact that much of this political unrest has been built on the
notion that teachers cannot be trusted for either their “quality” of teaching or
“commitment to education”. This language is the key. Who defines the quality of or
commitment to teaching, and what are the parameters for this judgment? The reliance on
non-referential qualitative measures has been illustrated by Bill Readings’ explanation of
“excellence” in the contemporary university. Michael Apple outlines the shifts in
educational policy achieved by the neoconservative movement:

1) proposals for voucher and choice plans and tax credits to make schools
more like the idealized free-market economy; 2) the movement in the
legislatures and state departments of education to “raise standards” and mandate both teacher and student “competencies” and basic curricular goals and knowledge, thereby centralizing even more at a state level the control of teaching curricula; 3) the increasingly effective assaults on the school curriculum for its supposedly antifamily and anti-free enterprise bias, its “secular humanism,” its lack of patriotism, and its neglect of the “Western tradition”; 4) the growing pressure to make the needs of business and industry into the primary goals of the educational system. (p. 20)

The neoconservative movements’ perceived failure of the public schools has been presented to the general public in various ways. The most notable failures, as we have seen several times throughout this dissertation, have come in a series of literacy crises. These failures are invariably accompanied by calls for reform, and, as in the CUNY case, these reforms are usually test-based reforms. The mistrust of “teachers and other ‘professional educators’” that Apple describes is more often than not rectified by external forms of assessment.

CVI for the purpose of writing assessment merges the sociocognitive model of learning with Cronbach’s notion of validity as a never ending process. Schools should be held accountable, but the methods of assessing this accountability need to be commensurate with contemporary theories of literacy and learning. The move towards testing “writing ability” through standardized, impromptu, timed-writing examinations is theoretically informed by a notion of literacy that is no longer accepted by the majority of
literacy scholars. Perhaps even more important is the research that shows these standardized writing assessments tend to privilege a certain kind of literacy and a certain kind of community. Finally, it has been shown that external forms of assessment effect classroom curriculum and teacher pedagogy and result in a skills based education that stresses drilling and rote learning. These non-recursive strategies for teaching have proven to be unsuccessful when used to educate a diverse population.

The pedagogy of drilling and rote learning doesn’t demand a teacher reflect on her pedagogical practices. However, a more nuanced understanding of teaching, one informed by a sociocognitive model of language and learning does. Brookfield’s expansion of the practice of reflective teaching, Critical Reflective Teaching (CRT), is another precedent for critical practice. Brookfield premises his notion of CRT with the belief that “we teach to change the world” (p. 1), and that “an uncritical stance toward our practice sets us up for a lifetime of frustration” (p. 1). The idea of a reflective practice finds its roots, arguably perhaps, in Schon’s *The Reflective Practitioner* but has been absorbed by many teachers and scholars interested in pedagogy. CRT is founded on the notion that we all make assumptions about the way the world works. As Brookfield says, “we are our assumptions” (p. 2). He categorizes three types of assumptions. The first is paradigmatic assumptions:

they are the basic structuring axioms we use to order the world into fundamental concepts…paradigmatic assumptions are examined critically only after a great deal of resistance to doing so, and it takes a considerable amount of contrary evidence and disconfirming experiences to change
them. But when they are challenged and changed, the consequences for our lives are explosive. (p. 3)

The second, prescriptive assumptions “surface as we examine how we think teachers should behave, what good educational processes should look like, and what obligations students and teachers owe to each other” (p. 3). The third, causal assumptions “help us understand how different parts of the world work and the conditions under which processes can be changed” (p. 3). For Brookfield “reflection becomes critical when it has two distinctive purposes. The first is to understand how considerations of power undergird, frame, and distort educational processes and interactions, and the second is to question assumptions and practices that seem to make our teaching lives easier but actually work against our own best long-term interests” (p. 8). Again, the focus on power and the manner in which it can “distort educational processes and interactions” is apropos of the situation of test-based reform at CUNY and DSP at Grand Valley State. While CVI might be used to uncover the ways in which CUNY officials use testing to socially engineer the student population at its more elite schools, CVI can help to illustrate that the decisions DSP allows students to make do not work to empower or otherwise provide students at Grand Valley State with increased agency or control over their educations, though DSP might certainly promise such outcomes.

The approach of CVI is concerned with the “individual and social consequences of decisions made based on test scores” and inquires into the presence of power and the misuse of power. Foucault’s concept of power as explained in *Discipline and Punish* is helpful when focusing on the relationship between power and educational assessment.
Foucault argues that “Discipline ‘makes’ individuals; it is the specific technique of a power that regards individuals both as objects and as instruments of its exercise” (p. 170). For Foucault, the examination is capable of both objectifying the individual, an idea that is echoed by Alan Hanson, as well as organizing individuals into a hierarchy of observation. The use of tests scores to rank students and to organize society in turn creates this Foucaultian “hierarchy of observation” that makes “an apparatus in which the techniques that make it possible to see induce effects of power, and in which, conversely, the means of coercion make those on whom they are applied clearly visible” (pp. 170-171). Foucault’s conception of power as a “network of gazes” (p. 171) aptly captures the power inherent in large scale testing and typified by test-based reforms. When Foucault refers to the examination as “a surveillance that makes it possible to qualify, to classify and to punish” (p. 184) he reveals the true power of assessment. When he adds that “It establishes over individuals a visibility through which one differentiates them and judges them” (p. 184) he cuts to the institutional use of assessment as a means of surveillance and dominance as opposed to a notion of assessment that seeks progressive social action.

CVI for writing assessment adopts a sociocognitive model precisely because of writing assessments history as a means to maintain dominance. As we have seen through the “literacy test” of the south prior to the Voting Rights Act of 1965, and as both the case of CUNY and DSP at Grand Valley State University illustrate in a contemporary setting, writing assessment is “power”. George Hillocks (2002), writing in the Testing Trap: How State Writing Assessment Control Learning, addresses the manner in which
standardized writing assessments interfere with the teaching and learning of writing because of the narrow view that these assessments are informed by. He writes:

   in the past 30 years, researchers and theorists have come to know that teaching writing entails teaching thinking. Further, they would argue that people learn through writing…writing is more than a mnemonic device. Its presence makes it an aid in thinking though highly complex processes and problems. But if writing is an aid to thinking, it is also obvious that thinking is requisite for most writing. The serious teaching of writing and thinking must go hand in hand. (p. 6)

Hillocks’ understanding of the relationship between writing and thinking underscore the negative effects that standardization in education represents. When we think about the rote learning and drilling that accompanies standardized tests, this relationship between thinking and writing illustrates the uneven playing field of educational opportunities. For Heath’s Trackton children, the wave of test-based reform spells trouble, and, as Nichols and Berliner explain in *Collateral Damage: How High-Stakes Testing Corrupts America’s Schools* (2007):

   Tests were chosen as the mechanism to measure productivity. Like those in the business community, legislators believed that productivity could be increased without more money needing to be spent simply by holding schools and educators accountable through the practice of high-stakes testing. Lazy teachers and their students would be detected and made to work harder. (p. 19)
The power that this investigation into the reproduction of the means of production locates can be considered “corporate power,” and Nichols’ and Berliner’s connection of high-stakes testing to the business community makes the corporatization of the university as described by Bill Readings a logical step forward in the test-based reform movement. This corporatization is evident in many contemporary institutions of higher-education and the modeling of these institutions on a corporate model of efficiency and productivity creates new conditions that need to be reproduced. The cases of both CUNY and directed self-placement illustrate how even the best intentions of scholars and administrators can be undermined by institutional pressures.

To this end, CVI is concerned with the manner in which assessment practices, intentionally or not, “distort educational processes and interactions.” To investigate the misuse of power through assessment, CVI is founded on tenets that blend institutional critique, validity theory, and critical theories of reproduction.

- Validity is a unified though faceted concept and that validation is scientific inquiry (Messick, 1989, p. 14).
- Tests do not have reliabilities and validities, only test responses do. Test responses are a function not only of the items, tasks, or stimulus conditions but of the persons responding and the context of measurement (p. 15).
- We can trace the social consequences of interpreting the test scores in particular ways, scrutinizing not only the intended outcomes but also unintended side effects (p.16).
- Determining what a test is measuring always requires recourse to other forms of evidence (p. 17).

- We will underscore the continuing need for validation practice to address the realities of testing consequences, including the often subtle and systemic effects of recurrent or regularized testing on institutional or societal functioning (p. 18).

- The advice is not merely to be on the lookout for cases your hypothesis does not fit. The advice is to find, either in the relevant community of concerned persons or in your own devilish imagination, an alternative explanation of the accumulated findings; then to devise a study where the alternatives lead to disparate predictions. (Cronbach, 1989, p. 14)

- We aim to change the practices of institutional representatives and to improve the conditions of those affected by and served by institutions (Porter et al., 2000, p. 611).

- Institutions are both material and rhetorical spaces, and our definition of them must encompass these elements (p. 625).

- The ideological apparatus which has been installed in the dominant position in mature capitalist social formations as a result of a violent political and ideological class struggle against the old dominant ideological State apparatus is the educational ideological apparatus (Althusser, 2001, p. 103).
The principles of CVI represent a view that educational institutions have a responsibility to protect society and its individual members from the misuse of this powerful technology of educational assessment. Within validity theory the role of validity inquiry “to address the realities of testing consequences, including the often subtle and systemic effects of recurrent or regularized testing on institutional or societal functioning” (p. 18) is expanded in its capabilities by CVI’s focus on “power.” By keying on traditional sites of educational exploitation like race, class, and gender, CVI is capable of recognizing abuses of power that might be missed in a more generally focused inquiry. This misuse of educational assessments that fail to represent an equitable process for all and, instead, serve the purpose of reproducing the social relations necessary for the perpetuation of a vast disparity in the distribution of wealth, health services, and educational opportunities will be challenged.

Critical Validity Inquiry and Writing Assessment

The low degree of validity which standardized writing assessments of both direct and indirect writing deliver have been repeatedly reported in the scholarship of writing assessment and educational measurement (Moss, 1994, 1998, 1996, 1992; Bracey, 2000, 2001; Haswell, 2001). The most difficult aspect of measuring “writing ability” as a construct is identified in the work of many scholars but no where more acutely than in Cherry’s and Witte’s “direct assessment”. As we have seen in both chapter one and chapter four, Witte and Cherry identify the “messy and complex” nature of writing and outline the pseudo-contextual nature of writing prompts while at the same time
identifying the importance of real context, both personal and historical, that defines a writer’s approach to writing. This deeply embedded process of social communication cannot be replicated by a timed, single sample, impromptu writing examination.

More so than any other area of assessment, writing assessment, and composition as a field of study, has produced an abundance of scholarship that details the sociocognitive nature of writing and language (Applebee, 1986; Durst, 1987; Freedman and Pringle, 1981; Langer and Applebee, 1987; Odell, 1981; Penrose, 1989; Pringle and Freedman, 1985), and is, therefore, more prepared to provide evidence of standardization’s inability to produce equitable decisions based on scores. As George Hillocks notes, “If states are concerned about encouraging high-level thinking in schools, that interests is more likely to be reflected in writing assessments than in assessments for any other subject matter areas” (p. 17) because “the kinds of writing tested may influence the kinds of writing taught in schools” (p. 19). Hillocks argues, in the tradition of Street and Heath, that “we learn by assimilating new experiences into our already known world” (p. 21) rather than arguing for the traditional view of learning that believes “truth is objective and may be apprehended directly through observation of the world” (p. 21). Hillocks, Cherry and Witte, Huot, Williamson, Moss, and Camp are representatives of view of writing assessment that demands the proper conditions for organic writing.

The use of large scale writing assessments, because they seek to create the traditional test conditions that value reliability more than they do validity, are more often than not informed by an autonomous model of literacy. These large scale writing assessments believe that writing is merely a means of producing a product and are not
complex and not embedded in individual and social practices. In short, the demands of psychometrics directly contradict the sociocognitive model of literacy. As Roberta Camp (1993) points out:

We attempt to minimize or compensate for those factors most likely to create variation in writing performance from one occasion to another. We try to present prompts that are immediately accessible to all test takers—whatever their individual interests, knowledge, or cultural experience—and that can be dealt with in the limited time available for writing. (p. 50)

The notion that writing prompts create a real context or produce a writing task that is similar for all groups and individuals who are presented with this prompt has been disputed (Cherry and Witte; Hillocks; Nichols and Berliner). Camp notes that the compromises made in the name of reliability and testing conditions are problematic for measuring writing ability:

Ironically, many of the efforts we made to enhance the reliability of the writing sample—especially those aimed at streamlining and universalizing the writing experience—appear now to limit the value of the assessment. Very likely we are also seeing the signs of growing incompatibility between our views of writing and the constraints necessary to satisfy the requirements of traditional psychometrics—in particular, of reliability and validity narrowly defined. (Camp, 1993, p. 52)

Critical Validity Inquiry, in response to this contradiction, approaches all forms of writing assessment by keying its inquiry on areas that have, in a socio-historical sense,
been sites where literacy has been exploited for the purposes of reproducing the social relations necessary for production. CVI also investigates the ability of the writing assessment to achieve a context in which writers can approach the “writing task” in an authentic way so that the “complex nature of writing” and a writer’s ability can be represented with a higher degree of validity.

An outline for the process of CVI contains several overlapping points of interest that represent the assessment practice as a complex interaction between the writer, the assessment instrument, the testing conditions, and the use of test scores to draw inferences and to make decisions. These points of interest are then analyzed in terms of a sociocognitive model of literacy to understand whether the conditions of the assessment as well as the consequences of the decisions made based on the test scores of the assessment are fair and equitable.
Table 5.1: Critical Validity Inquiry

Chart 5.1 introduces the foci of CVI and, as we can see, looks at race, ethnicity, gender, class, administration of test, and test conditions to understand the impact of the assessment on the individuals and social groups being assessed. CVI seeks to understand how “power” is exerted by the test design, test conditions, and decisions made based on the test scores, and how this power affects the particular groups and individuals assessed. CVI, therefore, locates power by identifying the forces behind the assessment (i.e. politicians and policy-makers, writing programs, instructors and teachers, a combination of these parties) as well as the purpose for which the assessment is undertaken (admission, placement, entrance, exit). The two cases at the center of this investigation, directed-self placement at Grand Valley State University and the ACT writing exam at
CUNY, provide opportunities to look at the relationship between power and writing assessment through the process of critical validity inquiry.

Critical Validity Inquiry in Action

The placement procedure of directed self-placement (DSP) argues for the rights of students to play an integral role in their own educational decisions. No doubt, Royer and Gilles have nothing but the best intentions for the welfare of their students. That said, the institutional forces informing the larger decision-making processes behind the university have distorted directed self-placement and molded it to suit the conditions necessary for that institution. Thus, DSP removes the need for summer faculty to read placement essays, or for the admissions officers to reschedule their orientation sessions, and places the decision in the hands of the student (customer). The theoretical underpinnings of pragmatism are undermined by students’ past experiences with educational assessments. So, although Royer and Gilles see pragmatism as a philosophical system that supports DSP, they are unable to see the larger influences of students’ experiences with educational assessment. The intentions behind DSP are laudable, but the subversion of these intentions by institutional powers that stress productivity and efficiency over social responsibility goes unseen and unaccounted for. If the process of CVI were undertaken by Royer and Gilles, for example, as a means to inquire into the validity of DSP, CVI would reveal some of the critical assumptions that inform Royer’s and Gilles’ scholarship on DSP.

To reveal just a few of the assumptions CVI’s process can uncover let’s consider CVI’s focus on both the consideration of rival hypotheses as well as the critical
examination of “power” as it pertains to gender. One rival hypothesis that a critical approach to DSP might raise is whether or not students’ decision-making processes are being informed by past educational experiences with assessment, such as the ACT/SAT. Critical theories of reproduction in education (Foucault; Bourdieu) and the work of cultural anthropologist (Hanson; Gould) could inform proponents of DSP of the need to reconsider the use of the ACT/SAT scores as one of the “general characteristics” for self-assessment. The fact that test scores impact a person’s sense of self-worth is a cause for concern. The removal of this information from the brochure isn’t enough to counteract, I would argue, the influence of students’ past experiences with educational assessment. I would argue, however, that the identification of students’ past educational experiences with assessment as a characteristic taken into account when determining one’s placement decision should become a focus for continued research and validity inquiry into DSP’s connection to other educational assessments.

As part of Royer’s and Gilles’ presentation of DSP to the writing assessment community, they also presented anecdotal and statistical evidence that provides an account of the reasoning and rationale that informed one female’s placement decision (Kristen) and one male’s (Jacob). The anecdotal evidence of Kristen reveals that although she had earned a higher high school GPA (3.68) and was ranked in the top 12% of her class as compared to Jacob’s 3.16 GPA and a ranking of 46% in his high school class, Kristen placed herself in the ENG 098 and Jacob in the ENG 150. Royer and Gilles argue that “both Kristen’s caution and Jacob’s determination seem to us excellent reasons for selecting the courses they chose” (p. 66). An explanation that Royer and
Gilles fail either to identify or to address is the manner in which the confidence of the male student and the caution of the female student represent traditional roles that many argue get reinscribed through the educational apparatus. As Schendel and O’Neill (1999) point out:

these stories illustrate potential problems with directed self-placement, primarily that students’ past experiences that implicate their gender (or their socioeconomic status or ethnicity, information not included in the article)—may be at least as influential on their decision-making process as their answers to the questions they are directed by. (p. 220)

Applying CVI to the case of DSP is interesting because the power being exerted in students’ decision-making processes is subtle, and, as Royer’s and Gilles’ omission evidences, is enmeshed in social practices and assumptions that need to be critically engaged to be revealed.

CVI applied to the case of CUNY at Kingsborough is less about the unintended outcomes of an assessment practice and more about assessment as a site of contention. The chairman of the task force was at the time of the report and is directly linked to a privatization of public education movement represented by the Edison Project. The then Mayor of New York was a conservative politician who had voiced his disapproval of open admissions and remediation. The ideological tenets of conservatism and traditional test theory mesh nicely with the conservative agenda for education. Like a business, a school should be run on principles of “productivity,” “accountability,” and “high
standards‖ (i.e. quality control). The implementation of broad test-based reforms is the obvious response of an institution that values “efficiency” above all other concerns.

As we have seen, the student population at CUNY at Kingsborough is revealed through both institutional statistics (See Chart 5.2) as well as evidence presented by Bruce Chadwick in “Tilting at Windmills: The City of New York’s Act Writing Exam.” Chadwick (in press) describes his ACT prep class:

   twenty five students registered of this course. Eighteen are students whose first language is not English; only seven students are first-language English speakers and writers…the breakdown is as follows: seven are originally from China; one is from Israel; one from Bangladesh; one from Bosnia-Herzegovina; one from Albania; one from Albania; one from Russia; one from Puerto Rico; one from the Dominican Republic and one from Mexico. (p. 2)

The diverse nature of both Chadwick’s course and the CUNY at Kingsborough student body reveals both the pedagogical challenges confronting the teachers at CUNY as well as the educational challenges confronting the students.
Table 5.2

<table>
<thead>
<tr>
<th>Year</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian/Pl</th>
<th>Am Ind</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>48.2</td>
<td>30.0</td>
<td>12.6</td>
<td>9.1</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>1999</td>
<td>47.6</td>
<td>30.0</td>
<td>13.2</td>
<td>9.1</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2000</td>
<td>46.7</td>
<td>30.3</td>
<td>13.4</td>
<td>9.5</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2001</td>
<td>44.8</td>
<td>31.7</td>
<td>13.2</td>
<td>10.2</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2002</td>
<td>43.4</td>
<td>32.5</td>
<td>14.0</td>
<td>10.1</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2003</td>
<td>41.3</td>
<td>33.0</td>
<td>14.4</td>
<td>11.2</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2004</td>
<td>41.7</td>
<td>33.1</td>
<td>14.2</td>
<td>10.9</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2005</td>
<td>41.0</td>
<td>32.6</td>
<td>14.4</td>
<td>11.8</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2006</td>
<td>41.1</td>
<td>32.0</td>
<td>14.4</td>
<td>12.4</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>2007</td>
<td>40.6</td>
<td>32.3</td>
<td>14.6</td>
<td>12.5</td>
<td>0.1</td>
<td>100.0</td>
</tr>
</tbody>
</table>

To confront the pedagogical challenges of teaching writing to such a diverse group, the writing program, as we saw in chapter four, developed a portfolio assessment practice that stressed process over product. This pedagogy embraces a sociocognitive model of literacy. It introduces and informs students about the complex and recursive nature of writing, while, at the same time, providing students with the necessary means of revising (peer review, instructor feedback, and self assessment). The stress this pedagogic model places on revision allows second language learners to approach writing as something that happens over time, instead of within a limited period of time, and keeps the focus on writing as a process rather than writing as a product. As we saw in chapter four, the effect of the CUNY ACT writing exam on the portfolio system was devastating.
In comparison, the theory of literacy informing the ACT writing exam can be understood through the inquiry methods of CVI for writing assessment. First, a sociocognitive to literacy would expect that a diverse student population, with its lack of “command of spoken and written English,” (p. 2) to struggle with the time constraints. The autonomous model of literacy that informs the ACT writing exam, however, ignores these important factors (race, ethnicity, second language learners) and aggravates the situation through test design and test administration. We have seen the problems with a single, impromptu timed writing expressed in each chapter of this investigation.

Foremost, single impromptu timed writings fail to view writing as a complex social act embedded in individual contexts and experiences. Instead, the design of the ACT writing exam assumes an autonomous model of literacy that understands “writing ability” as stable, measurable construct. The added stipulation that prohibits students from using a dictionary places second language learners at a further disadvantage.

The external origins of the test, its use as an exit exam, and the high-stakes attached to success and failure (a failure rate of 50% for ESL students) locate the power of the ACT writing exam in its ability to deny a large number of students access to degree earning programs and credit earning courses. The ability of the test to organize and rank students, to offer or deny opportunity, and to reinscribe social relations is directly tied to the external nature of the examination. This outside influence “distorts educational interactions” by influencing curriculum and undermining the writing faculty.

A rival hypothesis for the 50% failure rate that Chadwick records at CUNY at Kingsborough could be that the test conditions (time restriction, impromptu, no
dictionaries, no real context) when applied to second language learners fail to produce a genuine situation or conditions for authentic writing. The failure to produce authentic representations of student writing undermines the construct of “writing ability” and invalidates the writing assessment practice represented by the ACT writing exam.

Most importantly, the use of the ACT writing exam, the inferences it draws, and the consequences and the decisions made based on the test scores it produces have such a negative impact on such a large portion of the assessed that the continued use is unwarranted. Judging the ACT writing exam against the backdrop of Cronbach’s demands on validity reveals the cost of these consequences to be too high in several regards. First, Cronbach (1989) argues that “because psychological and educational tests influence who gets what in society, fresh challenges follow shifts in social power or social philosophy” (p. 4). The social shift towards a conservative agenda’s demand for accountability and test-based reform explain the use of testing and are highlighted in the task force’s report, but the fact that the ACT writing exam continues to be used at CUNY illustrates that validators have failed. According to Cronbach, “the bottom line is that validators have an obligation to review whether a practice has appropriate consequences for individuals and institutions, and especially to guard against adverse consequences” (p. 6). The consequences at CUNY are very real, as we have seen, and the continued use of the ACT writing exam ensures that more people will suffer the consequences that come with failing the ACT writing exam. Since neither students nor teachers ever get to see the exam again, even if they appeal their score, Cronbach’s assertion that “the worth of instructional test lies in its contribution to the learning of students working up to the test,
or to next year’s quality of instruction,” (p. 5) goes ignored by the assessment policy-makers of CUNY.

Conclusion

The expansion of validity inquiry into Critical Validity Inquiry (CVI) is an important step in uniting the field of writing assessment with the more powerful field of Educational Measurement. The history of writing assessment, which we have reviewed in chapter two, is closely linked to the positivist traditions that still inform much of educational measurement’s practices at this time. The scholarship of literacy and composition can provide educational measurement with a new and powerful lens through which to view the validity of its practices.

The necessity for the field of composition to confront the powerful technology of testing and the misuse of testing as a means of reinscribing socio-economic relations is crucial to the success of the field in gaining control over the design and administration of its own assessment practices. Scholar activists seeking institutional change through their research must design their methodologies properly and continually reflect on the ability of their research questions to produce results. Within the field of rhetoric and composition we have two models for our future research agenda into institutional change: Porter and Sullivan’s (2000) institutional critique and Phelps’ Practice-theory-Practice Arc (1993). Outside of our field we can turn to the unitary concept of validity and its stress on validity inquiry along with Bourdieu’s notion of “epistemological vigilance.”
Porter and Sullivan (1997) (quote Bourdieu’s notion of “epistemological vigilance” as a correlative concept to their own reflexive methodology. Praxis, for Porter and Sullivan, is

the move to problematize method—part of which requires overthrowing the neat compartmentalization offered by traditional methods, part of which requires a self-reflectiveness about process that is incompatible with traditional attempts to understand, undisturbed by the researcher, the cultures/phenomena they study. (p. 54)

As we have seen, a unified theory of validity, as described by Samuel Messick, is another correlative to critical practice and “epistemological vigilance.” A unified concept of validity does not differentiate in terms of the traditional categories of validity: criterion, construct, concurrent, and content. Instead, as Messick (1989) explains, “validity is a unitary concept. Validity always refers to the degree to which empirical evidence and theoretical rationale supports the adequacy and appropriateness of interpretations and actions based on test scores” (p. 13). In the same manner in which Porter and Sullivan discuss traditional understandings of methodology being “portrayed as a set of immutable principles, rather than as heuristic guidelines, masks the impact of the situation—of the practice—on the study in ways that could unconsciously reinscribe theory’s dominance over practice,” so too has the cultural dominance of testing as a method for sorting,

24 “The term test score is used generically here in its broadest sense to mean any observed consistency, not just on tests as ordinarily conceived but on any means of observing or documenting consistent behaviors or attributes” (Messick, 1989, p. 13).
placing, and inscribing individuals as gifted or deficient masked the original purpose for
the use of testing. The now (in)famous Binet test is the archetypal example of the
method masking the situation. Alfred Binet provided a list of cardinal rules for his test:

1. The scores are a practical device; they do not buttress any theory of intellect.

2. The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking children.

3. Low scores shall not be used to mark children as innately incapable.

(Gould, 1996, p. 185)

These rules were promptly ignored by American test theorist as they modified the original Binet test in order to first test members of the U.S. Army during WWI and later to make college entrance decisions (Gould, 1984; Elliot, 2005; Lehmann, 1999).

The benefits and usefulness of a unified theory of validity are found in the fact that validity is a degree and validation a process. More simply put, the decisions made based on a test score can be valid to a degree, and through an ongoing process of inquiry into the consequences, both for the institution as well as for the individual, that degree of validity can be continually tracked and the instrument of assessment tuned to increase the degree of validity.

The need for validity inquiry, a process of validating the decisions and inferences that are being made based on test scores, is Porter and Sullivan’s notion of reflexivity and Bourdieu’s notion of “epistemological vigilance” in action. That decisions will be made
based off of an assessment is one thing, it is a much more challenging process to validate whether or not those decisions have a high degree of validity. This notion of reflexivity and reflection is also understood through Phelps’ Practice-Theory-Practice (PTP) arc (1993) and is presented by Phelps as a way of understanding research, theory, and practice as ongoing, mutually dependent processes of amending praxes (classroom practices, responding to student writing, placement procedures) (p. 39). Phelps describes the PTP arc:

in composition practice, problems have been defined and tackled in a characteristic way. First, a situation arises where teaching breaks down in ways that don’t yield to trial-and-error solutions. In these circumstances, the educational system shares with its popular critics the tacit assumption that pedagogy is both responsible for the problem and capable of solving it unilaterally. So crisis generates methodology, or more accurately, a babble of competing methods, each defended by passionate adherents.

(p. 37)

The understanding that something isn’t working then leads scholars to inquire into “the constituent processes and activities that underlie surface behavior or its products. The scholars construct theories and models of these processes, often by borrowing or adapting basic research from other fields” (p. 37). An obvious example of such a theoretical shift in the common practices of compositionists would be portfolio assessments replacing conventional single draft essays. When socio-cognitive models of language came into fashion the traditional assessment practice broke down. Scholars, borrowing from the
fine art’s tradition of judging complex performances as process rather than product provided a new practice that was in league with a deeper, richer theory of language. Phelps writes that “ultimately theories are brought into more comprehensive networks of meaning and metacriticism develops to evaluate the methods, assumptions, conclusions, and roles of the researchers themselves. At this it is likely that theoretical frameworks may effect radical, even paradigmatic, changes in practice” (p. 37). A métis-based institutional critique transformed into an expansion of validity inquiry, critical validity inquiry (CVI), is an attempt to develop a reflexive, validation methodology that borrows from a number of disciplines in order to cut through the institutional cunning that reinscribes power relations through the use of educational assessment practices.
Appendices I

Writing 098 or 150?

A Guide to Placing Yourself in the Freshman Course that is Right for You

A key part of our educational mission at Grand Valley is to help you develop the communication skills needed in your future profession, whether that profession is engineering, accounting, nursing, physical therapy, graphic design, or any of the other dozens of majors we have to offer.

As a Grand Valley student you will write, both formally and informally, in the majority of your classes. At least three of your required classes will include a significant amount of writing instruction: Writing 150 (freshman composition) and two Supplemental Writing Skills (SWS) classes, usually taken after the freshman year. All students are also required to fulfill the junior-level writing requirement either by writing a passing two-hour essay or by taking Writing 305, a course introducing students to writing in specific academic disciplines. The overall aim of the writing program is to help you succeed in your college courses and ultimately graduate from Grand Valley as a fluent, confident, versatile writer.

Your first step is to select the freshman composition course that best suits your needs and abilities.

Writing 150
Grand Valley students may begin in either of two writing classes: Writing 098, a preparatory class that offers no credit toward graduation, or Writing 150, a four-credit class focusing on academic writing, with a special emphasis on research-based writing. Half of all WRT 150 class meetings take place in a computer classroom, where students use PCs in all phases of the writing process. As well as being equipped with current word processing and spreadsheet software, each computer is connected to the library, the Internet, and electronic mail. In WRT 150, students typically write five four-to-six page essays, at least two of which involve research.

You should think carefully about which class is right for you. Some students are prepared to meet the goals of freshman composition in a single term, while others need two terms or more of practice and instruction before they are ready to move on to the writing demands of upper-level classes. In any case, you must earn a "C" or better in WRT 150 in order to satisfy the freshman composition requirement, and in order to earn the "C," you must demonstrate a variety of abilities.

- read and discuss challenging material
- summarize and analyze what you've read
- develop brainstorming and planning techniques for writing
- conduct library research at the planning, drafting, and revising stages of writing
- participate in writing workshops and conferences
- write essays that establish and maintain a single focus or thesis
- develop your ideas with details, examples, and discussions
- use transitions and other devices to lead readers through your essays
• cite and integrate the ideas and information of others into your writing
• use style, tone, and sentence structure for strategic effect
• format and edit your writing to conform to standard academic conventions

A Typical Course Sequence

<table>
<thead>
<tr>
<th>Year</th>
<th>Credits</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>0-30</td>
<td>All students must pass WRT 150 with C or better. Students may begin with either 098 or 150</td>
</tr>
<tr>
<td>Sophomore</td>
<td>30-60</td>
<td>Take the first Supplemental Writing Skills Course</td>
</tr>
<tr>
<td>Junior</td>
<td>60-90</td>
<td>Take Junior Writing Assessment: if you pass go on to your second SWS class. If you don't pass, take WRT 305</td>
</tr>
<tr>
<td>Senior</td>
<td>90-120</td>
<td>Take second SWS course</td>
</tr>
</tbody>
</table>

Should I Take Writing 150?

GENERALLY speaking, you are well-prepared for WRT 150 if you have done quite a bit of reading and writing in high school. WRT 150 instructors will assume that you can read, summarize, and analyze published material from magazines, newspapers, books, and scholarly journals. They will also assume that you have written a variety of essays in a variety of forms, including narrative, descriptive, and persuasive writing. Here is a checklist of general characteristics that should help you decide if you are ready for WRT 150:
In high school, I wrote several essays per year
In the past year, I have read books for my own enjoyment
In high school, I wrote several essays per year
My high school GPA placed me in the top third of my class
I have used computers for drafting and revising essays
My ACT-Writing score was above 20
I consider myself a good reader and writer

Should I Take Writing 098?

IF you would not use the characteristics listed under "Should I Take Writing 150?" to describe yourself, or if you don't consider yourself a particularly strong reader or writer, you may want to consider taking WRT 098 before WRT 150. In WRT 098, you will focus on writing with a purpose—that is, writing in specific ways to reach specific audiences. You will write often—in order to develop comfort and fluency as a writer. And you will work on mastering the conventions of standard written Writing-spelling, grammar, punctuation, and usage. Here is a list of general characteristics that may indicate that WRT 098 is best for you:

☑ Generally, I don't read when I don't have to
☑ In high school, I did not do much writing
☑ My high school GPA was about average
☑ I'm unsure about the rules of writing—commas, apostrophes, and so forth
☑ I've used computers, but not often for writing and revising
My ACT-Writing score was below 20

I don’t think of myself as a strong writer

IN WRT 098, you will read successful samples of essays written by professionals and by other students. In a typical class, you will write five or six short essays—about two to three pages each. You may cite some of the essays you have read or people you have interviewed, but generally you will not write research-based essays. Indeed, the purpose of WRT 098 is to give you the confidence, organization, and command necessary to write the research-based essays demanded in WRT 150 and beyond.

The Writing Center

THE Writing Center, located in room 201 of the Student Services Building (895-3451), offers tutoring services to all writing students. In WRT 098, you work with a tutor in regularly scheduled meetings with two or three classmates, so you end up spending about twelve hours working closely with a tutor and several peers. As part of its service to WRT 150, the Writing Center assigns one tutor to every class. The tutor normally attends all fourteen computer-classroom sessions to assist you with all aspects of your writing process, from small-group brainstorming to editing workshops. Writing Center tutors are specially qualified and trained students who work closely with the full-time director of the Center. Writing Center tutors are also available daily for walk-in consulting.

Library Skills

IN WRT 150, you will be introduced to the university library through a Research Skills Instruction Video and a written assignment called the Library Research Profile, in which
you will begin researching a topic for a research-based essay. As a college student, you will use many different kinds of documents: books, of course, but also specialized encyclopedias, government publications, scholarly journals, and magazines and newspapers. In order to find information in these documents, you must become familiar with our various print indexes, our on-line catalogue, and our CD-ROM databases. The library skills you develop in your freshman year are crucial to your success in the classes you take in your sophomore, junior, and senior years.

The Writing Portfolio

A "portfolio" is simply a collection of work. Many professionals use portfolios to show other people what they are capable of producing. In WRT 150, the majority of your final grade will be based on a portfolio of three finished pieces of writing that will represent your capabilities as a writer by the end of the term. You will choose the three essays from your semester's work. Each portfolio is read and evaluated by at least two WRT 150 teachers. As faculty members, they represent the academic audience for which you'll be writing in college. Because you need to create papers that capture and hold the attention of college-educated readers, that present focused and supported ideas that challenge the intellect of such readers, and that conform to the general conventions of academic writing, we feel that basing final grades on the judgment of more than one teacher is the most reliable way to evaluate student performance. You will get more details about this grading system on the first day of class.

What to Expect on the First Day
IN both WRT 098 and WRT 150, you will be asked to write a brief essay on the first day of class. Your teacher will read the essay as an indication of your writing abilities and let you know whether he or she feels you are beginning at the appropriate level. On the advice of your teacher, you will have the opportunity to switch from WRT 098 to WRT 150, or vice versa, during the first week of classes. The final decision is yours, but your teacher will be able to help you assess your writing in relation to course goals and the abilities of other entering freshmen.

FROM now until the first day of class, you can do some general preparation to ensure that the first-day writing sample is indeed representative of your abilities. Most likely, you will be asked to write about a topic of general interest or of particular interest to you personally. That is, you will be able to use your own thoughts, experiences, and observations as support for the claims and ideas you present in the essay. You might begin your general preparation by discussing "in-class writing" with one of your past or current writing teachers. What strategies are most effective in such writing situations? What basic features should your essay include? You might also review your notes or textbooks from the writing classes you have had. Other than that, just try to be well-rested and ready to write on the first day of class.

Still Unsure?
If you are still unsure about which Writing course to take, talk with your academic adviser during orientation, or call Dan Royer at 895-3405. We will be happy to help you make the decision that is right for you.
Writing Department
107 Lake Superior Hall
Grand Valley State University
Allendale, MI 49401
Appendices 2

Survey on Potential Indicators of Writing Ability

8. 24% said “Generally I don’t read when I don’t have to.”
9. 23% said “I don’t think of myself as a strong writer.”
10. 15% said “My ACT-English score was below 20.”
11. 12% said “My high school GPA was about average.”
12. 12% said “I’m unsure about the rules of writing.”
13. 9% said “In high school, I did not do much writing.”
14. 6% said “I’ve used computers, but not often for writing and revising.”
References


-----.

