

1

TO TELL THE TRUTH

Beyond Rubrics

College writing research in the disciplinary period which began, roughly, in the mid-1960s has not told us much about exactly what it is that teachers value in student writing. Researchers who have used statistical methodologies to address this question have thrown little light on the issue. . . . And guidelines published by English departments—at least at places where I've taught—are even less specific. An “A” paper is one that “displays unusual competence”; hence, an “A” paper is an “A” paper.

FAIGLEY, *Fragments of Rationality*

Consider your favorite college or university writing program. Instructors in the program may include tenure-line faculty, adjunct faculty, graduate teaching assistants, and an administrator or two. Some are new to the program; some have been there thirty years. Several of them are trained in the field of composition and teach it by choice; others teach writing only when they can't teach literature; a few are on the writing staff mainly because it's a paying job. This diverse troupe probably delivers one or two required introductory composition courses to nearly every student who appears at your institution's door. Though they diverge considerably in their backgrounds, emphases, interests, and areas of expertise in teaching rhetoric, your program's instructors almost certainly teach a clearly established curriculum, including common readings, writing assignments, writing processes, and educational goals.

Now ask yourself about these teachers of college composition the question Lester Faigley implies in the epigraph above: What exactly do they value in their students' writing? More likely than not, your writing program's best answer will be found in a rubric or scoring guide, the “guidelines published by English departments” Faigley mentions. Hundreds of such guides for writing assessment are available in books and on the worldwide web, and many writing programs have their own. A prominent example can be found in the back of Edward M. White's *Teaching and Assessing Writing* (298). White's “Sample Holistic Scoring Guide” (“prepared by committees in the California State University

English departments, 1988”) identifies six levels of rhetorical achievement. At CSU, a student’s text qualifies for the highest rating (“superior”) if it meets the following five criteria:

- Addresses the question fully and explores the issues thoughtfully
- Shows substantial depth, fullness, and complexity of thought
- Demonstrates clear, focused, unified, and coherent organization
- Is fully developed and detailed
- Evidences superior control of diction, syntactic variety, and transition; may have a few minor flaws

As a statement of the key rhetorical values of CSU English departments, I find this guide admirable in its clarity, simplicity, and emphasis on intellectual and rhetorical substance over surface mechanics or format concerns. Furthermore, by presenting not only levels of achievement (“incompetent” to “superior”) but also the five specific evaluative criteria quoted above, it goes far beyond the tautological “A = A” formulation that Faigley protests.

But does it go far enough? The strength of the hundreds of rubrics like White’s lies in what they include; their great weakness is what they leave out. They present to the world several inarguably important criteria endorsed by the local writing program administrator as the criteria by which writing should be evaluated in the relevant program. They omit any mention of the dozens of other criteria (such as “interest,” “tone,” or “legibility”) by which any rhetorical performance is also likely to be judged. In pursuit of their *normative* and *formative* purposes, traditional rubrics achieve evaluative brevity and clarity. In so doing, they surrender their *descriptive* and *informative* potential: responsiveness, detail, and complexity in accounting for how writing is actually evaluated.

We need to critically examine such representations of our rhetorical values on the basis of what they teach—and fail to teach—students, faculty, and the public about the field of writing instruction. Theories of learning, composition, and writing assessment have evolved to the point at which the method and technology of the rubric now appear dramatically at odds with our ethical, pedagogical, and political commitments. In short, traditional rubrics and scoring guides prevent us from telling the truth about what we believe, what we teach, and what we value in composition courses and programs.

Theorists of knowledge from Nietzsche to Foucault and beyond have taught us that calls for “truth” cannot go unexplained. So I propose this

working definition of *truth* as I use it in this book: truth means *doing our epistemological best*. Before we make a knowledge claim (for example, Here is how writing is valued in our writing program) that carries with it serious consequences for students, faculty, and society, we need to conduct the best inquiry we can. In their rush toward clarity, simplicity, brevity, and authority, traditional scoring guides make substantial knowledge claims based on inadequate research.

A prime assumption of my work is that a teacher of writing cannot provide an adequate account of his rhetorical values just by sitting down and reflecting on them; neither can a WPA provide an adequate account of the values of her writing program by thinking about them or even by talking about them in general terms with her writing instructors. In this book I offer what I believe is a method of evaluative inquiry better grounded both theoretically and empirically, a method that yields a more satisfactory process of exploration and a more useful representation of the values by which we teach and assess writing.

WHAT WE REALLY TEACH

As a subfield of English studies, rhetoric and composition teaches and researches what educators generally accept as the preeminent intellectual skills of the university: critical and creative thinking, as well as interpretation, revision, and negotiation of texts and of the knowledge those texts are used to create (Berlin). Done well, this work prepares our students for success in personal relationships, careers, and democratic citizenship.

Most of us in the field would therefore likely embrace as part of our mission Marcia Baxter Magolda's call to help students move toward "self-authorship," defined as "the ability to collect, interpret, and analyze information and reflect on one's own beliefs in order to form judgments" (14). Unfortunately, the undergraduate experiences of participants in Baxter Magolda's study mainly lacked conditions that would have helped them develop in such sophisticated ways. As a result, during their early postcollege careers, they often struggled and stumbled in their efforts to "become the authors of their own lives." Reflecting on what higher education did and did not offer her study's participants, Baxter Magolda explains:

They would have been better prepared for these [early interpersonal, career, and citizenship] roles, and have struggled less, had the conditions for self-authorship been created during their college experience. (xxii)

The key academic principle that helps students move toward self-authorship is, according to Baxter Magolda, “*that knowledge is complex, ambiguous, and socially constructed in a context*” (195, Baxter Magolda’s emphasis). Theory, research, and teaching in rhetoric and composition strongly support such views of knowledge, including its social and context-specific character. However, rubrics, the most visible and ubiquitous tool of *writing assessment*—arguably the aspect of rhetoric/composition that impinges most powerfully and memorably on our students’ lives—teach our students an exactly opposite view of knowledge, judgment, and value. At the heart of our educational and rhetorical project, rubrics are working against us.

For all its achievements and successes over the past half century (see Yancey), the field of writing assessment has no adequate method for answering one of its most urgent and important questions: What do we value in our students’ writing? What we have instead are rubrics and scoring guides that “over-emphasize formal, format, or superficial-trait characteristics” of composition (Wiggins 132) and that present “generalized, synthetic representations of [rhetorical] performances . . . too generic for describing, analyzing, and explaining individual performances” (Delandshere and Petrosky 21). Instead of a process of inquiry and a document that would highlight for our students the complexity, ambiguity, and context-sensitivity of rhetorical evaluation, we have presented our students with a process and document born long ago of a very different need: to make assessment quick, simple, and agreeable. In the field of writing assessment, increasing demands for truthfulness, usefulness, and meaningfulness are now at odds with the requirements of efficiency and standardization. The age of the rubric has passed.

Regarding rubrics and scoring guides everywhere, I raise questions not only about their content but also about their origins and uses. To determine what we really value in a particular writing program, we must therefore pursue several related questions:

How do we discover what we really value?

How do we negotiate differences and shifts in what we value?

How do we represent what we have agreed to value? and

What difference do our answers to these questions make?

Compositionists willing to address these questions of inquiry, negotiation, representation, and consequences, will, I believe, find traditional rubrics and scoring guides lacking crucial ethical, pedagogical, and

political qualities. This book points the way toward a new method for discovering, negotiating, and publicizing *what we really value* in students' writing. We can reclaim what rhetoric and composition lost half a century ago when it adopted rubrics and scoring guides as its preeminent method of representing a writing program's rhetorical values.

THE BIRTH OF RUBRICS

Modern writing assessment was born in 1961 in Princeton, New Jersey. That year, Diederich, French, and Carlton of the Educational Testing Service (ETS) published *Factors in Judgments of Writing Ability* (ETS Research Bulletin 61-15). Basing their work on nearly a decade of research already done (at ETS and elsewhere) on writing assessment generally and especially on inter-rater agreement, Diederich, French, and Carlton declared that

The purpose of this study was to serve as a stepping stone toward closer agreement among judges of student writing . . . by revealing common causes of disagreement [among those judges]. ("Abstract")

Though the coauthors of the study emphasized that "It was not the purpose of this study to achieve a high degree of unanimity among the readers but [rather] to reveal the differences of opinion that prevail in uncontrolled grading—both in the academic community and in the educated public," still they found themselves "disturbed" by the wide variability in scoring among their fifty-three "distinguished readers": "[I]t was disturbing to find that 94% of the [300] papers received either seven, eight, or nine of the nine possible grades; that no paper received less than five different grades; and that the median correlation between readers was .31" (Abstract). The next half-century of research and practice in writing assessment was definitively charted in the single word "disturbing." Diederich, French, and Carlton had purposefully set out to investigate the geography, if you will, of rhetorical values among fifty-three "educated and intelligent" readers in six fields. Yet the truth they revealed "disturbed" them. Why? Because within the world of positivist psychometrics, the world in which ETS and other commercial testing corporations still operate, precise agreement among judges is taken as the preeminent measure of the validity of an assessment. Therefore, rather than seek to understand and carefully map out the swampy, rocky, densely forested terrain of writing assessment that they found lying before them, they quickly moved to simplify and standardize it thus:

A classification of comments was developed by taking readers at random and writing captions under which their comments could be classified until no further types of comments could be found. After many changes, 55 categories of comments were adopted and arranged under seven main headings:

1. Ideas
2. Style
3. Organization
4. Paragraphing
5. Sentence structure
6. Mechanics
7. Verbal facility

(Diederich, French, and Carlton 21)

Using factor analysis—one of the “statistical methodologies” Faigley views with skepticism—the ETS researchers eventually derived from those seven main headings a list of five “factors” that seemed to capture the values of their readers:

Ideas: relevance, clarity, quantity, development, persuasiveness

Form: organization and analysis

Flavor: style, interest, sincerity

Mechanics: specific errors in punctuation, grammar, etc.

Wording: choice and arrangement of words

And thus was born what became the standard, traditional, five-point rubric, by some version of which nearly every large-scale assessment of writing since 1961 has been strictly guided.

Two things leap off every page of *Factors in Judgments of Writing Ability*: the analytical and imaginative genius of the researchers and their sure command of methods of statistical analysis, most especially factor analysis. What is less obvious is that, because they were firmly committed to the positivist, experimentalist paradigm that ruled their day, they let a historic opportunity slip away. Confronted with an apparent wilderness of rhetorical values, they retreated to a simplified, ordered, well-controlled representation that would keep future writing assessment efforts clean of such disturbing features as dissent, diversity, context-sensitivity, and ambiguity. Three centuries earlier, officials of the Catholic Church turned away from Galileo’s telescope, so “disturbed” by what it showed them that they refused to look into it again. After showing the astronomer the Church’s instruments of torture to inspire his recantation and subsequent silence,

they returned eagerly to Ptolemy's comfortingly familiar millennium-old map of the heavens showing earth as the motionless center of the universe.

Within the paradigms of positivist psychometrics (Moss 1994) and experimentalist methods (North), dramatic evaluative disagreements like the ones Diederich, French, and Carlton carefully documented must necessarily register as failures. Disagreement is failure because positivism presumes a stable and independent reality (in this case "writing ability" or "writing quality") that humans try more or less successfully to "measure." The fact that "94% of the papers [in their study] received either seven, eight, or nine of the nine possible grades," was "disturbing" because if evaluators had been free of "error" and "idiosyncrasy" (Diederich's terms), they would have agreed in their judgments. The obvious psychometric solution to the problem of disagreement is to rebuild readers' evaluative frameworks so they can agree more consistently and more quickly. From the beginning this has been the precise purpose of the scoring guide and rubric, and these tools have served us well: when we use them, we can reach impressively high correlations among scorers, and we can judge students' writing with remarkable speed and relative ease.

From the standpoint of a constructivist or rhetorical paradigm, however, Diederich, French, and Carlton's achievement seems more *scientistic* than scientific. That is, it seems to provide an abstracted, idealized representation and to take on the appearance of objective truth by turning away from the messy facts at hand. Their epistemological frame and their statistical methods prevented them from delving into the powerful knowledge they revealed and clearly stated in their report: the "empirical fact" that

If [readers'] grades do not agree, it is not for lack of interest, knowledge, or sensitivity, but because competent readers with their diversity of background will genuinely differ in tastes and standards. (10)

The 1961 study by Diederich, French, and Carlton created an opportunity to chart in detail the axiological terrain of writing assessment among this group of "distinguished readers." The ETS team of researchers achieved the chance to show the world what real experts working in real professions in the real world valued in real college students' writing. This is the truth and the reality they discovered, and it could have provided them and us with a powerful authority and reference point for understanding writing assessment. Their positivist pre-suppositions and methods, however, compelled them in a different direction. Decrying their findings as "disturbing" and full of evaluators'

“error,” Diederich, French, and Carlton traded in the rhetorical truth confronting them (that readers value texts differently) in exchange for the grail of high inter-rater agreement. It is now time to trade back the grail and reclaim the truth.

First, however, we should acknowledge the benefits rubrics brought us during their fifty-year ascendancy. In the historical context of U.S. culture in 1961 and the following decades, rubrics may have done more good for writing assessment and the teaching of writing than any other concept or technology. During a time when educators were under constant pressure to judge “writing” ability using multiple-choice tests of grammar knowledge, the work of Diederich, French, and Carlton (and other researchers at ETS and elsewhere) legitimized direct assessment of writing (assessment that took actual *writing* as the object of judgment). In the 1960s, to become acceptable and legitimate, writing assessment first had to prove itself “scientific.” And scientificness, in the context of positivist psychometrics, required a level of standardization and agreement that would provide the impression that the thing being judged—in this case *writing ability*—was indeed an objective “thing.” If readers disagreed significantly—as Diederich’s study resoundingly proved they did—then the procedure was by definition “unscientific.” This is why Diederich, French, and Carlton treated their “empirical facts” showing the fabulous richness and complexity of rhetorical values merely as a “stepping stone” toward a cleaner, clearer, more agreeable representation of the judgment of writing. First and foremost, then, rubrics bought writing assessment legitimacy.

But that’s not all. Streamlining and reducing the possibilities for rhetorical judgment also speeded up what every writing teacher knows to be a mind-numbing and time-devouring task. So just as the agreeability rubrics promoted brought us legitimacy, the increased speed they provided brought us affordability. When time and money are short—as they always are in education—rubrics provide badly needed relief and enable faculty to assign and judge actual writing from large numbers of students with relative speed and ease.

Scoring guides yielded yet another set of advantages: documentation of the process of evaluating writing. Notoriously obscure in nature, the judgment of writing quality was at last explained in clear, simple, and public terms, so judgments could be held accountable to the values specified in the rubric. Students, instructors, and the general public could hold in their hands a clear framework for discussing, teaching,

and assessing writing. As public records of rhetorical values, rubrics made writing assessment less capricious.

Legitimacy, affordability, and accountability— these are the gifts brought to our field in 1961 by Diederich, French, and Carlton and by the rubrics they helped establish as a permanent fixture in large-scale writing assessment.

BEYOND RUBRICS

The 1990s brought dramatic changes to the fields of psychometrics and writing assessment. In this new theoretical context, we can look at rubrics with fresh eyes and raise critical questions about them that should open up new possibilities for writing assessment.

The key developments are in the area of validity theory. Because other scholars of assessment (Huot 1996; Moss 1992; Williamson; Yancey) have already addressed these changes in detail, I will touch on them only briefly before focusing on their implications for rubrics and for the practice of writing assessment beyond rubrics.

Assessment has become a public and educational issue, not solely a technical one. We no longer seem to be content to be told that assessments meet certain psychometric and statistical requirements. We want to know whether they help or hurt learning and teaching. Delandshere and Petrosky provide an apt synopsis of this shift in assessment theory:

In the last few years, there has been a shift in the rhetoric (if not yet the practice) of assessment (e.g., Gipps, 1994; Resnick & Resnick, 1992; Shepard, 1989; Wiggins, 1993; Wolf, Bixby, Glenn, & Gardner, 1991). Much more emphasis has been placed on the support of learning and teaching than on the sorting and ranking of individuals. (15)

To put it another way, researchers, theorists, and practitioners have joined together in their refusal to ignore the educational impact of assessments. As Wiggins states,

In performance tests of writing we are too often sacrificing validity for reliability; we sacrifice insight for efficiency; we sacrifice authenticity for ease of scoring. Assessments should improve performance (and insight into authentic performance), not just audit it. (129)

These researchers have thrown off positivist-psychometric prohibitions against considering contextual factors and thus have been willing

to challenge assessment practices that appear to undermine or corrupt best educational practices.

The key flashpoint in this controversy has been the issue of inter-rater agreement (a key aspect of “reliability”) and its relationship to validity. Moss addressed the question bluntly in the title of her essay “Can There Be Validity Without Reliability?” (Moss 1994). Her answer to that question is a qualified “yes.” Moss and others rely in turn for much of their argument on the work of psychometricians Lee J. Cronbach and Samuel Messick. On key validity issues, these theorists of evaluation have argued, for example, that “validity” is not a quality of an assessment instrument (a test or a portfolio assessment, for example), but rather a quality of the *decisions* people make on the basis of such instruments. This claim takes validity out of the technical realm to which psychometrics had relegated (and thus controlled) it for decades and puts it back into the realm of public discourse, in which anyone could speak to the educational effects and conceptual integrity of an assessment and thus to its validity.

Arguing for a “unified theory of validity,” Michael Neal (building on the work of Huot, Williamson, and others) presents five different issues that must be considered in assessing validity:

1. criterion-related validity
2. reliability [especially inter-rater agreement]
3. traditional construct validity
4. content validity
5. social consequences

(Neal 2002)

The key point here for my study is that validity has become more broadly and more multiply defined, and this has opened up new avenues of discussion, debate, and inquiry.

Grant Wiggins raises a particularly strong voice in the struggle to broaden and humanize assessment issues and, especially, validity, in his article “The Constant Danger of Sacrificing Validity to Reliability: Making Writing Assessment Serve Writers.” As the title makes clear, Wiggins takes part in a shift in validity theory away from nearly exclusive attention to inter-rater agreement (the chief consideration to which Diederich and his colleagues were devoted) and toward concerns driven by his core argument, that “Tests themselves teach.” The crucial question then becomes “What do our assessments teach?”

The fact is that almost all writing assessments I have seen use rubrics that stress compliance with rules about writing as opposed to the real purpose of writing, namely, the power and insight of the words. Writing rubrics in every district and state over-emphasize formal, format, or superficial-trait characteristics. (132)

The net effect of this validity revolution is to shift attention to the broad impact of assessments on teaching and learning and to judge the appropriateness of assessments based on the outcomes of high-stakes decisions affecting students' lives.

These recent shifts in validity theory cast writing assessment—especially rubrics and scoring guides—in a different light from that in which Diederich and his followers saw them. In 1961, agreement was the only discernible path to proving the legitimacy of direct writing assessment. Half a century later, in great part thanks to the work of Diederich, French, and Carlton and their followers, most assessment specialists (even those who work for testing corporations) support direct writing assessment. With a growing focus on the importance of connecting assessment to teaching and tapping into the pedagogical and epistemological value of differences among readers, Diederich's turn away from the disturbing complexity of real readers toward the standardized and simplified portrait presented by the rubric takes on new meanings.

Consider the validity double bind in which scoring guides necessarily find themselves. Remember that the function of the guide is to focus and narrow the factors or criteria by which readers judge in order to boost efficiency and agreeability. Whether the guide succeeds or fails in this task, it precipitates a crisis of validity in the assessment.

If, as often appears to be the case (Charney; Delandshere and Petrosky), the guide is "successful" in streamlining the messy mix of rhetorical values instructors bring to communal writing assessment, then that evaluation judges students according to values different from those by which instructors have taught and assessed writing in their classrooms. Therefore, the shared evaluation diverges from the "real" values of the writing program as enacted by the program's instructors in the program's classrooms. For the assessment to be relevant, valid, and fair, however, it must judge students according to the same skills and values by which they have been taught.

If, on the other hand, as other researchers have shown (Barritt, Stock, and Clark; Huot 1993), scoring guides often fail to narrow and

standardize the values by which teachers judge students' writing, then they significantly misrepresent the values at work in the writing program (both in classrooms and in the shared assessment process). In this case the scoring guide represents not what instructors value in students' writing, but rather what someone believes the program's rhetorical values to be, or what someone wants them to be, or what someone wants people to believe them to be. The rubric claims to represent the evaluative terrain of the program, but at the very best it documents only a small fraction of the rhetorical values at work there.

So rubrics may or may not succeed in narrowing the range of criteria by which we judge student's writing. The problem is that *either way*, they fail important tests of validity and ethics. Still we lack a rigorous, detailed, and accurate account of what instructors in the program value in teaching and assessing writing.

Not only as a document does the rubric or scoring guide fall short. It also fails us as a process of inquiry. Very rarely do rubrics emerge from an open and systematic inquiry into a writing program's values. Instead, they are most often drafted by an individual (usually the writing program administrator) and approved by a committee before being delivered into the hands of evaluators as the official and authoritative guide to judging students' writing. By predetermining criteria for evaluation, such a process shuts down the open discussion and debate among professional teachers of writing that communal writing assessment should provide. Delandshere and Petrosky describe the problem this way:

Ratings "force fit" performances into general categories; they mold them into abstractions that are defined solely to create scores and score variance, but they do not seem very helpful in describing or explaining the particulars of any one performance. As such, ratings are poor representations of complex events and, given the generic nature of the scoring rubrics, the danger exists that the assessors will attend to the most visible or superficial characteristics of the performance while neglecting the specific patterns and connections that give it meaning and integrity. (21)

More open and meaningful discussion and debate would not only improve the quality and legitimacy of the resulting document, but would also provide participants with one of the most powerful opportunities for professional development available to teachers of writing.

Instructors, administrators, and researchers of writing, as well as our students, our colleagues elsewhere in the academy, and the general public, all

deserve both a rigorous inquiry into what we really value and a detailed document recording the results of that inquiry. This book advocates a new method of evaluative exploration, demonstrates that method in the context of one university's writing program, highlights the method's usefulness to other writing programs, and suggests specific actions for instructors and administrators of composition who wish to move beyond rubrics to more robust truths about teaching and assessing writing.

DYNAMIC CRITERIA MAPPING

Having argued that positivist psychometrics and its scoring guides have cost writing instruction a great deal, I now propose an alternative method for inquiring into and documenting how we really evaluate our students' texts. The method I recommend and demonstrate in this book I have named "Dynamic Criteria Mapping." Dynamic Criteria Mapping (DCM) is a streamlined form of qualitative inquiry that yields a detailed, complex, and useful portrait of any writing program's evaluative dynamics. Later chapters provide a detailed example of the results of this form of inquiry, as well as recommendations for specific DCM strategies. Here I will explain the theoretical background for the method.

The interplay between two forms of educational inquiry, writing assessment and qualitative research, has gained increasing attention lately in books and journals devoted to improving postsecondary teaching and learning. Scholars in a variety of fields—composition and rhetoric (Huot 1996), evaluation theory (Moss 1996), and qualitative research (Guba and Lincoln)—have directed researchers' attention to the benefits of integrating qualitative methods into educational evaluation.

In *(Re)Articulating Writing Assessment for Teaching and Learning*, Huot outlines six principles that inform what he sees as the coming wave of writing assessment practices. Huot foresees that the new generation of assessment programs will be

1. Site-based
2. Locally controlled
3. Context-sensitive
4. Rhetorically based
5. Accessible

(178)

These five principles clash dramatically with those of positivist psychometrics, which privileges instead generalizability, context-independence,

statistical analysis, secrecy, and control (see “Beyond Rubrics” section above). Huot’s five principles also correspond closely to the features that set qualitative research off from the dominant, experimentalist paradigm. Especially in his call for a grounding in rhetoric and a sensitivity to context, Huot touches on the key distinguishing features of interpretive research. In other words, when Huot predicted an impending (now ongoing) historical shift away from positivist psychometrics in evaluation, he simultaneously predicted the ascendance of qualitative methodology in writing assessment.

Pamela A. Moss, a prominent theorist of educational evaluation, calls for a shift in methods quite similar to Huot’s. Distinguishing between a “naturalistic conception of social science” and an “interpretive conception,” Moss advocates for including interpretive research methods in educational assessment.

A second way in which interpretive research traditions might inform assessment theory and practice is by providing theory, illustrations, and methodologies that illuminate a complex social reality, thus enabling us to better understand the ways that assessments work within the local context. (1996, 27)

And Guba and Lincoln, in their landmark *Fourth Generation Evaluation*, make a similar argument. Here they explain why they term their approach to evaluation a “constructivist” one.

The term *constructivist* is used to designate the methodology actually employed in doing an evaluation. It has its roots in an inquiry paradigm that is an alternative to the scientific paradigm, and we choose to call it the *constructivist* but it has many other names including *interpretive* and *hermeneutic*. (39 Guba and Lincoln’s emphases)

These authors have pointed to a shift in assessment theory and practice toward an interpretive or qualitative paradigm. To date, however, the effort has remained mainly theoretical, and the limited application of such methods to writing assessment has been conducted by researchers, not practitioners.

The field of composition needs a workable method (supported by a well-developed theory) by which instructors and administrators in writing programs can discover, negotiate, and publicize the rhetorical values they employ when judging students’ writing. Huot, Moss, and Guba and Lincoln have clearly called for this need to be satisfied by combining assessment theory and practice with qualitative research. Grounding

these scholars' proposed integrative project in my study of a specific university's first-year writing program, I demonstrate and propose such a method: Dynamic Criteria Mapping.

The following chapter, "Studying Rubric-Free Assessment at City University: Research Context and Methods," introduces a distinctive first-year writing program featuring a portfolio assessment system that was, at the time of my field research, metamorphosing from the psychometric paradigm to a hermeneutic one (see Broad 2000). Several progressive elements in this program—including the absence of a scoring guide or rubric—made possible my inquiry into rhetorical values and the construction of the DCM model. Chapter 2 also describes my research methods, which extend from Glaser and Strauss's grounded theory approach. Chapters 3 and 4 ("Textual Criteria" and "Contextual Criteria") present the findings that emerged from my study of what they really valued at City University. Chapter 5, "A Model for Dynamic Criteria Mapping of Communal Writing Assessment," highlights the benefits of DCM and offers specific suggestions for writing programs that wish to take up the sort of inquiry advocated here.

I aim to engage readers in questions of truth and method in program-wide writing assessment, persuade them of the need for change, and demonstrate specific techniques that will solve urgent problems we currently face. The immediate outcome should be a more productive process of evaluative inquiry and negotiation as well as a more detailed, accurate, and useful document to result from that inquiry. The long-term outcome should be better learning for students of composition, enhanced professional development for writing instructors, and increased leverage with the public for writing programs that can publicize a complex and compelling portrait of what they teach and value in students' writing.