# 4

# DESIGNS ON ASSESSMENT
## *University of Nevada, Reno*

Jane Detweiler and Maureen McBride

## LOCAL CONTEXT

The University of Nevada, Reno is the state's flagship research university, with a long tradition of excellence in providing a liberal arts education. At most recent count, about 15,000 students are enrolled (about 12,000 undergraduate and 3,000 graduate). Its "vertical" Core curriculum was created/elaborated over a number of years beginning in the 1980s, with first-year math and writing courses, a three-course humanities sequence, distribution requirements in fine arts, social sciences, and sciences, and junior- and senior-level general and major capstone courses as writing-intensive, culminating experiences.

The Core writing program administers a three-course sequence. By standardized test scores, students initially place in English 098 (Preparatory College Writing), English 101 (Beginning College Writing), or English 102 (Intermediate College Writing). Since we think that students' actual writing provides a better indicator of their proficiency and practical experience, we also have an alternative portfolio placement process wherein students compile a collection of at least three samples of their best recent writing. Students may also place into or out of English 102 by their scores on advanced placement exams.

English 102 is the course required by the Core curriculum, and students must pass it to enroll in the Core humanities sequence and move on through to the capstone portion of their general education requirements. In English 101, students gain greater experience with the writing process, peer reviewing, focusing their writing on topics, reading critically, analyzing and shaping their writing for a variety of rhetorical situations, and understanding writing genres and conventions. Building on this experience, English 102 challenges students to conduct research and to craft arguments based on evidence; this course is one in "general composition," to the extent that it doesn't take a discipline-specified approach, and that it emphasizes flexibility of response to a variety of writing contexts and conventions. The course is theme-based, with no set reader, rhetoric,

handbook, nor syllabus—each instructor designs his or her course around the program's student learning outcomes for English 102. The program runs just under one hundred sections of English 102 per year (with about the same number of English 101 sections, thirty-five sections of English 098, and a handful of English as a Second Language, honors, and other special courses). The teaching community at UNR is a mix of full-time faculty, teaching assistants (in writing/rhetoric and composition, literature, and literature and environment), and contingent faculty (temporary full-time lecturers and part-time instructors).

Under the leadership of Kathy Boardman, the Core writing program conducted a comprehensive portfolio assessment of the English 102 course in 2000, with a follow-up study in 2001 that focused on a few features that the initial study had revealed as possible areas for improvement in the curriculum. This was well in advance of our accrediting body's new interest in assessment, and Kathy's "closing of the loop" by changing instructor preparation and inservice training provided an impressive model of effective assessment practice.

### Our Assessment Project's Rhetorical Situation: A Comedy Of Expectations

Jane Detweiler took over leadership of the Core writing program in the summer of 2004, just shortly after the new director of the Core curriculum assumed his position. Amidst the usual pratfalls of beginning work in an administrative position, she realized that the Core director had designs–assessment project designs.

In a meeting that first fall semester, anticipating the accreditation cycle which would begin with a self-study in 2006-07, the Core director explained that he wanted to design and implement an assessment of the Core curriculum as a whole. As one might expect, he had already encountered a number of frustrations. The math and science departments were still in the midst of substantial restructuring of their programs; along with the social sciences, they plead inadequate time to prepare a curriculum *and* do an assessment of that curriculum. These disciplines would only be able to muster something like surveys of the "match" of student and teacher expectations for specific courses (read: substantive assessment in these disciplines would have to wait.) In areas of the core where curriculum was not undergoing wholesale revision, assessment would be more feasible and more necessary, given the upcoming accreditation. Hence, the Core curriculum director approached the directors of programs in Core writing and Core humanities, as well as the chair of the capstone committee, to propose a "vertical" assessment of general education: a study of writing and critical thinking in first-year writing, humanities, and the general capstones.

What follows is the Core writing program administrator's recollection of the dialogue (emphatically not verbatim, with events condensed, stylized, or omitted to suit her purposes as the teller of this tale):

*Core Curriculum Director*:      What I would like to do is to conduct a longitudinal study that would follow a cohort of students from Core writing through to the general capstones, using e-portfolios as data. I would like to study writing and critical thinking in English 102, the Core humanities sequence, and in those junior- and senior-level capstone courses.

*Core Writing Program Administrator, thinking to herself, explicating his simple declarations over the next few seconds:*      What? E-portfolios? I haven't heard about any e-portfolios. . . . You want to assess writing AND critical thinking? I was planning to assess my program, but you want me to collaborate with all those other departments across the disciplines? This had damned well better not turn out to be a value-added kind of assessment . . . .How in the heck do I assess critical thinking? I mean, it's part of what we do, but I've never learned about how to do assessment of THAT. Do you have any idea how hard this is going to be, and how much money it will cost? And keeping a "cohort" is harder than you think. Given just regular attrition, you're going to lose your cohort in no time, and the students don't take the courses in sequence or over a predictable number of years . . . it'll be seven or eight years before all of them complete everything . . . and, wait a minute, are you going to ask their permission to use their work? This whole e-portfolio databasing of student work is kind of creepy, especially if they don't know that we're using their work for assessment. And you want this done in what timeframe? The self-study is only a couple of years away!

*Core Writing Program Administrator, aloud*:      But the accreditation visit will be in 2007-08, which means that we need results in 2006-07 for the self-study. And, as far as I know, there are no e-portfolios going on in Core writing or in Core humanities.

*Core Curriculum Director:*      Right. So we can only *start* the longitudinal study, planning and getting the e-portfolios under way. In the meantime, we can do 'snapshots' of writing and critical thinking in English 102, Core humanities, and the general capstones. The faculty in each program will need to develop an assessment project that is 'local,' that examines how *they* teach and evaluate writing and critical thinking . . .

To his credit, the Core curriculum director steadfastly funded the locally-developed assessment projects he requested, using a line built into the general education program budget for this purpose. There were only the most

minor stumbles as he tried to shepherd along various local, faculty-driven assessment projects, like when he mused, "And, since they know how to assess writing, the faculty in English can figure out how to assess writing in the other areas . . . "

*Core Writing Program Administrator, worrying about how this conflicts with context-sensitive assessment, among other things, she points to her desk in the corner:*    You know, I have a day job. If you want me to assess other programs, that would be a full-time job in itself—what Core writing duties do you want me to let go, to make time for your larger study? All of them? I am ready and willing to *collaborate* with other program directors on a larger general education assessment, and to get cracking on the 'snapshot' you want me to do . . . .

Following these conversations, and over the course of several others, the Core curriculum director and his Core writing program administrator came to an agreement about what was feasible, given the timeframe and other constraints. The Core writing program administrator arranged to run a graduate-level internship in program assessment in the spring semester, for which she hastily assembled a packet of background readings and acquired two recent books on writing program assessment: Bob Broad's *What We Really Value* (2003) and Brian Huot's *(Re)articulating Writing Assessment* (2002). She planned that her graduate students would help design and implement an assessment project, in much the same way that Kathy Boardman's crew of interns had done years previously.

**Designs on Assessment at UNR**

As the UNR team began designing a portfolio assessment project intended to measure a general education program's success in preparing students to write effectively and think critically, we faced a number of difficult considerations. Perhaps the biggest challenge was our relative lack of experience with writing assessment and our even greater unfamiliarity with assessment of critical thinking. Still, we had the strong (financial and logistical) support of a Core curriculum director and an English department that valued our contribution to improving writing instruction for the roughly three thousand students who would pass through some or all of our three-course sequence in a given year.

In spring 2005, the Core writing program administrator and six interns[4] initiated their project, emphasizing an intensive study of the latest in

---

4.    The project described in this chapter benefited from the contributions of the interns who participated in the assessment coursework (Meg Cook, Michaela Koenig, Kara Moloney, and Eliot Rendleman), some of whom also later acted as the graduate assessment coordinators. Maureen McBride, Sarah Perrault, and Doug Walls helped to implement the project as designed, to interpret the results, and to write up the final report.

writing assessment, including recent assessments that focused on critical thinking (specifically, the work of Bill Condon and Diane Kelly-Riley at Washington State). Although we used Condon and Kelly-Riley (2004) as a starting point, our descriptions of critical thinking (and ultimately the features we identified as "critical thinking" features) were developed in our focus groups from the language that our instructors had used to describe critical thinking.

The six interns divvied up the background resources into broad areas, annotating key studies and sharing them via web courseware. Seeking to understand what would be a useful model, they also studied carefully the report and process records from Kathy Boardman's study (2000), and inquired further from her as necessary.

Their deliberations centered on some central insights:

- As Huot (2002) persuasively suggests, assessment should be a locally-driven, contextually-situated *rhetorical* enterprise, designed with the needs and interests of various audiences in mind.

- As Broad (2003) illustrates with his study, whatever assessment activities are to be conducted, they should begin with efforts to describe carefully and thoroughly what teachers in the program value and should result in representations that are useful and valid to those teachers.

- As Condon and Kelly-Riley (2004) point out, writing and critical thinking do not necessarily absolutely coincide (a piece of writing *can* be an effective response to an assignment and not demonstrate critical thinking at all).

- As Boardman's group found (2000), the information that those in the field found compelling and useful was not necessarily what central administration would find compelling and useful (indeed, it seemed that some "up the food chain" needed to be regularly reminded that an assessment *had* been conducted in the Core curriculum).

Further, the team had to bear in mind some central tasks or constraints:

- To design a study that would be a valid assessment

- To meet the demands of a key stakeholder (the Core curriculum director) and assess "writing" and "critical thinking" in English 102

- To link the program's outcomes for English 102 to Core curriculum-level outcomes in general education

- To propose a study with a reasonable-but-substantial-enough budget

- To complete the study in the time allotted (a little over one year from design to implementation to reporting of results)

Fairly early in the design process, the team encountered some difficulties that would need to be resolved for the assessment project to move forward. First, although there were plenty of materials that described the Core curriculum, there were no Core outcomes *per se*. In the accessible archives around the office were mission statements, program descriptions, course outlines, and even some self-study documents related to the Core curriculum. The writing program had its own course outcomes, designed to move students toward our definition of "effective writing" (which was the only apparent "outcome" articulated for Core writing in the Core curriculum documents). At the level of the general education program across the university, there were no measurable outcomes. The Core writing program administrator sought to move the Core board (the committee overseeing general education) to articulate Core outcomes (this finally occurred in late fall 2005 and early spring 2006—well after the design process for the assessment was completed).

A second difficulty arose as the team considered how to make any case that the Core writing program's teachers were meeting the expectations of our external audience of central administrators. Broad's arguments (2003) that we should carefully describe and document what we valued were profoundly compelling, and his thorough approach—*dynamic criteria mapping or DCM*—offered an exemplary way to begin the sort of contextually-valid, locally-driven assessment Huot advocates (2002). As we looked at the actual criteria map that resulted from Broad's study of a writing program, though, it seemed to us that a description like that would not be recognizable as the result of an assessment. It would be immensely useful to us as a teaching community, but in that form, it wouldn't easily allow for evaluation of a program's success.

This question of how to work with a map was one we confronted at IUPUI. When we tried to create one collaboratively with our writing program colleagues, we found that the complexity Bob Broad (2003) had represented in his map was missing in ours because as our faculty worked to create a graphic representation of our values, they also worked to simplify the result. Apparently, their previous experience with graphic representations of important ideas prompted this simplification. Later, as we worked toward a grading document we all could use, many asked, "Where is the rubric?" *Harrington and Weeden*

We found ourselves in a quandary: Broad (2003) quite rightly faults universalized, de-contextualized rubrics used nationwide for assessment projects, but our own community had a tradition of assessment using a rubric that had been drawn from our "local" course outcomes. Was this homegrown rubric subject to the same critiques? To add another wrinkle: the results of past assessment activity apparently had not been particularly interpretable by external audiences (aside from supporting assertions that "assessment activity has been taking place").

Boardman and her team (2000) had definitely used the assessment project's findings to improve teacher preparation and other program functions, and their rubric was a well-designed, comprehensive measurement tool for study of Core writing courses. Yet their findings had not been received and used beyond the program—how could we avoid this pitfall?

A third difficulty presented itself as the team studied the Boardman rubric in light of the Core director's mandate that both *writing* and *critical thinking* be examined. While this tool was extremely effective as a measure of writing, it did not specifically focus on *critical thinking*. When Diane Kelly-Riley visited UNR in spring 2005, she emphasized the need to develop operational definitions of this concept for each field or discipline, and described in some detail the process of articulating just what, exactly, a given community considered this intellectual activity or creative activity to be. The Core curriculum mission statements and other materials described "critical thinking" to be a key goal, but didn't really articulate measurable student outcomes by which the general education effort to teach critical thinking might be evaluated. Initially, the Core writing assessment team followed Boardman's team in understanding critical thinking to be manifested broadly but measurably as "critical reading" and "rhetorical awareness." Drawing on the experience of the Washington State University Critical Thinking Project (2002), the Core writing assessment team worked to create additional, more narrowly-specified locally-valid, contextually-sound, measurable definitions of habits of mind we could designate as "critical thinking." At the same time, the team worked to reconcile the local, highly-contextualized rubric used in past program assessments with Broad's (2003) more recent theoretical discussions about the limitations of rubrics.

After much discussion, deliberation, and design process, we arrived at what we considered a productive middle ground: in addition to providing a process for describing a community's values with regard to writing, *dynamic criteria mapping* (DCM) might provide a process for developing and validating the contextual soundness of any measurement tool; for purposes of our local effort, this process might result in a rubric that might allow easier "translation" of our community's criteria (what we value with regard to writing and critical thinking) for external audiences post-assessment. So,

we decided to use DCM to describe what we valued in writing and critical thinking, to map our criteria for that valuing, and to work toward a measurement tool (*even a rubric*, possibly) that we could use to talk with administrators and students about *what we really valued* in Core writing.

---

The challenges faced here by our UNR colleagues about how to use DCM, to what purposes, for what audiences, and at what moments resonate strongly with our own processes. As an active, collaborative, process-embracing practice, DCM helps give a rich shape to uncovering communally held values. As a rhetorical construct, the maps play an altogether different role. *Adler-Kassner and Estrem*

### WHAT WE REALLY DESIGNED

DCM had encouraged us to work from within. We wanted our assessment to be connected with the 2000 assessment conducted by Boardman and to reflect current values of our instructors. To access what our instructors valued in March 2005, we conducted an informal survey based on the 2000 assessment rubric features. The survey asked instructors to rate sixteen features on a scale of 1 to 10 (10 being highest). Instructors were also asked to describe what they value in student writing, how they recognize critical thinking in student writing, and what a successful paper for our English 102 course might be. The response rate was a depressing twenty percent. Even with the low participation, we took the responses and started to discuss which features from the 2000 study were still valued, which features could be combined, and which features needed to be added.

To get more detailed responses and to nurture a sense of inclusion, we decided to hold focus groups with instructors. Initially, we held two sessions in May of 2005. Twelve instructors participated in the two sessions, primarily graduate teaching assistants and part-time lecturers (who cover the majority of our core writing courses). These sessions were primarily designed to open up discussions about what our instructors valued in writing for our English 102 course.

One of our primary fears was that instructors would feel attacked by the assessment and resist participating in the process, so we tried to provide a space for instructors to discuss their perceptions of assessment. We used the focus groups as the opportunity to voice these concerns about assessment by having small groups of instructors create movie posters depicting visual representations of assessment.

To complete this project, members of each group had to discuss their perceptions of assessment and agree on the representation. We then let the other members of the focus group interpret the movie poster before allowing the designers to discuss their process. The movie posters ranged

from large brains in a high noon showdown to robotic monsters attacking a piece of writing. This activity was fun, funny, and allowed instructors to let go of their negative views of assessment and work toward creating an assessment process that they valued. After presenting the visual aspects of the posters and allowing participants to discuss what they disliked about "assessment," we moved the discussion toward what they valued in writing.

---

Jane and Maureen's movie poster activity provides a DCM heuristic for uncovering educators' values about writing. DCM-inspired activities are inherently visual, and they provide alternatives to traditional interviews, surveys, and questionnaires. DCM-driven methods, such as this movie poster activity, make visible aspects of rhetorical values that traditional research methods may not be able to reveal. In fact, Broad (personal communication, October 6, 2006) explained that UNR's innovative application of DCM reveals educators' "latent rhetorical values" using a "psychoanalytical approach." *Stalions*

> The participating instructors engaged in small-group discussions of what they valued in student writing, what they wanted from their students, and what they looked for when assessing student writing. As the discussions developed, a recorder tried to capture the essential features identified during the discussion. Each small group's list of features was discussed with the entire group to ensure accuracy and involvement. Following this discussion, participants were asked to review samples of student writing and identify what they valued and found problematic in each sample. Each participant read the samples silently, marking comments and writing notes. Small group discussions and then a large group share were used to open up discussions about values connected with instructors' assessments. A comparison between the features initially identified by the instructors, and those they had marked postitively or negatively in student samples revealed that evaluating writing created complications in our process. When the instructors were discussing the writing features they identified in the student samples, the discussion moved toward features that were easily identifiable. What participants could see in the student samples did not always align with the values of writing the group had initially identified. To bring all of the ideas together, the entire group generated another list of values associated with writing. There were many overlaps in the features, but it was important to us to capture the language that instructors were using, so all features were recorded using the language of individual instructors.
>
> From these lists and the discussion, participants were asked to design an assessment tool to evaluate student writing. The word "rubric" was purposefully avoided, to allow participants to think outside of that form, to

allow the possibility of dynamic criteria mapping to emerge. The guidelines were purposefully vague. Some instructors asked if this was meant to be a rubric. The basic response was that the instructors should design a tool that would help them assess student writing based on their personal lists of values. Some participants worked individually; others formed pairs or small groups. The assessment tools that participants created resembled traditional rubrics, though the lists were color-coded in neon or had elaborate groupings of features.

> The fact that faculty wanted the new instrument to be called a rubric, and even to look like a rubric, suggests how deeply rubrics are embedded in the culture of writing assessment. We saw this, too, at IUPUI. Moving from rubrics to something else can prove difficult. *Harrington and Weeden*

Other tools listed and coded features. Some assessment tools were extremely simple, listing only five to ten features. Participants created flow chart tools, descriptive paragraphs, and a cluster of star patterns with features written between arrows or along the lines of the star.

With all of the information gathered from the initial focus groups, the assessment team reviewed the posters, the lists of values, and the various assessment tools. From these documents and the responses to the survey, we developed a "rubric" draft. This initial draft had twelve features, which was a reduction from the 2000 survey's sixteen scored features and three comment features. Features were easily identified: many features reverberated throughout the process; however, the idea of visual form became a conversation. Part of the discussion about visual representation of the features was in response to the assessment team's discussions of the hierarchical structure of the 2000 rubric that seemed to privilege the initial features. The top-down structure of the 2000 rubric also seemed to leave little room for assessment readers wishing to to start any where other than the top and move through the features. Responding to various patterns of the focus groups' imaginative assessment tools and some of the assessment team's doodlings, the fairly-final draft rubric assumed a star shape: a group of numbered rays, one for each "feature area," linked at zero and radiating outward to the maximum score of six. (See figure 1.)

This star shape seemed to us deeply appropriate, since it reflected our community's sense that all the aspects of writing we were describing were integrally linked, inseparable, flowing together. The rubric also allowed a way to evaluate writing (in a shorthand way, to be sure) and generate numbers to translate our findings for external audiences. When a portfolio was scored using the chart, and lines drawn to connect the hatchmarks on each of the rays, we would have a visual representation suggesting the

"shape" of the whole (higher scores all around would make a "fuller" circle around the star; an area of lower scores would appear as a divot or flattening in that circle).

The star-shaped rubric presents a rubric/DCM hybrid assessment tool; this rubric incorporates features of a traditional rubric; however, this rubric becomes a full or misshapen circle depending upon how evaluators connect the hatchmarks in judging portfolios. In his introduction, Broad explained that "the rubricity or non-rubricity of the results was not of prime importance" in the DCM endeavors discussed in this book. *Stalions*

Finally, and perhaps most importantly, the star-shaped chart helped us to avoid privileging some aspects of writing over others. Unlike typical rating sheets, which list features to be scored from top to bottom and from one page to the next, the star allowed readers to see the whole rating system at once on one page, and to begin their evaluation wherever they wished on the star diagram. Somehow, the thing just fit the way we wanted to guide our process of evaluation for the project.

In the fall of 2005, we facilitated two additional focus groups to help us refine the assessment tool and to keep instructors involved in the development process. The first of these test-runs had participants applying the rubric to samples of student writing. We opened the focus group with instructors sharing their initial reactions to the rubric.

We were concerned about how instructors would react to the star shape that had emerged from our earlier processes. We began the focus group with a brief discussion of first impressions of the rubric, and then participants wrote about their responses to the rubric. Most of the responses were immediately positive—especially from group participants who had designed a star-shaped representation in the initial focus groups. One remarked, "I think that it is something that can be worked with rather easily and guide response," while another added, "Circular design allows some representation of values that are discipline-specific and those that cross disciplines." We did have a few participants who expressed concern about the design being too complicated. A participant wrote, "Initially, the diagram looked a bit confusing, but after explanation of its use, it appears quite simple, straight-forward" and another agreed, "Looks complicated at first glance, but makes more sense as I begin to understand how it will be used and applied." After this initial exercise, participants read student samples and scored two student papers using the rubric. Participants discussed their scores and comments in small groups. There was a lot of discussion in the small groups about overlaps in features and potential difficulties

assessing entire portfolios using the rubric. Small groups tended to focus on features that were either very apparent in the student writing or obviously absent.

As a larger group, we discussed general impressions and had discussions about overlapping feature descriptions that had led to varying interpretations of features. At this point in the process, participants discussed the importance of the feature that rated whether a student's writing fulfilled the requirements of the teacher's assignment (which we planned to include in the portfolios during scoring). Where the needs of a *program* assessment dictated solely a focus on the students' work as representing how well both teachers and students were working toward meeting stated course outcomes, these readers (as teachers themselves) felt it was important that they evaluate how well the writing answered an assignment (much as they would do in grading their own students' writing in a class). As a compromise position, the assessment team decided not to include the "answers assignment" feature as a scored item, but to offer space for discursive commentary on this aspect of portfolios on a "comment only" page.

The tool was revised based on participants' comments: some features were combined (specifically features that addressed focus and purpose), while other features were given fuller descriptions (features such as "problem and its complexities" and "rhetorical awareness" received more descriptors to help our readers recognize the features in student writing). In some cases, at this stage, we borrowed *names* for features (some of the critical thinking features, for example, were based on Condon and Kelly-Riley's rubric [2004], and some of the writing features were borrowed from Boardman's previous assessment project at UNR). These we carefully combined with feature *descriptions* from what teachers said in focus groups, making sure to use the language that was most identifiable to our instructors.

A final focus group session was held to dry-run the assessment reading planned for spring 2006. Timing issues and scoring variances were of particular interest to us for the planning of the official assessment. Participants applied the revised rubric to student portfolios. We had a discussion about usability of the rubric for portfolios, visual design, specific features, and general responses to the process. During this discussion, participants brought up many important considerations, such as the influence on portfolio evaluators of instructors' grading criteria as presented on assignment sheets (which we planned to remove from the portfolios to be scored). This final focus group helped us to narrow our features down to nine scored features and three comment-only features.

Participation in the focus groups was essential to the process of our assessment. There has been a true buy-in to the assessment, and to the rubric specifically, among focus group participants. Since the initial

presentation of the rubric to instructors, feedback has been extremely positive. Instructors in the final focus groups have even changed their own classroom/student assessments to be more reflective of their values and those listed on the rubric. Initial participation was basically induced through bribes of food and a small stipend to each participant; however, instructors then began to ask to participate to gain experience with assessment and to participate in discussions about evaluation of writing with other instructors.

---

UNR's inclusive, collaborative assessment endeavors produced ample opportunities for faculty to develop collegiality. Broad (2003) explained that criteria mapping "offers tremendous potential for writing instructors' professional growth and feeling of professional community" (121). *Stalions*

The initial focus groups were a starting point to introduce instructors to an assessment process that listened to instructors' values, incorporated their ideas into the assessment plan, and sought their feedback throughout the process. The focus groups were the foundation for creating our "rubric" or assessment tool. Wording for features and the descriptions for each feature were taken directly from focus group participants' feedback, to encourage instructors to identify with the features and with what each feature would look like in student writing.

Obviously, an assessment tool was created through the focus group process; however, the groups also offered our department and instructors time to look at their own processes and at how their pedagogical approaches fit into the department. Feedback from participants include comments such as: "I thought that this assessment focus group was most beneficial in how participants more clearly articulate the diverse values we bring with us to the classroom as an instructor;" and "I think this focus group experience will be beneficial to my teaching practices, in addition to being beneficial to the Core writing program. . . . It reminded me of the values I hold for writing, and how I need to improve my assignments and class discussions in order to meet those values and writing goals. . . . It also made me conscious of values and practices I held/hold but haven't noticed/don't notice."

The process that evolved was certainly inspired by DCM even if the final product is not the sort of criteria map Broad (2003) produced based on his study. It is the organic nature of DCM that we applied in our assessment design process. We have basically produced a non-traditional rubric, but this is what came from the instructors in the program. And they own the rubric. They connect with the star pattern and features. The rubric also allows us to take our outcomes to administrators in terms that they

can identify with (the numbers, of course, along with very condensed versions of our extensive descriptions of "what we value in writing") and translate to their audiences as well. Just as we could report to our external audiences the precise ways that the program appeared to be succeeding "by the numbers," so our external audiences could also point to how the program, a key part of the Core curriculum, appeared to be accomplishing some of the stated goals of UNR's general education effort.[5]

### The Rubric

The rubric is a nine-pointed star. Each axis represents one of the nine scored features, and there are six scores (1-6) marked on each axis.

**Fig 1. Nine-pointed Star**



Portfolio Code: _____

---
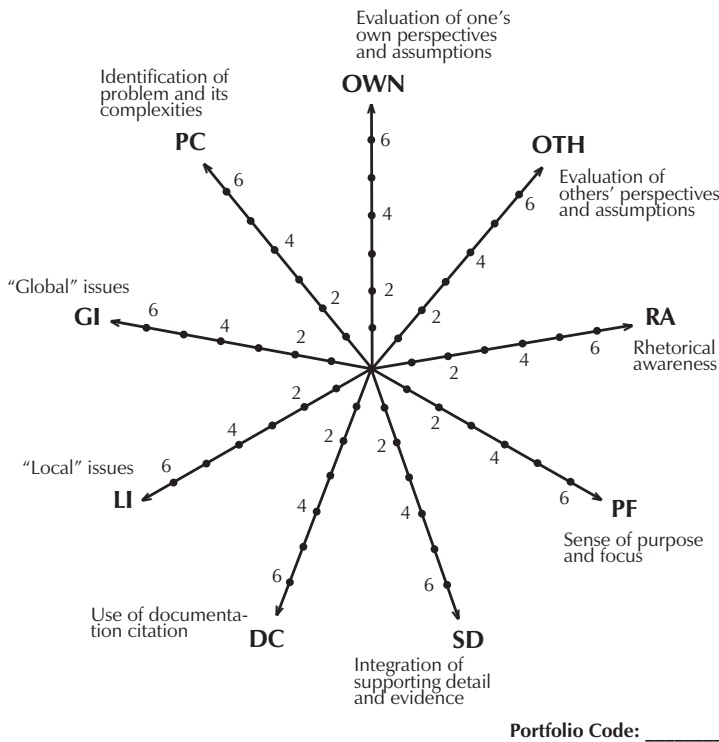
5.    As is noted above, the Core outcomes came after much of the design stage for the assessment project. Once these outcomes were articulated, they established that the Core writing program provides a crucial introduction to writing process and experience with conventions of various writing communities (Core Curriculum Outcome #1), as well as practice with research process and effective argumentation (Core Curriculum Outcome #2). With our assessment project, we demonstrated that the writing and critical thinking involved in these curricular objectives could be systematically described and measured, with statistically significant results. The team was invited to present our study at a regional assessment conference, at which we assisted other departments with beginning the process of describing what they valued in student writing and other work.

Readers marked a score for each of the nine features, with half-scores allowed between any scores except 3 and 4. In other words, a valid score might be a whole number such as 4, or a half number such as 4.5, but it could not be a 3.5. Readers were told not to use 3.5 in order to encourage them to make a judgment by choosing one of the middle scores on the continuum instead of opting for the exact halfway mark.

### The Comment Sheet

The comment sheet had spaces for three comment-only features[6], and three blank spaces for writing comments on scored features. Readers used the comment-only features areas to comment on issues they noticed but were not taking into account while scoring portfolios. Readers used the three blank spaces to comment on scored features. For example, a reader wanting to make notes about a writer's use of documentation and citation (DC) would write "DC" in the left-hand column and the comment in the right-hand column.

---

Both UNR and EMU have an interesting outcome from DCM: an approach to rating portfolios that is easy to use yet generates rich descriptive data. Both systems are wonderful illustrations of the creativity that emerges from the notion that a single score from a single rubric must be the summary of an assessment. Both programs encourage the use of numerical data or check-off sheets to provide some stability across readers, and both programs encourage the use of short, open-ended assessment spaces to allow readers to provide unique commentary on each portfolio. These systems are easy for readers to use, and they preserve the complexity of reading and judging writing. *Harrington and Weeden*

### The Feature Descriptions

To help readers use the rubric and the comment sheet, we also provided a scoring guide. This matrix contained descriptors generated by teachers as we designed the rubric (e.g., "avoids easy dichotomies" or "develops a line of thought"), combined with brief descriptors for the six-point rating scale (e.g., "4 = fully meets the requirements of the feature").

---

6.     The "Comment Only" section contained spaces for the following kinds of response: 1) Requirements of Assignment: Addresses assignment; form and format; 2) General comments regarding how assignment addresses requirements; 3) Overall Portfolio: Sense of the writer (i.e. experiments, plays. makes conscious choices, breaks with convention intentionally, shows engagement); overall impression of the portfolio and writing samples; general comments on your overall impression of the portfolio; 4) Anomaly/Outlier: Not applicable to the English 102 portfolio assessment; not enough evidence to draw any conclusions; general comments on why assignment(s) cannot be scored or does not seem applicable to assessment.

## WHAT HAPPENED: INTERPRETATIONS AND MORE INTERPRETATIONS

With our non-traditional, now-validated rubric in hand, in late spring 2006 we began ramping up for the actual portfolio readings in early summer. Our primary goals—since we already had selected our readers—were preparing the community for the logistical challenges of portfolio collection and the readers for the rigors of the actual evaluation process (slated to take place over five days just after semester's end).

We held two hour-long information sessions to further allay fears and explain, over and over, the details of the collection process. Teachers were encouraged, not required, to attend—and we made sure that all the information was conveyed multiple ways (over email, in hard copy, in person). In every case, we tried to reassure members of the community that this was indeed a *program assessment*, and not an evaluation of them as individual teachers.

As might be predicted, there were concerns logistical ("What do you do if the randomly-selected students dropped?"), practical ("What if the random selection only selected the students who were doing poorly?"), and protective ("How will you NOT know the student's name, or mine?"). The assessment team patiently explained that the selection process contained a healthy margin for attrition (choose five, need to net three from each section), and that, because the selection *was* random, it would necessarily mean that all students were equally likely to have their work included in the sample. Even more patiently, we detailed how portfolios and the attendant assignments would have identifying information removed (the instructors could even do it themselves, and just note the student's identification number) and a code number applied for tracking purposes during the actual portfolio evaluation readings.

Our patience and diligence were rewarded when, in early May, one hundred percent of our instructors submitted at least the minimum of three student portfolios. Only three out of thirty-nine instructors handed in portfolios after the deadline; of those, two had notified us in advance that the portfolios would be late. The portfolio assessment team took particular pride in this response, feeling that we had managed to reassure our colleagues that this was indeed a *program* assessment.

The actual readings proceeded very smoothly. Ten readers met for six days. We held norming sessions on the Thursday before the official reading week began, and on the first two mornings (Monday and Tuesday) during the reading week. Norming (training readers to evaluate consistently and according to the stated criteria) also took place on the Thursday of the reading week. For norming activities, readers were given copies of "spare" portfolios (these were complete, processed portfolios from each class

beyond the three "have-to-haves"). For each portfolio, the readers each did a separate assessment, then discussed their scores on each feature. The Thursday norming session also allowed the assessment team to do some final fine-tuning of the wording in the feature descriptions, while the discussions allowed readers to discover when their understanding of the feature descriptions was different from their peers' understanding.

Perhaps more important than the precise details of the norming sessions was our insight that readers' scoring became more divergent after the first two hours of reading, converged again after the lunch break for another couple of hours, then diverged more wildly as the afternoon wore on. Once we figured this out, the two-hour reading periods became the rule, and we strongly emphasized taking breaks and quitting soon after the day's second reading period.

According to standard portfolio evaluation procedure, the reading process ensured that each portfolio was read at least twice, with discrepancies of more than one point on any one feature (out of nine) prompting a third reading. With our norming, our "two-hour-insight," and, perhaps most important, a rubric that was contextually well-validated, we managed to achieve an inter-rater reliability of .77.

### RESULTS AND WHAT WE MAKE OF THEM: NUMBERS AND DESCRIPTIONS

Now that we had all kinds of numbers; what in the heck did they mean? Not being true measurement-types, we turned to a colleague from the Math and Statistics Department, Danelle Clark, for help in analyzing the data we had generated. With her able assistance, we tried to get a general sense of how students were doing *on average* in the writing and critical thinking activities demarcated by our key features. She determined whether the scores on each feature were normally distributed (they were), and whether the differences between the feature scores were indeed real differences (they were), and proceeded apace with other tests to check for statistically significant relationships between sample scores (and found some interesting correlations).

Careful consideration of the numbers on our key writing features suggested that the UNR Core writing program has been generally fairly successful.

---

In addition to qualitative studies, DCM facilitates quantitative analyses. Broad (2003), for instance, calculated criteria frequencies in addition to producing qualitative criteria maps. In addition, *Stalions* (2007) calculated the approximate frequencies of placement evaluators' criteria use, which were depicted in qualitative/quantitative criteria maps. *Stalions*

The readers assessed 192 portfolios, or fourteen percent of the 1,379 students in English 102, and sample mean scores ranged from a low of 3.6 to a high of 4.0, on a scale of 1-6 with 6 as the highest score.

We decided to pair the numbers with some careful, detailed descriptions of a range of portfolios, chosen for aggregate or "overall" average scores ranging from low to high on the six-point rating scale we had used for scoring the nine features.

We also did linear regression analysis, checking whether there were any statistically significant relationships between pairs of features. When these correlations were displayed in a matrix, we noted that there was a small cluster of pairs that seemed to be highly correlated (above .75, or closest to 1.00). Making a command decision, we decided that these were statistically significant correlations—and the more scatter-plotted, less-well-correlated pairs were not.

With all the numbers and some useful analyses in hand, we prepared to argue that the results of our study suggested that most English 102 students were adequately competent or more than competent in the kinds of writing and critical thinking activities that the assessment measured. As compared with previous Core writing assessments, the assessment team was able to evaluate more, and more specified, domains of critical thinking, and fewer, less specified aspects of writing, with the newly-designed rubric.

Since statistical tests determined that the various features were normally distributed, we felt that the findings warranted some cautious claims about how current students were doing or similar students would do "on average." To make our case, we drew together the results of both the scoring and the statistical analysis, and made some specific observations.

We decided to make a series of points, using the scores. As soon as our number-laden charts were complete, instructors were invited to come and offer comment and interpretation before we took our assessment show on the road (a few buildings over, for a command performance at a meeting of the Core board). The assessment team facilitated a lively discussion of the data, took notes on the commentary, and wove the community's interpretations into a final report.

## WHAT DID IT MEAN? SHAPING A REPORT

When we began to prepare our report, we anticipated the needs of various audiences.

DCM can feel unwieldy, and overly focused on process, input, and conversation. However, as Detweiler and McBride report here, the result of this kind of careful process is that assessment takes on a life of its own: it feels truly collaborative. The lively discussions with different groups of interested parties (or stakeholders, as we call them) are one of many significant benefits. *Adler-Kassner and Estrem*

The teachers in our community might want more details about the process of the project, or the actual numbers on the various features. Some other readers, like local secondary teachers and students, might be studying our website for insights into what we teach (and value) at UNR. Certainly, the director of the Core curriculum had indicated an interest in a careful descrip-

---

Jane and Maureen anticipated the needs of various interested audiences--teachers, students, scholars, professionals, and external stakeholders. They anticipated and constructed what Ruth Ray (1993) coined "local knowledge,…knowledge for the community of teachers within one's department, school, community, district, or state" (86), and "global knowledge,…knowledge…for the field of composition" (90). *Stalions*

tion of the validation process, the results, and the implications we drew from our findings. He wanted the report to be a model for other disciplines (no pressure there) that might also want to adopt a portfolio evaluation process for their assessment efforts. The report has since been forwarded "up the chain," containing detailed discussion of process and results, and has provided information useful in the accreditation process, as well as in ongoing discussions of curriculum changes proposed in response to the growing population of incoming students who need developmental writing courses.

In the main report, we offered an interpretation of the numbers and an evaluative description of several portfolios, to lend a sense of what the numbers "might look like" for the readers. We then developed different versions, some of which would be made available on the program website for local teachers, students, and other interested parties. In preparing the report, we also generated other versions of the information that would be potentially publishable in the profession's journals and in collections like this one.[7]

Since Kathy Boardman and her 2000/2001 assessment team had done extensive linking of substantial reader comments with specific scores, our assessment team did not necessarily need to do this kind of qualitative documentation again. Further, our mandate was to study writing *and* critical thinking; hence, we focused on developing and validating constructs for this kind of assessment. Without Boardman's substantial qualitative data, however, the team felt the need to be especially careful about making

---

7.    We decided that we would be very careful in circulating details of the rubric and scoring sheet, since we are committed to contextually-valid assessment. To encourage others to engage in developing their own assessment tools—and to discourage simple transfer and application of our rubric—we will make the rubric and feature descriptions available in carefully limited forms, and lead discussions for other groups around campus and elsewhere. In this way, we will emphasize the need for locally-developed assessment tools, and assist others in this development process.

overbroad and ungrounded statements about what any particular student would be able to do. With the quantitative analysis, our 2006 study gained in explanatory power, and lost the greater nuance and descriptive-interpretive depth of the 2000/2001 study.

To address this problem, Maureen selected portfolios representing the full range of scores, to show how well a given portfolio demonstrated the overall score range (low, mid, high) for all nine features. The portfolios that were selected did not have any discrepancies in scoring (a third reader was not needed). She read through each of the portfolios looking for features that were exemplified by the writing samples; for example, the feature for examining one's own beliefs was easily identifiable in the high-scoring portfolio. After reading through the example portfolios, comparisons were made between the score ranges. In addition to looking at individual features, the assignments and sequencing of assignments, especially for the high-scoring portfolio, were discussed.

## CONTINUING THE CONVERSATION: IMPLICATIONS FOR THE PROGRAM AND FUTURE ASSESSMENT

After the completion of the report for our administrative audiences, we developed a range of continuing projects to carry on the cycle of interpretation, evaluation, and reflective programmatic change. The most immediate projects had to do with sharing our results and revising teacher preparation in several ways.

We revised the summer orientation and the teaching practicum to reflect what we found in the course of our assessment project. Specific results and our interpretations of these became a session on "teaching critical thinking," with a special emphasis on how our programmatic focus on rhetorical awareness might be more explicitly tied to helping students demonstrate this intellectual practice. Various other sessions were similarly adapted to take advantage of patterns we discerned in the assessment data.

We held meetings in which teachers were invited to help us figure out what the numbers might mean (see the sections above for examples of how our colleagues helped us understand the numbers). These responses were folded into the final report, and into a range of other program materials.

We involved our instructors in revising teacher resources. For example, we asked for volunteers for a working group to look at examples of assignments from high-scoring (4.5 to 5.5 range overall) portfolios. We chose only high-scoring portfolios to ensure that the assessment did not become a critique of teaching. We had six volunteers meet for a Saturday with only a small bribe of homemade snacks and potential cv lines. Instructors read through six sets of assignments, making notes, commenting on similarities between assignment sequences. The findings of this working group have become a handout on

"features of assignments and sequences that invited high-scoring responses," which has been shared with current teachers and will be crucial to preparing new teachers in the program. The discussion about the assignments led many instructors to talk about their own teaching practices, sequencing, and values for writing. In their written sign-offs, instructors said, "I found this session to be extremely helpful because it exposed me to a variety of successful assignment strategies. I think it would be helpful for instructors to see these sequences as models upon which they might base their courses, or as inspiration for designing their own sequences."

We plan to design and conduct a study of how well English 101 articulates with English 102. While this prerequisite course has its own curricular goals, most of them point toward the required, culminating course in the sequence. Since the program seemed to have been pretty successful in teaching "rhetorical awareness"—judging by the scores on portfolios in our sample—we also might be able to emphasize critical thinking more heavily in English 101, and better prepare students for English102. A follow-up study to see how well our articulation efforts are working is certainly in order.

We also plan to share the results of our study, as well as our process of designing and conducting it, in various forums, both on campus (at a regional assessment conference) and beyond (at the Conference on College Composition and Communication and in publications).

With these changes in teacher preparation and our various assessment working groups, we have begun to "close the loop." We are beginning to bring the assessment back to the instructors and into the classroom.

---

By closing the loop between writing program administrative and pedagogical matters, Jane and Maureen's UNR assessment model invokes what *Alford* (MMCC) calls the "fractal" nature of organic writing assessments; in other words, such an assessment model should operate at both a macro and micro level without losing its essence. *Stalions*

It is our goal to continue the conversations, allaying assessment concerns as we go, and enriching our understanding of *what we really value* in, and as, "writing" and "critical thinking" at UNR.