

California State University, San Bernardino

CSUSB ScholarWorks

Theses Digitization Project

John M. Pfau Library

1999

Large-scale portfolio assessment: Pitfalls and pathways

Candace June Gilbert

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/etd-project>



Part of the [Rhetoric and Composition Commons](#)

Recommended Citation

Gilbert, Candace June, "Large-scale portfolio assessment: Pitfalls and pathways" (1999). *Theses Digitization Project*. 1524.

<https://scholarworks.lib.csusb.edu/etd-project/1524>

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

LARGE-SCALE PORTFOLIO ASSESSMENT: PITFALLS AND PATHWAYS

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
English Composition

by
Candace June Gilbert
June 1999

LARGE-SCALE PORTFOLIO ASSESSMENT: PITFALLS AND PATHWAYS

A Thesis

Presented to the

Faculty of

California State University,


San Bernardino

by

Candace June Gilbert


June 1999

Approved by:



Edward M. White, Chair, English

5/13/99
Date



Carol Haviland



Kellie Rayburn

ABSTRACT

Traditional methods of writing assessment, such as multiple choice and essay exams, seem to have little connection to classrooms where writing is taught as part of an ongoing process. Because of this, teachers often feel that institutional tests are not aligned with what they are actually teaching. Portfolio assessment, if properly implemented, is a way to align institutional assessment with actual teaching methods. However, before a plan for using portfolios can be viable, issues of validity, reliability, feasibility and accountability must be examined and resolved. These issues, along with the need for faculty and institutional "buy in," must all come together if a portfolio assessment plan is to work.

The experiences of the school district discussed in this thesis points to these issues. Determined to use portfolio assessment as a means to satisfy state requirements for accountability, this district formed "Stepping Stones," a group of teachers and administrators working together to produce a district-wide assessment plan. Stepping Stones participants, after two years of discussing, researching, inventing, revising, and collaborating, formulated a tentative portfolio plan for their K-12 school district, a plan which at the beginning of the 98/99 school year had yet to be implemented. The inception of the Stepping Stones plan, from beginning to end, serves as an

enlightening lesson on the trials and tribulations that occur when a group of well-meaning educators and administrators attempt to create a tenable portfolio assessment. Thus, the Stepping Stones "journey" is a touchstone for examining many of the possibilities and problems that occur with large scale portfolio assessment.

The problems with developing an authentic assessment through portfolios, which, by definition, include more than one sample of student writing, are many. When assessing a large number of students, the questions of storage, cost, assessment techniques, and common structure must be answered. Along with these questions, issues of reliability and validity must also be answered.

For an assessment to be valid, it must align with the definition of the construct being assessed. If writing is defined as that which is developed over time through the writing process, then portfolios, unlike multiple choice or essay exams, can be a valid method for assessing writing ability. Although there are many methods for determining the validity of an assessment (predictive, concurrent, face, content, construct), there is one type of validity that is often neglected: consequential validity.

Positive consequences can be gained when teachers gather together to develop portfolio criteria that have enough commonality to be useful for large scale assessment, yet grow out of individual classroom lessons. Inevitably,

these brainstorming sessions include discussions of one another's lessons, beliefs about writing, and methods for assessing writing. These discussions permit a community of teachers to learn from one another, learning that can be extended to students as they make decisions concerning their writings and their portfolios. Portfolio assessment, thus, can have positive consequences for the student/teacher learning connection. Consequently, while there are many valid reasons for using portfolios to assess writing, creating a portfolio assessment that is both valid and *reliable* can be problematic.

For a writing assessment to be reliable, there must be consistency in results. Without both validity and reliability, an assessment is meaningless. While this makes sense, not all writing experts agree with this conclusion. Some proponents of writing portfolios argue that the importance of validity overrides the need for reliability. This argument stems from the fact that establishing reliability, particularly "scoring" reliability, is especially difficult when facing the magnitude of a large-scale portfolio assessment. With careful training and guidelines, reliability is possible to achieve; still the question remains whether or not portfolios are always the best choice for a writing assessment.

If issues of reliability and validity can be resolved, participants must then decide whether or not the knowledge

gained merits the extra cost and time in pursuing a large-scale portfolio assessment. For an assessment plan to be effective, it must have institutional support from both the administration and the faculty. If the participants decide that the information gleaned from portfolio assessment does not merit the extra cost and time, it is unlikely that the assessment will have the necessary support.

In sum, portfolio assessment is sometimes worth the effort, and sometimes not. The purpose for each assessment must be carefully examined before the decision is made to use portfolios. In the end, if an institution decides to attempt large-scale portfolio assessment, no matter the outcome, the experience itself, "pitfalls and pathways" included, can prove to be a worthwhile learning experience for all involved.

ACKNOWLEDGMENTS

Having started writing my thesis five years ago, it has become a strangely significant part of my life, and, by proxy, my thesis advisors' and my family's lives as well. Five years is a long time to listen to me talk about portfolio assessment; yet, I can't recall anyone but me complaining. Lucky for me, I had wonderfully patient and kind thesis advisors. Kellie Rayburn offered me honest criticism; her suggestions helped me focus and unify my thesis. Carol Haviland stressed the importance of writing and writing theory; her revision and editing comments were invaluable. Edward White, mentor extraordinaire, started me on the path of writing assessment; his never failing dedication, support, and enthusiasm kept me going and led me to the finish.

As part of an extended family, I may have written the thesis, but it took every one of us to produce the final product. For all the babysitting, support, and encouragement, without which, I would never have made it to this point, I thank my parents, Robert and June Fingler, my grandmother, Mary Crawford, my sister, Connie Mugford, and her family, Dale, Ken, and Robert Mugford. For being the kind of kids who never made me feel as if my education was getting in the way of their happiness, I thank Rich and Breanne. For whisking me away from thesis writing to the world of romance, I thank my suitor, Greg. For joining our

family and returning me to thesis writing, thus becoming partner, collaborator, editor, and computer expert, I thank my husband, Greg.

A family is a wonderful thing.

A finished thesis is pretty good too.

TABLE OF CONTENTS

ABSTRACT:.....	iii
ACKNOWLEDGMENTS:.....	vii
INTRODUCTION:.....	1
CHAPTER 1: STEPPING STONES: ONE JOURNEY.....	5
CHAPTER 2: PORTFOLIOS: THE ROAD TO AUTHENTICITY.....	20
CHAPTER 3: VALIDITY: THE ENLIGHTENED PATH.....	30
CHAPTER 4: CONSEQUENTIAL VALIDITY: DETOUR TO SUCCESS..	47
CHAPTER 5: RELIABILITY: THE LONG AND WINDING ROAD.....	61
CHAPTER 6: FEASIBILITY: OVERPASSES AND TUNNELS, OR "GETTING THERE".....	80
CONCLUSION: THE LIGHT AT THE END.....	87
WORKS CITED:.....	90

INTRODUCTION: Mapping the Journey

Methods for accurately assessing writing among large groups of students are not easily found. Multiple choice exams, the most widely used method of assessment, do not assess actual writing. Essay exams do assess writing, but it is writing produced in one sitting, a method that clashes with the commonly held belief that writing is part of an ongoing process. Writing portfolios are one way to satisfy the desire to assess writing as a process, and, according to Edward M. White, "Any large-scale effort to develop a new model of writing assessment ought to devote itself to ways of using portfolios as a major measurement device" (*Assessing* 177). But the road to developing a successful portfolio assessment is rarely straight and narrow. Like most journeys, the experiences gleaned along the way transform the goal of a destination into a culminating event, preferably, one in which colleagues collaborate around issues of authentic assessment and students engage their individual writing processes and, thus, become active participants in their own writing assessment. The best reason for any kind of assessment is that students learn to assess their own skills. Portfolio assessment offers this opportunity.

The following thesis is devoted to an examination of the writing portfolio as a major measurement device for writing assessment. Chapter one offers a narrative of one

K-12 district's effort to develop a large-scale portfolio assessment plan that would satisfy state requirements for multiple assessment; be developed, accepted, and used by teacher practitioners; and be a positive learning device for students. The district's portfolio would also align with state standards that require that students learn to employ a writing process that includes drafting, revision, and editing within all content areas. The "Stepping Stones" project is an example of a well-meaning group of educators and the problems they encounter as they attempt to revamp and reinvent an outdated and unsuccessful portfolio plan, an example that will serve as an introduction to many of the "pitfalls and pathways," that any group turning to portfolios as an alternative to traditional methods of large scale writing assessment will encounter. Each chapter will revisit Stepping Stones in light of the issues encountered and discussed.

Chapter Two looks at portfolios and authentic assessment. The need for accountability in writing assessment often seems at odds with a teacher's need for autonomy in a classroom. Portfolio assessment offers a way to align accountability with autonomy. This alignment, however, is not without problems. A large scale assessment, by definition, involves a large student population. The time, money and space involved in managing this kind of massive assessment disallows the kind of depth that is

desirable in any kind of assessment. The chapter also examines authentic assessment in relation to the pros and cons of essay tests versus portfolio assessment and the possibility of negative consequences for teaching when teaching and assessment are at odds.

Chapter Three explores validity (predictive, concurrent, face, content, construct) and its role in writing assessment. When an assessment is valid, it *honestly* measures what it purports to measure. In other words, the assessment aligns with the construct being measured. When writing is defined as a process, portfolio assessment, with the proper considerations, can be considered a valid assessment of writing. The chapter concludes with an introduction to consequential validity and its role in portfolio assessment.

Chapter Four focuses on consequential validity in relation to teachers and students. Teachers involved in portfolio assessment have the opportunity to develop a high level of collegiality with other teaching professionals; this, in turn, provides opportunities for teachers to learn from one another's teaching practices. Students involved in portfolio assessment learn to be metacognitive about their writing processes and their learning styles. An assessment that is consequentially valid will be an extension of the student/teacher learning experience. Portfolios when used for assessment purposes, can be such an extension.

Chapter Five defines and discusses issues of reliability in writing assessment, particularly, *scoring* reliability. Included in this discussion is Peter Elbow's and Edward M White's disagreement over the importance of reliability and portfolio assessment. Also included are two case studies that look at scoring reliability and portfolio assessment. This chapter concludes with a consideration of the importance of purpose and cost-effectiveness in assessment.

Chapter Six examines cost effectiveness and institutional validity. For an assessment to be feasible, the results, along with the amount of money and time devoted to achieving them, should be worth the cost. Portfolio assessment may not always be the best choice for an assessment, as Peter Elbow points out in this chapter. For an assessment plan to work, it must have support from both the administration and the teachers. This is institutional validity. No matter how valuable an assessment plan is, if it does not have institutional validity, it will not work.

The conclusion looks at the light at the end of the tunnel, the positive impacts of the entire process of implementing a portfolio assessment plan. Even if the plan does not work exactly as expected, the benefits gleaned along the journey are worth the trip.

CHAPTER I

STEPPING-STONES: One Journey

In March, 1996, Morongo Unified School District board member Sue Meader and MUSD Assistant Superintendent Beverly Willard worked with Greg Gilbert, a local community college faculty member, to bring the Inland Area Writing Project's Stepping Stones Institute from the University of California, Riverside to the Morongo Basin. MUSD's area of concern involved a federal mandate that requires districts to maintain records of student assessment in all content areas. Discovering a way to assess writing ability at all grade levels and to keep accurate records of this assessment was to be Stepping Stones challenge. John Trimbur, an editor of the MLA series, *Research and Scholarship in Composition*, provides a rationale for this type of challenge. Trimbur says

that there is growing pressure from university administrators, state legislatures, and national panels of experts for accountability, for hard data to convince the public that writing instruction merits the resources devoted to it. This line of thought holds that unless we devise ways to assess writing, someone else will do it for us. (46)

Administrators in K-12 districts are under the same pressure for accountability. Willard's decision to use Stepping Stones and teachers from MUSD as a means to discover ways to satisfy part of the state's federal mandate encompassed not only the desire to produce "hard data" on MUSD students'

writing abilities, but also the desire to keep the assessment within the district, rather than "letting someone else do it for us." Stepping Stones participants were to develop a portfolio assessment that could produce appropriate data and, at the same time, be acceptable to teachers.

The Stepping Stones project evolved through three phases, Stepping Stones I, II, and III, and involved more than thirty teachers in approximately 75 hours of work over a period of 26 months. The first two weekends were theory based workshops led by Edward M. White, one of the nation's leading authorities on writing assessment, and Carol P. Haviland, Director of the Writing Center at California State University, San Bernardino. The second phase of Stepping Stones involved five weekend meetings dedicated to the formulation of the portfolio plan, and the third phase devoted three additional weekends toward reviewing and revising the project's recommendations to the district.

The Stepping Stones Project, whose members met for the last time during the 98/99 school year, reflects many of the trials and tribulations which occur when a group of well-meaning educators and administrators attempt to create and implement a writing assessment which conforms to current writing theory; is valuable to students, manageable for teachers, accepted by administrators; and succeeds in fulfilling both state and federal mandates for

accountability. The following pages concerning Stepping Stones inception will serve as an introduction to some of the issues--validity, reliability, feasibility (in relation to costs and time), and accountability--which all must come together along with faculty and institutional "buy-in"--in order for large-scale portfolio assessment to be a viable alternative to such traditional assessment methods as multiple-choice and impromptu essay exams. These issues will be examined in more detail, and in light of other portfolio projects, in the chapters which follow this discussion of Stepping Stones.

Stepping Stones I (SSI) originally came about because Greg Gilbert and Sydney Tibbetts (teacher, Victor Valley High School) desired to establish a more unified continuum between school districts, colleges, and universities, with regard to the instruction and assessment of writing. Gilbert approached Beverly Willard with the idea of an institute for teachers in MUSD. Willard agreed and proposed that the seminar focus on MUSD's existing portfolio so that it could be used as a viable assessment tool. Funding for Stepping Stones was raised through a combination of efforts on the part of Gilbert's employer, Copper Mountain College, and the MUSD. SSI was attended by one high school teacher, two middle school teachers, nine elementary school teachers, Willard, and a local town manager, along with facilitators Gilbert and Tibbetts. Dr. Edward M. White spent the first

weekend discussing connections between writing and assessment. His lectures, discussions, and activities were based on theory, practice, and the need for balance between issues of validity and reliability. Because of the undeveloped potential of MUSD's existing portfolio, his instruction provided an important foundation upon which the rest of the project could build.

All elementary teachers and secondary English teachers in MUSD are required to keep writing portfolios for each of their students. Originally teachers were to provide a sample of each of the three styles of writing required for their particular grade level, based on existing content standards. This portfolio was then to be handed on, at the beginning of the next school year, to the student's new teacher. This storage and transfer system soon proved to be too cumbersome and work intensive for teachers who made their complaints known both by word and by their general refusal to comply. Requirements for portfolios were then toned down and changed for each grade level. For example, seventh grade teachers were requested to place in the portfolio the autobiographical essays that their students had written according to current content standards. Eighth grade teachers would, theoretically, receive these portfolios, complete with essays, and attach district approved rubrics completed by the seventh grade teacher, and they would then have their eighth grade students build on

these autobiographical essays which, upon the student's completion of the eighth grade, would be sent along to the high school for the beginning of ninth grade.

While well intended, this sequence of events set off a chain of problems. Seventh grade teachers received their students' sixth grade portfolios from multiple elementary sites. The portfolios contained a variety of student writing, anywhere from one to ten pieces, depending on the decisions of the particular elementary sites. It was obvious, in many cases, that the students had worked hard on these portfolios, and many seventh grade teachers did take the time to sort through them in an effort to discover useful information about their students' writing abilities. Yet without specific district guidelines for continuity between site portfolios, there was no way to accurately compare or assess the portfolios. In addition, further down the line, not all of the eighth grade teachers chose to participate in extending the autobiographical essays which had been written by their students during the previous year. Even further down the line, when high school teachers were questioned about the portfolios they received from the middle school, they *all* said that they had not looked at them. At present, the portfolios are sitting in a high school storage room. The original MUSD portfolio project, lacking focus and purpose, was destined to fail. Nevertheless, according to a survey of MUSD teachers, while

the frustration level was high, most teachers still believed that the portfolio project was worthwhile and most were interested in continuing. However, they were only interested in a system that could avoid needless bureaucracy and be of direct benefit to their students. The MUSD survey reflects Whites' statement that "[teachers] are perfectly ready to adopt assessment . . . when they are convinced that it will enhance student writing and support their teaching" ("Power" 13). The best way to provide teachers with the kind of assessment they are looking for is to involve teachers in the development of the assessment. Mary H. Sawyer, in discussing one successful portfolio project directed by the late Alan Purves, professor of Education and Humanities, and developed by graduate students and grade 6-12 teachers, says that an "important component of the project's success was that teachers . . . were not research 'subjects,' nor were they implementors of any pre-designed portfolio system" (66). The teachers were given the opportunity to produce an assessment based on their classroom experiences. Sawyers believes that this sent "the message that their research and what they were doing in their classrooms was interesting to other researchers, teachers and administrators . . ." (66). Willard's decision to involve teachers in all aspects of the Stepping Stones project sent the same message. Stepping Stones was to be the vehicle for uniting teachers in their effort to salvage

MUSD's portfolio project.

During the first weekend of Stepping Stones, Dr. White's presentation on such issues as development of writing prompts, focus on criteria, holistic scoring of both essays and portfolios, importance of validity and reliability in large scale assessment, and other important topics (which will be discussed at length throughout this thesis), sparked notes of enthusiasm and caution among the group and also brought forth one of the main problems which was later to plague SSII, and that is the problem of focus. With so many issues at hand, how was a small group, meeting for a relatively short period of time, to accomplish such an enormous task?

The second weekend of SSI, focused, with the help of Dr. Carol Haviland, on practitioner development of connections between writing assignments and the assessment of writing. Dr. Haviland had institute members divide into groups which worked collaboratively in developing writing prompts that were later discussed in terms of their value and effectiveness. With Haviland's guidance, the members spent the day discussing and debating various problems and solutions concerning portfolio assessment. This presentation helped the members to see more clearly some of the uncertainties that arise when assessing writing and how these uncertainties become increasingly problematic with large scale assessment. Stepping Stones I provided a

necessary forum for valuable information gathering; however, two weekends was clearly not enough time to formulate, focus, and effectively begin the process of revising MUSD's portfolio system. It was decided at this time that the members of SSI would try to sort out data in order to develop a more cohesive plan for development, and this is what the group did at their last meeting. What came out of SSI was a proposal to MUSD's board that the board provide funding, by way of a one time payment to facilitator Greg Gilbert and stipends for participating site teachers, for Stepping Stones II, which would take place over the '96-'97 school term for a period of four meetings and one practice scoring session. In return for this support, SSII members promised to come to the board's closing meeting in June of 1997 with a portfolio project proposal. The school board's approval for funding was also an implicit approval of the teachers' work in Stepping Stones I. This kind of encouragement exemplifies what Sawyer sees as "critical to the success of [Purves'] project" (67). Sawyer explains, "The districts all chose to invest in their own teachers rather than spend district money on outside experts or publishers' pre-packaged portfolio systems" (67). MUSD's investment in teachers provided the impetus and motivation for Stepping Stones II.

Of singular importance to Stepping Stones II was that the MUSD Writing Portfolio provide the district with some

kind of form that would satisfy federal funding conditions which require proof of multiple assessment within all content areas. Aside from satisfying the district's needs, Stepping Stones was guided by one primary principle: to create a portfolio plan that allows for a maximum of site autonomy while imposing a minimum of constraints on classroom practices and teacher schedules. The ideal portfolio would be one that would adapt to an individual teacher's curriculum. According to White, "when teachers are forced--as they often are--to choose between teaching to an inappropriate institutional test and helping their students learn to write, they are bound to consider evaluation an intrusion into the classroom" ("Power" 12). Stepping Stones II participants concurred with White's statement, and they were determined to de-institutionalize--as much as possible--the MUSD portfolio.

Clearly a concise plan of development was needed in order to achieve qualitative results. The members of the group, with the help of their facilitator, Greg Gilbert, began to outline a collaborative picture of the ideal portfolio. It soon became apparent that the following portfolio requirements were non-negotiable: the required portfolio contents should not guide a teacher's classroom, but, rather, the contents should flow naturally out of a teacher's normal curriculum; the paperwork involved in assembling the portfolio should not be too time consuming;

teachers should receive recompense, in the form of in-service days, for the time they spend assessing portfolios. From these few agreed upon requirements, the members put together a survey which was distributed to teachers throughout the district, and which, when returned, tended to support that these requirements were ones shared by many MUSD teachers. With this information, along with a variety of portfolio systems and sample rubrics (including the "New Standards" portfolio funded by the federal government, which includes all content areas, but which we judged as too complex and time consuming), including the rubric developed for the current MUSD portfolio project, SSII members formed collaborative groups and went to work developing the revised portfolio plan.

SSII participants came to the next meeting armed with ideas which they presented and debated. From these presentations, a tentative portfolio plan was developed. SSII spent the remainder of their meetings revising and editing the plan for presentation. At the last gathering of the SSII, members practiced grading and scoring portfolios, and thereby increased their understanding concerning the important link that should exist between assignment and assessment criteria. At SSII's conclusion, the teachers had finalized a plan in which they took genuine pride. Greg Gilbert and Beverly Willard presented the revised plan at the MUSD board meeting in June, and it was approved to be

piloted in a few schools during the '97-'98 school year.

Prior to the beginning of the 97 - 98 school year, Assistant Superintendent Beverly Willard retired. Her successor, Dale Mitchell, decided to postpone implementation of the pilot program until he could meet with Stepping Stones participants and discuss it with site principals throughout the district. In the meantime, Mitchell acquainted himself with assessment issues district-wide, including the use of the existing district writing portfolio. In March, 1998, he gave the go-ahead for Stepping Stones III and invited all interested teachers to meet so that he could be part of a decision that would follow regarding the use of writing and portfolio assessment in the district. SS III involved twenty teachers, all of whom had participated in SSI and/or SSII, and while the first meeting entailed a review of the previous two phases, teachers were quick to reacquaint themselves with the issues, and, perhaps most importantly, to express a willingness to reconsider the portfolio plan they had completed at the conclusion of the previous academic year. Clearly, Mitchell's focus was accountability, and because he demonstrated a broad grasp of the issues and an earnest commitment to reliable and authentic assessment within MUSD, his leadership was happily endorsed by SSIII participants.

Though development of a portfolio plan should occur from the ground up, solid and authoritative support from the

top is vital. While the former superintendent, Mrs. Willard, endorsed the efforts of Stepping Stones, Mr. Mitchell's involvement was hands-on at all levels, and he made it clear that accountability was important, that he had to report to the state that the district was meeting standards. Mitchell's solid direction was just the thing to help create focus for Stepping Stones III.

Mitchell and the Stepping Stones institute agreed that using portfolios is a way to get away from complete objectivity, or the "multiple choice syndrome" of writing assessment, and to get, at least, a little closer to the actual assessment of writing. Along with objective tests and classroom grades, portfolio assessment could be used as a way to account for and to assess student performance. The next step was deciding on the minimal content requirements for the district-wide portfolio.

While a substantive number of suggestions were offered as to portfolio contents, Stepping Stones III had to pare down their choices to a select minimum in order to accommodate all teachers and all grade levels, K - 12, in the district. After much debate, the following was chosen:

- 1) one selection of one-shot writing, an impromptu response to a prompt;
- 2) writing process alone, student revision and editing without external assistance;
- 3) writing process collaborative, revision and editing

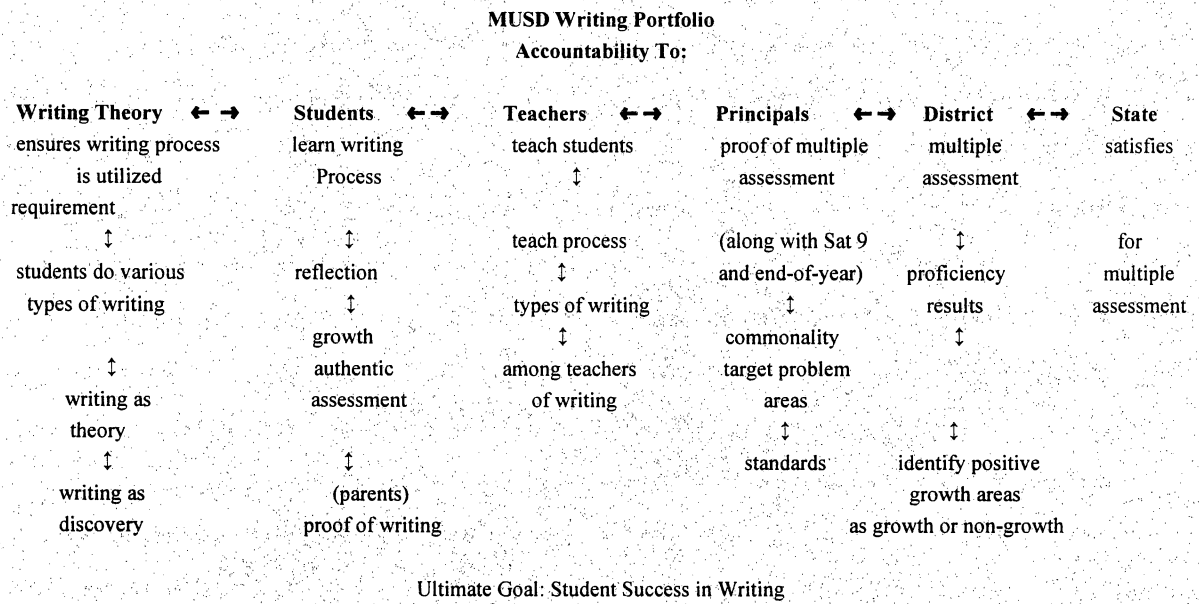
workshops

Stepping Stones III determined a number of optional entries, both recommended and suggested for sites to add to their portfolios. The two recommended entries were: 1) student reflection and 2) writing across the curriculum. Other suggestions included: 1) the original assignment description attached to the completed assignment, 2) time capsules (suggested for k-3, could include writing and/or drawing samples, to be opened upon the completion of 6th grade for sentimental purposes), 3) rubrics (charts breaking down the various parts of the assignment, along with the points possible for each part, 4) examples of various writing styles (letters, modes). Also discussed were instructor assessment forms and forms that report to the district on student proficiency.

While Stepping Stones final recommendations may appear modest, it is worth noting that participants had considered the federal government's "New Standards" portfolio, portfolio plans from various school districts, and portfolio systems as detailed in a wide range of articles. In the final analysis, based on group discussions and more than a hundred teacher surveys within MUSD, it was decided that a smaller *imposition* might allow for greater cooperation on the part of teachers and individual sites. In as much as the participants at Stepping Stones had come to appreciate issues of process, cross-curricular writing, and authentic

assessment, they were willing to believe that the same discoveries would be appreciated by their colleagues as a culture of assessment accountability evolved within the district. By starting small, SSIII made an investment of faith in their fellow teachers.

One of the participants of Stepping Stones composed the following informal flow chart to show the recursive direction of accountability. At the bottom of the chart, the stated "Ultimate Goal: Student Success in Writing" represented a view point with which all of the Stepping Stones participants agreed enthusiastically.



Stepping Stones teachers remained optimistic over the possibility of including authentic assessment through portfolios as one of the multiple assessments accepted by the State of California. All participants agreed that no

matter the outcome, the opportunity to meet with other teachers and talk about writing and writing assessment was valuable and more than worth all of their time and effort.

The Stepping Stones portfolio project thus serves as an introduction to the confusion, controversy, and gratification that occurs when trying to create a portfolio system for large scale writing assessment. All of the problems and solutions discussed are those which happened in the *planning* stages of this project. The implementation of this assessment plan itself has not even begun. When MUSD decides to pilot the project, the Stepping Stones system will be put to the test. It will be interesting to see whether or not this plan succeeds and whether or not those involved will find that the benefits outweigh the problems. In the chapters ahead, I will discuss portfolio assessment in relation to validity, the problems with achieving reliability, the benefits and consequences of using portfolios as a measurement device, the problems with cost and time in relation to portfolio use, whether or not portfolios are *always* the best tool for assessing writing, and how all of these considerations work together. Portfolio assessment, at face value, appears to be the logical way to assess writing, but is it?

CHAPTER 2

PORTFOLIOS: The Road to Authenticity

Teachers have often required students to keep a file folder of their writing assignments. This folder offers, throughout the course, a dynamic picture of a student's efforts and growth as a writer. At the end of the course, the folder becomes a testimonial to the student's progress. Ostensibly, MUSD's original writing portfolio existed for the purpose of a district-wide assessment; in reality, the original portfolios were simply "testimonial" folders. The "new" Stepping Stones system employs methods, methods that were missing in the original plan, for utilizing the contents of these folders as tools for large scale assessment. It is when we, as instructors and writing specialists, begin to develop means for using authentic assessment for large scale accountability, that we enter a new arena in writing assessment. "Portfolios," according to Edward M. White, "offer to the world of assessment a view of student learning that is engaged, and dynamic, as opposed to the overwhelming passive concept that still dominates the assessment movement" (*Portfolios* 27). This "engaged and dynamic" form of assessment reflects our current definition of writing as a "rich and multifaceted activity deeply immersed in the context that surrounds it" (Camp 45). "This complex view of writing," according to Camp, "is not easily reconciled with traditional approaches to assessment" (46).

Traditional methods of assessment did not directly assess writing. Prior to the 1960's, indirect methods, such as multiple choice exams, were used to assess or measure writing ability. In the 1970s, "American educators were inundated with legislative requirements for testing that were part of the 'educational accountability movement'" (Popham 471). Because of this inundation, teachers and educational programs became subject to assessment, and they began to be held accountable for student assessments. The result of this accountability focus was that teachers began "to emphasize in their instruction the knowledge and skills that were being tested" (Popham 471). Teachers, rather than focusing on instruction for the purpose of teaching writing skills, instead, focused on instruction for the purpose of preparing students to pass the "accountability" tests. Today, this accountability focus (as evidenced by Stepping Stones) is still an important issue; however, with the advent of authentic assessment, the opportunity to align teacher autonomy in teaching writing with acceptable and accountable assessment measures is increasingly possible.

Historically, teacher autonomy (teachers making the decision to teach subject matter in what they see as the best way possible) and institutional accountability (students passing state tests) appear to be competing interests. Ideally, they should be compatible interests built from the ground up and involving authentic assessment.

While portfolio assessment appears to be the impetus to create the ideal, the reality is that the ideal is not so easily achieved. For example, in order to create a reliable assessment there must be some sort of continuity and commonality in the portfolios being assessed. Therefore, criteria must be developed and *imposed* on teachers and students. This imposition threatens autonomy. Teachers then may feel as if they and their students are more pawns than participants in the assessment process. Stepping Stones tries to alleviate this problem by making sure that teachers develop the criteria, and that this criteria leave sufficient room for site and teacher autonomy. More than likely, teachers will continue to feel somewhat imposed upon. However, the issue is not whether or not teachers feel imposed upon, but to what degree that imposition is acceptable, and ideally, empowering. It remains to be seen whether or not teachers will feel these efforts were adequate.

Furthermore, by definition, a large scale assessment, such as one that includes an entire school district, has the kind of breadth (large amounts of students involved) that makes depth (extensive, personalized assessment) difficult to achieve. Depth involves individualized attention: learning styles, maturity, personal dynamics, and other such considerations that could be taken into account when assessing student writing. This kind of depth, while making

assessment more valid, is impossible to achieve when dealing with the breadth of a large population of students. The balance between breadth and depth is difficult to achieve and therein exists the opportunity and the challenge for teaching professionals.

Stepping Stones is an example of the genesis of a large scale assessment plan which hopes to strike a balance between depth and breadth, reliability and validity, teacher autonomy and institutional requirements, cost and budget, and, while seeking this balance, to attain the goal of student learning. While these aspirations are commendable, are they achievable and/or practical?

To answer these questions, we must discover whether or not the benefits of portfolio assessment outweigh the problems. For example, are essay exams, which fulfill the need for authentic assessment in so far as actual writing is being used, a better choice for large scale assessment?

For an essay test to be valid, the content of the test must have a "high degree of match . . . [with the] definitions and interpretations assumed by the reported score" (White, *Teaching and Assessing Writing (TAW 1985)* 186). The current definition of writing theory has expanded from the simple view of writing as a way to *display* knowledge to a more complex view of writing as a way to *create* knowledge. Current writing theory sees writing as a process, one that develops over time in a recursive manner

utilizing revision and editing. Certainly essay tests that do not allow for reflection and revision, are not a validation of what a student learns in a process oriented classroom.

Essay exams are used often by universities as a way to determine a student's writing proficiency. Elbow and Belanoff question the validity of this type of assessment:

[When] a proficiency exam embodies a university requirement, the whole university can be seen as saying to students, "Here's a serious matter Tell us what you think about it in approximately five hundred words; we know you can give it the attention it deserves; and then you can go home." The writing is unconnected to any material and cut off from connection with any conversation. Is that how we want students to approach serious intellectual issues? (5)

The type of assessment Elbow and Belanoff describe is one that is separate from anything students have been taught about learning, thinking, and writing in a process based composition classroom. When an assessment does not correlate with instruction, then, it is misrepresenting instruction. If an assessment does not represent writing, as defined by the instruction, then the assessment is direct and may not be valid. And, if the assessment is not valid, then we must question the purpose and the consequences of the assessment. Students, who have been taught that writing is a process may well wonder why an assessment does not include a process approach. And students will surely feel a

disconnection between what they have been taught and what they are having assessed. They will certainly, and rightfully, question, at least in their own thoughts, the validity of such an assessment, and by virtue of those questions, their relationship to writing instruction, perhaps even to writing itself.

When assessment is a natural outgrowth of classroom instruction, teachers no longer feel the need to teach to a test, and students no longer feel a disconnection between what they have learned about writing and how they are being assessed. Furthermore, students involved in portfolio assessment "will," according to Daiker et al., "gain self-confidence both as writers and persons, they will develop critical thinking and evaluative skills, and they will become more independent and self-sustaining" (2). Students involved in portfolio assessment are required to make choices about which writings to revise and which to include in their portfolios. It is this requirement that allows students to attain the qualities that Daiker et al. describe. Portfolio assessment teaches responsibility because it requires that students take ownership of their own assessment; portfolio assessment offers opportunity for lifelong learning because it teaches students to become independent, critical thinkers, and portfolio assessment validates teaching because it is an extension of a student's learning experience.

While there are many compelling arguments for choosing portfolios as large scale assessment tools, Richard Larson cautions against choosing this tool without first examining the "political implications of such an assessment" (272). Larson's definition of "political" in this instance means "the relationships between the people who wield power in a given situation--those who have, or have asserted, authority over what happens in that setting--and the people whose behavior is directed or influenced by the people in power" (272). Because portfolios grow out of classroom assignments and are "the products of an interchange between teacher and student," as opposed to an assessment that is imposed from external sources, Larson suggests that in some ways, using these portfolios for assessment other than in the classroom, is "like an invasion of privacy--an invasion of the teacher's classroom" (272). Larson's observation points out the fact that some teachers, rather than seeing portfolio assessment as a way to bring authentic assessment into the realm of large scale assessment, may see this type of assessment as a political ploy on the parts of administrators to judge their classroom practices, a judgement that would more than likely be based on some sort of outside imposed criteria. Larson, in looking at these political issues of authority and consent, poses the question, "What happens, for instance, if the decision to use portfolios is not the teachers' decision" (275)?

More often than not, teachers are not involved in the selection of large scale assessment tools. Standardized tests are presented to teachers who act as proctors for the test and, once finished, get on with the business of teaching. Prompts for essay exams may be developed by teachers, but the essays themselves do not evolve out of a unit the teacher has developed for the individual classroom. On the other hand, portfolios are developed over time in the classroom. They may require that teachers adjust their day to day teaching styles to conform to the needs of the required portfolio. Portfolios impose on teaching in a way that no other writing assessment has. "These issues," according to Larson, "become particularly acute when portfolios are introduced primarily as a means for assessing students systematically, at arm's length" (276). Teachers may feel that, rather than teaching writing in their own way, they are instead being used as pawns for helping students to produce the perfect required portfolio. Furthermore, teachers may believe that if their students do not produce these portfolios, their teaching may be judged as inadequate. One way to prevent teachers from feeling this sort of "invasion" into their classrooms is to make sure that teachers are included in the process. Larson concurs:

The implication seems clear: an institution whose leaders want to see portfolios used for assessment or for teaching must engage faculty members in

adequate, open and democratic discussions of what is involved; the leaders should not rely on mandates of indoctrination. (277)

Certainly, MUSD administrators bring their awareness of the importance of including teachers in the process into the Stepping Stones portfolio project. The Stepping Stones portfolio plan, although developed by teachers and administrators, still faces all of the challenging questions to which Larson alludes: How do we create portfolios which conform to standards that do not impose on teacher autonomy? How do we get all teachers to agree to portfolio assessment? How do we involve all teachers fairly in the assessment? How do we involve all students fairly in the assessment? Larson advises, "Wise administrators will recognize that the possible benefits of portfolios assessment will not be achieved without the cooperation of the faculty and maybe of the students, too" (283). Larson continues by describing the benefits of portfolio assessment that is realized through such a democratic procedure:

[Administrators] will bring into the open, for departmental and even campus- wide discussion, questions about what constitutes literacy, what the term "Writing" embraces, how "ability to write" may be understood, and even what "reading includes. They will discover that in this process portfolio assessment becomes for all participants not threatening, not political, but educative. (283)

The kind of discovery Larson describes, discovery that educates, will only come about if a group of interested

professionals are willing to face and accept the challenge of answering the tough questions that large scale portfolio assessments pose, questions that have no easy answers, questions that will take everyone working together for a common goal to answer. Herein lies the best reason for pursuing a method of assessment that seems at times overly political, complex, time consuming and "bulky" — administrators, teachers and students can work together developing and enacting portfolio assessment. If an assessment can be a means to developing a community of people working together to understand each other and to extend their learning experiences, then this assessment has the kind of validity that deserves our attention and efforts — the kind of validity that cannot be ignored.

CHAPTER 3

VALIDITY: The Enlightened Path

Peter Elbow believes that we should "[use] all [of] our professional influence and rhetorical skill to persuade institutions to refrain from making significant assessment decisions except on the basis of well-furnished portfolios" (*Writing* 121). He goes on to make the compelling charge "that any other method of writing assessment is unfair, untrustworthy, and unprofessional" (*Writing* 121). Elbow's dogmatism is probably based on the fact that a "well-furnished" portfolio can complement current teaching practices which view writing as a process. Because assessment based on portfolios, according to Elbow, is "a huge improvement over assessment based on single samples of writing," he feels confident in proposing that portfolio assessment is the *only* worthwhile form of assessment when the assessment is going to be used for any meaningful purpose (*Writing* 120).

It seems logical to agree with Elbow that portfolios are the most valid tool for assessing writing when we define a valid assessment as one which most represents or matches the construct (writing) being assessed. Validity, according to Roberta Camp, "is now seen as a single unified concept in which the construct to be measured is central to all other considerations" ("Changing the Model" 60). If we choose an assessment method that does not parallel the construct we

are purporting to measure, then we must question the purpose of the assessment. Consequently, says Camp:

[Our] concerns about the possible deleterious effects of conventional assessment formats [such as multiple choice tests and essay exams] on students . . . no longer appear peripheral; they are central to validity, especially if these effects derive from misrepresentation of the construct of writing. (61)

Camp points out the possibility of negative consequences for students when assessment and teaching are at odds, and she is concerned that this is exactly what is happening when we continue to use traditional methods of assessment to measure non-traditional, or current, methods of writing.

Elbow's confidence in portfolio assessment along with Camp's concern over traditional methods of assessment seem to make an indisputable case for the validity of using portfolios to assess writing. However, before we accept this pronouncement as the *only* truth, it is important to step back and try to understand exactly what we mean when we speak of validity and exactly how validity works in conjunction with other aspects of writing assessment.

When we talk about an exam or assessment being valid, we are saying that a test is actually measuring what we say it is going to measure. Validity has historically been determined in a number of ways such as matching the content of the assessment to the construct being measured, determining if by the looks of it (face value) a method

seems fair, comparing scores from two different types of writing assessment for concurrence, and by judging the efficacy of an assessment in predicting future assessment results.

Predictive validity attempts to predict how well a student will perform in a given situation. "[The] point of reference," for predictive validity according to Edward M. White, "is the degree of accuracy the test scores exhibit when used to make predictions about student performance in another setting" (TAW 185). For example, Student Aptitude Tests (SATs) attempt to predict how well students will do in college. The problem with this type of thinking in relation to the SATs is that, because the test has been criticized for catering to the middle class, the test may actually be measuring "middle class cultural conditioning" rather than college success. If this is true, if the SATs function more as a barrier than an opportunity for potential students from the working class (whether or not that is their intended use), then this test is not honestly assessing what it purports to assess; it is not sound, well-grounded, or valid. Predictive validity is not a useful tool for determining the validity of an assessment when predictions are made based on false assumptions. However, when applied properly, predictive validity serves its purpose.

Colleges and universities very often use an impromptu essay exam in order to determine or predict how well a given

student will do in one of the offered writing courses. "One very common use of [this type] of writing assessment [procedure] is to determine student placement into different levels of a writing curriculum" (Williamson "Validating" 12). These students have already been accepted into the university; therefore, the most serious consequence of a misused or misread exam is that some students may be incorrectly placed. This mistake is relatively simple for an instructor to rectify by conferring with the student and possibly assigning a new class. Using an assessment to predict, when the assessment is flexible and designed with various production outcomes in mind, is valid. Predictive validity is often used in conjunction with correlative or concurrent validity.

Concurrent validity "refers to the degree of agreement between scores of two different tests of the same skills" (White TAW 185). If students achieve similar outcomes for more than one test of writing, then the assessment scores, according to the theory of concurrent validity, are valid and can be used to make judgments about students' writing abilities. Concurrent validity is sometimes used to support the use of multiple choice exams as a tool for assessing writing. "The claims for the validity of using multiple-choice tests to determine writing competence," according to Roberta Camp:

are not entirely without foundation. Most students who do well on carefully designed and relatively comprehensive multiple-choice tests of grammar, sentence structure, and usage are likely to perform well in response to well-designed prompts for writing, as the correlation studies indicate. ("Changing the Model" 47)

In this instance, Camp is looking at scores of students who have taken both a multiple choice exam and an impromptu writing exam and then compares these scores to find concurrence between them. Her statement hypothesizes that, because there is an agreement between the two scores, a multiple choice exam may be as valid a way to assess writing as is an essay exam. However, teachers of writing have long known that the best measure of writing is writing; furthermore, the kind of complexity of thought required in composing a piece of writing cannot be measured with a multiple choice exam. Concurrence between multiple choice tests and other writing assessments is probably a determination of something, but exactly what that something might be is up for debate.

Concurrent validity, in the case of Camp's previous example, may be offering a picture of how consistently given students perform on a variety of writing tests (in this case, the student can be assessed as to her or his prowess as a test-taker); however, this concurrence does not give an accurate or valid picture of these same students' writing abilities. Camp states:

The historical development associated with emerging views of validity is moving the field away from a primary emphasis on patterns of relationships such as the correlational [concurrent] studies of predictive validity so often used to justify multiple-choice tests of writing, toward methods more likely to be supportive of complex performances in writing. ("Changing the Model" 62)

While multiple choice exams may be able to assess a student's comprehension of the patterns of English, they cannot assess this same student's ability to write sentences, paragraphs or essays that reflect the student's thought and findings in a logical, ordered and meaningful way. Complex writing performances cannot be assessed by multiple-choice exams, even when these exams concur with other tests. Face validity, conversely, offers a step toward the direct measurement of writing by focusing on actual writing.

While predictive validity predicts, and concurrent validity correlates or compares, face validity assumes an assessment is valid by how it appears, on the face of it, to the assessor. According to Edward M. White, "[The] use of face validity represents one of the major reasons that research projects are difficult to replicate; writing that seems obviously better on the face of it to one observer, may look quite different to another" (TAW 186). Face validity is, as White suggests, subjective. It is a "common sense" measure in that a writing professional determines

validity by whether or not the assessment has traits common to what he or she views as writing. However, what constitutes good writing to one professional, may not be the same as what constitutes good writing to another. While all direct assessments of writing are somewhat subjective, face validity, because of its highly subjective and non-transferable nature is of little value in large scale assessment.

Stepping Stones is, in many ways, still at the face validity or, a "step in the right direction," phase. While actual writing is being assessed, the assessment criteria is still completely up to the individual teacher. What one teacher regards as a "6," another might regard as a "4." The MUSD portfolio requirements do not as yet involve any kind of common criteria other than the three particular types of essays. However, including actual writing as a part of the state's assessment plan is still, on the face of it, a step in the right direction toward authentic assessment. While face validity is a questionable forum for serious assessment, it can be used as a starting point for choosing an appropriate method of assessment. Assessors can determine, on the face of it, which types of assessment methods seem to be the most suitable for a given situation. Their choices can then be narrowed through discussion and negotiation.

The use of face validity, while of limited value, still demonstrates a positive move away from indirect assessments

of writing; however, according to Gail Stygall et al:

[while] the move away from indirect writing assessment . . . and concomitant move to direct assessment . . . answered questions of "face" validity, issues of content and construct validity continued to loom . . . Thus the discipline's first move toward "face validity," assessment of an actual piece of writing, was only a partial solution. (1-2)

Stygall's admonition alludes to the fact that while using direct assessment was and is still a step in the right direction, it was only a partial step. It is not enough for writing specialists to simply agree "that direct assessment is the most valid approach to writing assessment"; they must also "[focus on] important concerns about the current limits of direct assessment tools and the need to continue refining them" (Williamson 14). In order for an assessment of writing to be valid, a further step must be taken to ensure that the content of the assessment tool chosen matches the definition of the construct being measured.

Content validity is defined "as the judgment of experts about the adequacy of the content of a test and the testing procedures themselves to measure the phenomenon of interest" (Williamson 11). In other words, the content of an assessment must match or represent the construct (the phenomenon of interest) being measured. For example, suppose students who are taking an impromptu essay exam for the purpose of assessing the quality of their writing

(writing being the phenomenon, or construct, of interest)

receive the following prompt on which to write:

Write an essay comparing and contrasting the first 100 days of Franklin D. Roosevelt's presidency with that of the first 100 days of Ronald Reagan's presidency. You have exactly one hour to complete your essay.

After students recall facts and organize their responses, the time that remains for writing varies with each individual; therefore, it is not valid to say that each student spent the same amount of time actually writing. Furthermore, if assessors are unable to separate writing ability from historical knowledge, they cannot claim to be assessing strictly for writing ability as "poor" writers with prior historical knowledge may produce a more impressive factual essay than "good" writers who do not possess the same knowledge. And, if assessors ignore the knowledge and look simply at the writing (if such a thing is possible), what is the purpose of requiring students to devote time to recalling facts? "[Some]" according to White, "will challenge the validity of scores derived from a single essay on certain kinds of topics because they feel that differences in scores may be related to differences in the amount or quantity of relevant knowledge the student writer had available in formulating a response" (TAW 188-189). White's point must be conceded, and those who would challenge the validity of the scores based on the aforementioned prompt would be correct to do so. It is

clear that the content of this prompt, while possibly valid for assessing specific historical knowledge, is not necessarily valid for assessing the ability to write well on a generic topic.. "Some questions," according to White, "give an unfair advantage to students with a particular kind of knowledge The ideal question will allow the weak students to write comfortably enough at their level while it challenges the best students to produce their best work" (TAW 111). It is essential that the writing prompt be conducive to eliciting "good" writing, rather than "good" knowledge of a specific topic. The prompt should also "be carefully developed, with an eye to the stated test criteria, by a committee constantly refreshed by new members" (White, *Holistic* 93). Only when all of these measures have been taken, can a prompt for an essay exam be considered content valid.

While a good writing prompt is essential, it does not solve the problem of content validity in relation to current needs for writing assessment. According to Edward M. White:

Even when careful test committees establish test criteria and specification and offer a well-developed set of questions . . . , we remain uncertain that we have defined the representative content of the material we are examining. And when a writing test offers students only one topic . . . and one short period of time for response, our uncertainties are compounded. (TAW 187)

Current methods of defining and teaching writing usually

involve process as well as product, and these methods are not adequately represented by one-shot impromptu essay exams. Certainly, essay exams "in contrast to the fill-in-the bubble tests" are a more valid way to assess actual writing, but "the claim that an essay test represents real writing now seems questionable" (White, *Portfolios* 32-33). If we define "real" writing as writing that is produced on demand in a short period of time without any time for revision, then a well-written prompt will more than likely produce this type of writing. However, we should be aware that preparing for an impromptu essay exam may force students to produce what Joan K Wauters calls "the lockstep of formulaic writing: the five paragraph essay" (68). While the five paragraph essay does not conform to current theories of writing which view writing as a process that begins with an unpredictable outcome (and an unpredictable paragraph count), it does seem to conform very well to the conditions imposed by a timed impromptu essay. Elbow and Belanoff's concern over the message we are sending to our students, "Here's a serious matter Tell us what you think about it in approximately five hundred words; we know you can give it the attention it deserves; and then you can go home." which is voiced earlier in *Chapter Two*, can be revisited (5). Elbow and Belanoff ask, "Is that how we want students to approach serious intellectual issues?" (5). This question indicates that the pressure to produce a well-crafted "formulaic" essay in a short period of time may

override a student's concern that the essay be creative and thoughtful. According to White, the "'reality'" of impromptu writing "is of a peculiar kind: first draft (usually), pressured, driven by external motivation rather than an internal need to say something" (White, *Holistic* 90) White's comment reinforces Elbow's and Belanoff's concern that impromptu essay exams may actually have a deleterious effect on student writers, a concern that is reiterated by Roberta Camp as she comments on essay exam prompts:

The prompts that we so carefully designed for equal accessibility are now seen to cut off the opening explorations of a topic in which writers find a way into it that engages their interests and allows them to use their knowledge and skills to best advantage. (52)

Ironically, one-shot impromptu essay exams, rather than assessing our students' writing as reflected by what we teach in process-based composition classrooms, may instead be reflecting and assessing our students' abilities to conform to external, unrehearsed standards. In order for an writing assessment to be valid, it must conform to a specific definition of writing; essay exams do not always conform to these definitions.

Our definitions of writing have expanded to include such metacognitive activities as exploration and discovery, both of which require a commitment of time. A commitment by students to engage in this exploration, this writing

process, requires a fair and equal commitment by assessment professionals to develop a new means of assessment, one that will contain and support this process of exploration, discovery, and metacognitive learning. Writing portfolios are ideal for such assessment.

According to White:

The great advantage of portfolios for assessment is that they can include numerous examples of student writing, produced over time, under a variety of conditions. . . . [They] can showcase several kinds of writing and rewriting, without time constraints and without test anxiety. Whereas most evaluation instruments provide a snapshot of student performance, the portfolio can give a motion picture. (*Assigning, Responding, Evaluating (ARE)* 63)

This "motion picture," however, must change according to the needs of each assessment purpose. In other words, simply because a portfolio contains a variety of writing does not make its contents valid for all writing assessment. The content of a portfolio is similar to the prompt for an essay exam in that both must match the particular construct being measured. For example, if the construct for which students are being assessed is writing improvement, their portfolios might contain several drafts and the final products of one or more essays written throughout the semester. If students are being assessed for their ability to write well in a variety of modes, portfolios may contain such essays as a reflective piece, a comparison/contrast piece, a summary, a

research paper, etc. (It can be assumed that, although actual drafts are not in the portfolio, included essays are the products of revised work). Portfolios can be adapted to include a myriad of topics, modes, drafts, and genres, depending on the construct being measured for each particular assessment. In any case, the content of a portfolio has a much better chance of matching, and thereby validating, writing as a recursive and continuing process than does an impromptu, one-shot, essay exam.

Content validity cannot, however, be established without first establishing construct validity; the two co-exist. Construct validity is the extent to which an assessment tool embodies a theory of writing, and in order for an assessment to be valid, the content of the assessment must also support this theory. "What we are experiencing," according to Camp:

is a mismatch between the complexities of the conceptual framework for writing that we find in current research and practice and the simpler construct implied by traditional approaches to writing assessment, including the writing sample.

(52)

In other words, while our definitions of writing have changed to involve a level of thinking that includes the recursive process of writing and re-writing, our means for assessment have not; therefore, current means of assessment, such as the impromptu essay exam, do not necessarily

validate or match the construct (writing) we are purporting to measure. In order for an assessment to be valid, "research [must confirm] . . . the link between the construct (writing ability) and the test" (White TAW 188). If there is no such link (as Camp suggests is often the case), then the assessment is not valid. As writing professionals, we must work together in developing methods of assessment which *do* match the theory that writing is a thoughtful evolving process, one which requires time for revision.

Because defining "good" or "bad" writing is in itself a subjective notion, measuring writing ability will never be an exact science. The closest we can come to an honest, valid assessment is to choose a tool that matches, or validates, current theories of writing, and current methods of teaching writing. "Thus," according to White, "to measure student competence or achievement in composition requires a workable definition of "good" writing and a theory of how competence in writing develops" (TAW 189). White continues, "Every teacher and writer knows, and writing process research is continuing to confirm, that revision is an essential part of writing. Every time we give an important grade for first draft writing, we deny in practice what we say about revision" (TAW 189). White, a leading authority in both the teaching of writing and writing assessment, is affirming and helping to create a workable definition of writing by stating that competence in writing usually comes

about through a lot of practice, by way of revision. Therefore, it seems viable that the revision process should be a part of any *valid* assessment which claims to be measuring writing competence. Roberta Camp affirms this claim with the following concern, "[We] once regarded a piece of writing in a single mode or for a single purpose to be a sufficient sample, we now see it as insufficient to represent the variety of modes and purposes for writing . . ." (Camp 51-52). Portfolios, which celebrate revision and include more than one writing sample, are a method of assessment that alleviate Camp's concern about variety and satisfy White's desire for revision.

An added benefit to using a method of assessment which involves many drafts of writing, as well as a variety of modes, is that assessment can grow out of classroom assignments, thus creating a link between assignment and assessment. This link helps create a better overall idea of a student's writing ability. While Kern argues that "no one has yet identified or explained what quantity and variety of writing 'will' provide a valid picture" (51) of a student's writing ability, it seems clear that we will come closer to that "picture" by looking at more than one sample of writing, writing which has been produced over a period of time through a number of revisions. As Elbow and Belanoff state: "We cannot get a trustworthy picture of a student's writing proficiency unless we look at several samples produced on several days in several modes or genres" (5). A

"trustworthy," or honest picture of a student's writing ability is what we need in order to call an assessment valid, and portfolio assessment offers this type of validity.

Portfolio assessment, which includes time for process, allows us to take the meaning of validity one step farther; it allows us to look at the opportunities for students and teachers alike to benefit from and find personal purpose in writing assessment. According to Brian Huot:

With the portfolio as a gauge of writing quality, being able to write well implies the ability to compile a representative sample of work, which reflects not only the writer's ideas, goals and interests but also her knowledge and awareness of readers' criteria and expectations. In other words, portfolios exemplify to students, teachers and testers that writing is an ongoing process (Beyond 329)

When this process becomes the basis for assessment, when this assessment involves students reflecting metacognitively about their writings, when it involves students and teachers working together to make decisions, when it requires teachers gathering together to discuss writing issues, and when it involves both teachers and students in the assessment process, then this assessment is providing the kinds of consequences that have never before been attributed to a writing assessment. The next chapter will examine a few of the positive consequences that portfolio assessment offers teachers and students.

CHAPTER 4

CONSEQUENTIAL VALIDITY: Detour to Success

"Consequential" validity is probably the most important type of validity an assessment can offer because of the positive repercussions which work to unify issues of "teaching, learning, institutional goals, and student individuality" (Belanoff and Dickson xxiii). Portfolio assessment embodies this unification. "Portfolios enable assessment, but they also reach out beyond assessment and engender [change]" (Belanoff and Dickson xxiii). This opportunity for effecting change makes portfolio assessment uniquely situated among other methods for assessing writing.

Portfolio assessment sessions are particularly valuable for teachers because they offer teachers the opportunity to work together collaboratively. While composition instructors everywhere are enthusiastically promoting collaboration for their students, collaboration does not seem to have caught on as a means for developing interaction among these same instructors. Very often, instructors in the same school have little or no idea what is going on in classrooms other than their own. In order to keep from growing increasingly isolated and stagnant, and in order to keep what Stephen M. North calls "The Great Debate" alive, we need to search continually for new ways to develop community among instructors, communities that will encourage growth and understanding through conversation, negotiation

and collaboration. One such "new way" is the collaborative assessment of portfolios. While collaborative assessment readings of essays have also brought teachers together, portfolio assessment is unique in that it offers teachers a sort of "mini-view" into one another's methods of teaching.

In portfolio assessment sessions, faculty gather together to read, assess, and discuss portfolios from one another's composition classrooms. It is not enough, however, to simply passively discuss and accept the contents of these portfolios. In order to validate the consequences of portfolio assessment for teachers, they must become active participants in these sessions; they must question one another's teaching methods and motives. According to Anne Gere, "Participants in collaborative groups learn when they challenge one another with questions, when they use the evidence and information available to them, when they develop relationships among issues, when they evaluate their own thinking" (69). This kind of challenge is met in portfolio assessment sessions where instructors are given the opportunity to discover and discuss their colleagues' classroom pedagogies. Peter Caccaveri considers the prospective value of one of these sessions, "Teachers question the criteria, even values of other teachers, and have theirs questioned in turn. Teachers 'learn,' rather than just teach, and they get a sense of community which is reassuring as well as unsettling" (50). Caccaveri's and Gere's theories on collaboration and community combine to

validate the consequences of portfolio assessment for instructors; they not only develop community through negotiation and conversation, but they also gain insight into one another's classroom pedagogies. The following study exemplifies these consequences for instructors at a large mid-western university.

Marjorie Roemer, Lucille M. Schultz, and Russel K. Durst worked together in order to study the effects of portfolio assessment on teachers and administrators. Their study included three pilot groups. Roemer worked with teaching assistants; Schultz worked with experienced full-time teachers, and Durst worked with part-time teachers. Each pilot study lasted one quarter, and each group was encouraged to "find its own way" in their respective portfolio assessment sessions.

Roemer's pilot group of teaching assistants appreciated the opportunity to discuss issues and assignments with each other. They concluded that they would strongly endorse the system because portfolio assessment "[created] a sense of community among portfolio group members" and it "[encouraged] interaction among teachers and more discussions regarding students' texts. (460) These sessions helped those involved by developing a community where beginning teachers had "someone else . . . to help shoulder the burden" of their first year teaching, and by offering insight into one another's teaching practices.

Schultz's pilot group included eight experienced teachers who, from the beginning, "seriously interrogated portfolio assessment" (461). Accountability was a major concern for the majority of these teachers. "[T]eachers said that they were worried that their standards, standards they had used for years, might be different from those of their colleagues; . . . they also worried that there could be pressure to adjust their standards" (461). This concern springs from the fact that portfolio assessments, unlike impromptu essay exams, are based on classroom work. Teachers, therefore, are asked to publicly examine and discuss with colleagues an aspect of their teaching which had previously been relatively private. The participants of this pilot found that their fears were unfounded as is displayed in the following sample comments: "It was good in that I saw what [my colleagues'] students were doing. I felt less isolated . . ." and, "I like the portfolio system. I know there are a myriad of details to work out, but it feels right. It feels right to talk to each other about what we do" (462). Schultz reiterates these comments in her summing up of the participants overall assessment of portfolios: [The] teachers who completed the project [two dropped out for personal reasons] . . . were enthusiastic about it. In particular they reported that they enjoyed the collegiality and the support . . . " (462). The teachers involved in this study were at first concerned that their classroom pedagogies would be challenged in ways that they

would find uncomfortable. Instead, they discovered that sharing and discussing their classroom practices with colleagues could be a rewarding as well as helpful experience.

While not as far along in the process as Shultz's group, Stepping Stones participants still enjoyed the benefits of this communal experience. Teachers from across the district, many of whom had never met before, spent hours discussing and sharing their viewpoints on writing, writing assessment, and district policies. An aura of enthusiasm was pervasive throughout each sessions as participants agreed, disagreed, argued, collaborated, compromised and found eventual consensus on issues of writing. The merging of voices brought a kind of solidarity to this small group of teachers and district officials which had not existed on the first day of meeting. Stepping Stones participants all agreed that this unexpected communal benefit was one worth sharing and promoting through future meetings.

Communities founded on portfolio assessments substantiate that the quality of discussions and the general "bringing together" of interested professionals qualify this type of assessment as being consequentially valid. These communities involve a dialectic which encourages disagreements and discussion over one another's pedagogical practices. Because teachers are grading each other's students' work, and not an impersonal test, grading can have "painful moments" (Belanoff and Elbow 21). Some teachers,

Elbow and Belanoff observe, are "pleased to discover the striking disparity of standards that sometimes emerges [Others] are disturbed to feel moving sand under the foundation--as though everything is arbitrary and anarchic" (20). However, as painful as it may be, this type of disparity, and the questioning of standards that is likely to come about because of this disparity, is necessary for communal growth. Kenneth Bruffee, one of the early proponents of collaborative methods, believes that without these types of challenges a community will become "stale [and] unproductive" (648). Much of the excitement generated during Stepping Stones meetings came about because teachers, frustrated with "stale" assessments such as the California Test of Basic Skills (CTBS), began to feel as if they were breathing life into an otherwise stagnant, unproductive system. The idea of a writing assessment based on actual writing developed through a process motivated the Stepping Stones group to find ways to work together towards this common goal. By facing these challenges and working toward consensus, teachers involved across the country in portfolio assessment are engendering positive changes in their communities.

The last of the three pilot studies, which included Durst and three very busy part-time composition instructors, also found portfolio assessment to be consequentially valid. The participants of this study, while having to find ways to overcome time constraints, draw the following conclusion:

"[We] got closer to sharing goals and expectations by looking at portfolios of student writing than we ever could have in more abstract ungrounded discussions of single papers in a single mode" (467). These part-time instructors, by reading portfolios produced in one another's classrooms, were able to glimpse a little of their colleagues' pedagogical goals. This glimpse allowed even these busy teachers the opportunity to belong to, and benefit from, the kind of community that grows through shared experience and portfolio assessment.

Roemer, Schultz and Durst draw the following conclusion from their study:

[Portfolio assessment] is a system that gives power to teachers [Teachers] come together and hash things out. . . . We have learned that portfolio assessment is a system that depends on trust, on flexibility, and on a willingness to decentralize; but then, so does a lot of the best teaching. (468)

Decentralization, trust, and flexibility, are positive consequences for teachers who work together in communities founded on portfolio assessment sessions.

Along with validating communal growth, portfolio assessment offers the opportunity for teachers to bring teachers and teaching back into the assessment loop. "Many writing tests," according to Edward M. White "currently are imposed from outside the classroom . . . and scored in more or less mysterious ways" (TAW 1). This is a problem, according to White, because "[we] cannot separate teaching

from . . . measurement" and still call the measurement valid (TAW 189). Portfolios are developed within the classroom; they validate teaching and assessment as being part of a continuum, not a separate and alien entity, and because portfolios are developed in the classroom, students also become an integral part of the teaching, assessment continuum.

One of the most important and exciting consequences of portfolio assessment is that it brings students into the assessment loop: "By shifting responsibility to our students, we ask them to be more than mere recipients of someone else's paper and-pencil tests. They must be active, thoughtful participants in the analysis of their own learning" (Murphy and Smith 58). This participation begins in the classroom where the contents of writing portfolios are developed by students throughout the semester.

"Students know," according to Peter Elbow,

that their portfolios of finished pieces will have a better chance of passing or getting a higher evaluation if they have made use of all elements of a rich writing process. Thus, portfolio grading helps the learning climate because it reinforces continuing effort and improvement . . .
.. (*Introduction* xv)

This continuing effort through discussion, revision, editing and choice, makes portfolio assessment an integrated internal part of a student's writing program rather than an external entity beyond a student's control. Assessment then becomes "but one function of a student's writing, reducing

the importance, power and significance of the testing moment as a legitimate reason to write" (Huot, *Beyond* 329). Students may begin to see assessment as an ally, an opportunity for growth, rather than a final irreversible, "unrevisable" external judgment. "Portfolio assessment," according to Elbow, takes the stance of an *invitation*: "Can you show us your best work, so we can see what you know and what you can do--not just what you do not know and cannot do?" (*Introduction* xvi).

This *invitation* makes portfolio assessment consequentially valid for students in that assessment becomes "not just an ending, a final exam so to speak. It [becomes] a beginning" (Murphy and Smith 58). Students, through portfolio assessment, begin to challenge themselves as writers. They begin to examine their own writing processes and search for ways to improve. In the course of this search, they begin to communicate with others who can support and help them, and through this, students begin to create a community of writers.

Portfolio assessment encourages students to become active members of a community of writers who are working together to better themselves. Writing has traditionally been thought of as an isolated occupation, one that involves only the word processor and the writer. The development of an assessment that invites interaction among writers sends "[the] message to students . . . that thinking and writing are enhanced by conversation with peers and teachers . . ."

(Elbow and Belanoff 15). Students learn to interact with others in order to revise their work thoughtfully and honestly. Conversation, negotiation, and interaction are means for developing community, and a community founded on portfolio assessment empowers its members by affording the opportunity for choice.

Daiker et al, in describing a portfolio writer's choice say that:

Writers choose 1) which of their pieces to submit, a choice which sometimes means deciding between a piece on hand and composing a new one; 2) how extensively, if at all, they will revise each piece; and 3) when and where they will do their writing and revising. (2)

Choice provides students involved in portfolio assessment with feelings of control and ownership. The power students feel when given control of their portfolios is akin to the power Schultz, Durst and Roemer say teachers feel when they are given opportunity through portfolio assessment sessions to learn to trust one another by "hashing things out and decentralizing" (468). Portfolio assessment empowers communities of both teachers and students to learn to trust and depend on themselves and each other. With this power comes responsibility, the responsibility to make *good* choices. In order to make good choices (and to have writing worth choosing), students must begin to examine closely both their writing and their writing processes.

Portfolio assessment has undoubtedly valid and

important consequences for students because inherent in the assessment is the need for students to use critical thinking skills in order to examine and make good choices about their writing. According to Mills-Courts and Amiran: "To the extent that students are asked to select and justify their contents, the first, most obvious advantage of portfolios is that they require a "doing" of learning that demands [an] intellectual self-consciousness" (102). This "doing" of learning goes beyond simply making choices for inclusion in a writing portfolio; this "doing" of learning requires intellectual self-consciousness-- metacognition--about one's method of learning, and about one's process of writing. When students' writing processes are unveiled, and they begin to make metacognitive connections between these processes and their processes of learning, then their writing becomes a vehicle for discovery rather than a disturbing road block on the way to their degrees. According to Edward M. White:

Process evaluation argues for complex judgments of competence based on more than the correctness of process. The process model sees writing as a series of overlapping activities, all of which have to do with critical thinking and problem solving: invention and prewriting, drafting, refining and rethinking, connecting, revising, and (finally) editing. The metacognitive activities associated with some definitions of critical thinking are an inherent part of the writing process, which requires that writers assess their work in order to revise. (*Assessing* 111)

Revision is embedded in portfolio assessment. Students will usually want to include their best writing in their final portfolios, and "best" writing often implies one or more revisions. In some instances, portfolios may actually contain drafts of work so that the assessor can measure progress. In any case, students must involve themselves in the critical thinking that White describes in order to produce and assemble their portfolios. Students take this act of critical thinking one step further when they are required to include in their portfolios one or more reflective pieces.

The reflective essay can include reflections on a myriad of topics. For example, reflective essays may require that students describe and discuss their reasons for included writing choices, their writing processes, their growth (or stagnation) as writers, and/or their experience assembling the portfolio. Whatever its contents, the reflective essay requires that students examine some part of their writing experience. "Students who learn to reflect on their writing . . . engage in a form of assessment that has greatest potential effect on their learning because it addresses directly their own awareness of what they have done and what they can do" (Camp and Levine 200). Students' awareness of "what they have done and what they can do" is a step toward self-assessment:

The reflective piece usually invites writers to evaluate their own work . . . and this implies the

value of self-assessment. . . . We need to remind ourselves again that we assess students today so that they can assess themselves tomorrow. (Daiker et al. 2)

The ability to assess oneself is a skill which has implications that reach far beyond the writing classroom. Self-assessment implies an ability to think critically about one's own work. This type of critical self-assessment requires that students think about *how* they think. "Most of the best research on cognitive development," according to Mills-Courts and Amiran, "suggests that it is extremely important to create situation in which students must think about their own thinking" (103). Portfolio assessment creates exactly this opportunity for growth.

Cognitive development is further enhanced when students learn to integrate their learning experiences into a focused whole [They learn to] reflect upon their own thinking--not just as college students but as lifetime learners, as citizens of the world. (Mills-Court and Amiran PPP 104)

When students have learned to become metacognitively aware of how they learn, then they have acquired skills which enable them to become lifelong learners--the ultimate goal of any writing program, indeed, of any educational program. These types of lifelong learning skills can be added to the list of positive consequences which can come about through portfolio assessment.

At the beginning of the previous chapter, I describe a

valid assessment as one that is sound, well-grounded in theory, and producing the desired results. Portfolio assessment can fill all of these requirements and then go beyond to include positive consequences for teachers and students, consequences that have the ability to produce "lifelong learners." There is little disagreement that portfolio assessment, when properly administered and carried out, is valid; and almost any way one looks at it, if validity was the only measure needed for determining the best method of assessment, portfolios would be chosen each time. It is for this reason that Peter Elbow makes the intriguing claim that any significant writing assessment is one that has been based on portfolio and that any other method is unfair (*Writing* 121). However, Elbow is wrong if he believes that validity should be the sole determiner in choosing the best method to assess writing. Issues such as reliability and cost-effectiveness (among others) must also be considered in relation to validity in order to determine the overall effectiveness and usefulness of a writing assessment tool. While issues of validity in portfolio assessment are relatively easy to resolve, issues of cost-effectiveness and reliability are not. The following chapters will focus on these issues and their consequences for portfolio assessment.

CHAPTER 5

RELIABILITY: The Long and Winding Road

While those in the field of composition readily accept direct measures of writing such as essay tests and portfolio evaluations as more valid ways of assessing and measuring writing than indirect measures such as multiple choice tests, there is still some disagreement about which direct measure is the *best*. Because portfolios can contain more than one sample of writing--writing that has developed through a process--there is not much doubt that portfolios are a more valid way of measuring writing than are impromptu essay tests. Validity, however, while extremely important to any measurement, fulfills only some of the necessary requirements for a good assessment. Reliability, or consistency, is equally important. "While high reliability in a measure or in a study ensures only consistency--not validity--no results can be meaningful if the measure used is not consistent" (White, *TAW* 177). Meaningful results are the ultimate goal of an assessment. Without both validity and reliability, an assessment is meaningless.

Validity research utilizes assessment scores in order to discover whether or not assessments are actually measuring the construct they purport to measure--are we sure the scores reflect writing ability? Reliability research uses scores to discover whether or not assessments are consistent in their measurements--are we sure the scores are dependable? For an assessment to be reliable, we have to

be able to depend on it being fair and reasonable. "[Reliability]," according to White, "is in a sense a technical term to describe fairness or simple consistency . . ." (TAW 22). While validity means "honesty" in assessment, reliability means "fairness." For an assessment to be fair, it must have some means of scoring that is consistent. A variety of assessors who are judging the same material, should have a high degree of agreement or consistency in scoring. Without consistency in scoring, an assessment is not reliable. "Reliability," observes White, "is a major concern for text-based writing research, but all research, indeed all measurement of any sort, needs to consider and demonstrate reliability if results are to be convincing" (TAW 180). Convincing, consistent results confirm the validity of an assessment for both teachers and students. While it seems clear that portfolio assessment is a valid tool for measuring writing ability, can the validity of this tool stand up to questions of reliability?

Peter Elbow, as we know, is impressed with the "improved validity" of portfolios because, for one reason, they "give a better picture of students' writing abilities . . . [than do] most writing assessments . . ." (*Introduction xi*). Elbow, however, is at the same time, concerned with issues of reliability that this type of assessment faces:

When a portfolio increases validity by giving us a better picture of what we are trying to measure (the student's actual ability), it tends by that very act to muddy reliability--to diminish the likelihood of agreement among readers or graders.

That is, if we are only looking at single pieces of writing by students--all written under the same conditions, all in exactly the same genre, all answering the same question--we are much more likely to agree with each other in our ranking than if we are looking at pieces by each student, all of them different kinds of writing written under different conditions. When all writing is alike it is easier to agree about it.
(*Introduction* xii)

In other words, Elbow believes that, because of their "sameness," they are much more likely to obtain consistent, reliable, readings of essay exams than can portfolios. While Elbow does a good job of identifying one of the problems among readers of portfolios who have to reliably assess "different kinds of writing written under different conditions," he himself does not view this as a problem. Elbow believes that disagreement among readers is an asset because, he claims, "if assessment is to bear any believable relationship to the actual world of readers and responders [people who 'consistently disagree in their interpretations or evaluations of texts . . .'], then reliability *should* be under strain" (*Introduction* xiii). Elbow continues his argument by stating that

[g]iven the tension between validity and reliability--the trade-off between getting good pictures of what we are trying to test and good agreement among interpreters of those pictures--it makes most sense to put our chips on validity and allow reliability to suffer. (*Introduction* xiii)

Elbow's willingness to "trade-off" and to gamble on validity while leaving reliability by the wayside seems odd in light of his strong belief in the importance of students. For

students to be assessed in a way that is as free from prejudice and subjectivity as is humanly possible, we should attempt to establish a method of assessment that is as fair as possible, and, asserts White, "if we are not interested in fairness, we have no business giving tests or using test results" (*Holistic* 93). White goes on to rebut Elbow in the following passage: "[Elbow] makes the grave error of asserting that reliability of measurement is not only unimportant, but actually in conflict with validity" (*Review* 538). White continues by saying that Elbow's separation of validity and reliability is a "false dichotomy" because "[reliability] means consistency and fairness, and no measure . . . can be more valid than it is reliable" (*Review* 538). Cherry and Meyer reiterate White's sentiments:

Reliability and validity are two of the most basic concepts in measurement theory. . . . In order for a test to be a valid measure of a trait such as writing ability, it must be both reliable and valid: it must yield consistent results, and it must actually measure writing. A test cannot be valid unless it is reliable. (110)

In light of measurement theory, Elbow's belief that reliability is unimportant, theoretically, negates his belief in the increased validity of portfolios.

While Elbow's notion to "throw away" reliability in favor of validity is unrealistic, his concern that reliability may compromise validity is not. Elbow is correct in stating that "the tension between validity and reliability" makes necessary some form of "trade-off" (*Introduction* xiii). There is always some form of trade-off

when assessment leaves the personal enclosure of the classroom and becomes the sole property of measurement theory.

When assessment takes place in the classroom, it is done by teachers who have personal knowledge of their students and their students' writing abilities. Teachers should be capable of assessing the consistency of their students' work by looking at such things as improvement and level of effort. Therefore, it is not important that students' portfolios are similar to one another, nor that they are evaluated in exactly the same way. Ideally, the teacher's personal knowledge of the individual student and professional judgment concerning the student's work can be the basis for determining reliability. It is this type of assessment that comes closest to what Elbow calls the "actual world of readers and responders" (*Introduction* xiii). However, the farther away we get from the classroom, the individual student, the farther we get from the ideal. And large scale assessment is about as far from the ideal as one can get. The "actual world of readers and responders" becomes an artificial world of assessors who must come to some kind of consensus based on common criteria in order to produce results that are both reliable and valid.

In order to maintain as high a level of reliability and validity as is possible in large scale assessment there must be some kind of trade-off--and this trade-off usually comes in the shape of uniformity of assessment standards, both in

criteria and contents. It is this uniformity that Elbow fears will destroy the aspects of validity ("different kinds of writing produced under different conditions") that we value in portfolio assessment. While some aspects of validity are decreased with uniformity of standards, validity is not abolished, and portfolios, even with this decrease, still remain the most valid tool for assessing writing.

While Elbow's desire to abolish reliability is misinformed, his belief in the value of portfolio assessment is not. It seems safe to assume that in order not to lose this exciting new form of assessment, it would be to our advantage to search for ways to increase the reliability of portfolios as an assessment tool. According to White, "We do not have to throw away fairness [reliability] to be honest in our measurement" (Review 538). One way to maintain validity and reliability--honesty and fairness--when portfolios are used as a way to assess mass quantities of writing in a forum where readers have no personal knowledge of the writers is to read and assess portfolios holistically.

With the move from machine-scorable indirect measures of writing such as multiple choice tests to more direct measures of writing such as essay tests, came the need to discover a different way of assessing writing. Holistic scoring was developed in order to fill this need. According to White:

The holistic approach argues against reductionism and denies that the whole is only the sum of its parts . . . It is the most obvious example in the field of English of the attempt to evoke and evaluate wholes rather than parts, individual thought rather than mere socialized convention. (TAW 18-19)

While holistically scored essay exams are an improvement over multiple choice exams in that all of the "parts" of a student's writing are displayed at once in a single "whole" essay (for example, sentence structure, organization, thesis, development), they are still reductive in that they assess only one sample of a student's writing which has been produced under extremely artificial circumstances. Writing ability, then, has been reduced to one-shot, one sample, and one mode. Holistically scored portfolios offer more. The "parts" of a portfolio consist of all of the "parts" of an essay exam, except these "parts" are repeated over and over in a variety of essays which have been produced over a period of time in more natural settings. Writing ability, then, has been expanded to a process, a process which can be used to construct many different types of essays. The "individual thoughts" that White hails as one of the benefits of holistic scoring become even more accessible through holistically scored portfolios.

While the opportunity to assess "wholes" rather than "parts" is an obvious advantage of holistic scoring, it is not the only advantage. Holistic scoring has also the advantage of high interrater, or reader, reliability. In fact, according to Brian Huot, "[Holistic] scoring emerged

as a primary practice solely on the strength of its interrater reliability coefficient" (*Reliability* 204). The interrater reliability that has worked so well for the holistic scoring of essay exams can be adapted to the scoring of portfolios, a practice which can help establish reliability in portfolio assessment. White reiterates this conclusion by posing the following question and answer:

How can we approach reliability in our assessment of portfolios? The obvious answer is to adapt the measures that have led to high reliabilities for essay test scoring: develop a collegial discourse community for assessment, use a scoring guide to describe the measurement criteria, and agree on sample portfolios at different score levels to illustrate the scoring guide.
(*Portfolios* 36)

These "measures" have been part of successful holistically scored essay sessions for decades, and these same measures work to help ensure high interrater reliability for holistically scored portfolios.

In direct correlation with successful holistic scoring sessions is effective reader training. Readers in holistic scoring sessions are not "real readers in natural settings," but rather are readers reading in "artificial settings with imposed agreements" (Elbow, *Ranking* 189). As such, these readers must not rate portfolios simply on their own personal judgments of writing ability. They must learn to rate portfolios based on agreed upon criteria (however, these criteria should not be "imposed" on the readers, but rather developed by the readers). "The procedures used for rating texts holistically can," according to Cherry and

Meyer, "directly affect the reliability of the scores that result. It is well-known that careful training of raters can improve interrater reliability" (120).

According to Charles R. Cooper,

We all know how unreliable ratings of essays can be: a group of raters will assign widely varying grades to the same essay. This phenomenon has been demonstrated repeatedly for decades. It is an incontrovertible empirical fact Curiously, another fact that often is ignored or slighted in discussions of the unreliability of essay grades is that we have known for almost as long as we have know about unreliability that reliability can be improved to an acceptable level *when raters from similar backgrounds are carefully trained.* (18)

Because, the validity and reliability of any assessment depends on the particular purpose of the assessment, careful training must first begin by discussing this purpose. Once the purpose for the assessment is clear, sample papers or portfolios must be selected, read, and discussed by those who will be doing the rating. "In most scoring sessions," according to Cherry and Meyer, "raters review and judge a number of sample texts, discuss the criteria that inform their judgments, and gradually move toward greater agreement about how to score the papers" (121). The criteria must be developed anew for each assessment situation because criteria that work for one assessment purpose may not work for another. Each assessment must have standards that correlate with its particular purpose. Criteria must be developed that balance the needs of the assessment with a reasonable amount of validity and reliability. Because of

this, it is essential that those involved hold responsible holistic standardizing sessions in order to establish the proper criteria for reading reliably.. According to Cherry and Meyer, "Using two or more raters to score papers and conducting careful training sessions are both sound ways to increase the interrater reliability of holistic sessions" (121). In order to create a productive training session with readers who can learn to read with a fair amount of consistency, trainees must learn to work together as a community. "The essential problem," in a holistic scoring session, according to White,

is to create a coherent working community of highly educated specialists, with common goals and procedures to accomplish intense, difficult, boring and (usually) low-paid labor . . . [The] scoring guide should not be presented as a fixed and revealed truth, but as a guide that can always be improved--until the scoring begins . . . In short, the readers must develop a sense of ownership of the [assessment] and the scoring guide if they are to score willingly and reliably. (Holistic 97)

Once the difficulties of establishing this type of community (a community which has the added benefit of creating consequential validity for teachers--Chapter 4) are surmounted, the chance for a reliable reading is increased.

Are the difficulties of establishing reliable readings for holistically scored essays the same as those found in establishing reliable readings of holistically scored portfolios? According to Laurel Black et al. of the Miami University Portfolio Program, the answer is "Some

difficulties are exacerbated by the portfolio approach, while some difficulties are unique to portfolio evaluation" (8). The "crucial difference" lies in the contents of the portfolios. Because portfolios may contain "several different pieces . . . [portfolio] scoring guides simply cannot describe ideal portfolios with the same precision as a scoring guide for a single-sitting assigned topic essay" (Black et al. 8). In fact, Belanoff and Dickson call portfolio assessment, "Messy" (xx). This "messiness" very often causes problems for readers first attempting to use portfolios as an assessment tool. Liz Hamp-Lyons and William Condon discover such problems in their "study of how [teacher-evaluators] handle the cognitive task of making what [they] initially thought would be a 'holistic' judgment of the multiple texts in portfolios in one composition program" (*Questioning* 177). In their attempt to look closely "at the process of reading a portfolio," the two ask the readers to keep a formal log which "attempt[s] to get at such questions as how and when a reader makes the decision about a score on a portfolio . . ." (*Questioning* 179). In their study, Hamp-Lyons and Condon find,

a clear suggestion that readers do not attend equally to the entire portfolio [R]eaders' self-reports indicate that readers arrived at a score during their reading of the first paper [Readers] tend to reduce the load in portfolio reading by finding short cuts to decisions (*Questioning* 182-3)

If readers are "not attend[ing] equally to the entire portfolio," then these readings cannot be considered

consistent, reliable, or fair; in other words, the entire reading is invalid. These findings cause Hamp-Lyons and Condon to conclude that "holistic reading, in the case of portfolios, is highly unlikely, if not impossible"

(*Questioning* 180). While it is easy to understand the desire to reach this conclusion based on the given facts, the conclusion is still a wrong one. To understand the problem with Hamp-Lyons and Condon's study, we must look at their scoring criteria and their training sessions.

At the beginning of the study conducted by Hamp-Lyons and Condon, readers who are "well-trained and experienced in formal writing assessment" gather together to discuss "what the goals of standardizing should be" (*Questioning* 179-80). Generalized criteria are established, and it is decided that readers should look for such things as "evidence of awareness of viewpoints other than the writer's own; complexity in the issues the writer discusses etc" (*Questioning* 179). As the study progresses, more criteria are added; however, there is no specific criterion added that might counsel readers to read holistically. And, even with all of the careful attention to criteria and standardizing, there is never an attempt to rigorously train readers in *how* to read portfolios holistically (as Black et al. do successfully in a study I will address later in this chapter). The readers, because they are not trained to read the portfolios in a consistent and holistic manner, instead read in a way that is completely "helter skelter." It is

impossible to obtain any kind of reliable scores from such a reading.

Hamp-Lyons and Condon conclude from this study that the problem is in the nature of the portfolios. However, it seems clear, upon reflection, that the problem is not caused by portfolios, but rather lies in the nature of reader training, a finding Hamp-Lyons and Condon hint at when they say, in their concluding thoughts, "Our readers have told us over and over that they feel the need for criteria and standards against which to measure portfolios" (*Questioning* 187). The need for criteria and standards is not peculiar to portfolio assessment; it is an essential element of all assessment. Condon and Hamp-Lyons come to this realization in their second study, a study that rebuts the findings of the first. Condon and Hamp-Lyons, convinced that portfolios are worthwhile assessment tools, and concerned that their first study negated the reliability of portfolio assessment because "as portfolio assessment became routine, readers went on a kind of 'automatic pilot,' leading them to become less attentive to the whole portfolio," decided to conduct a second study with the "reformation of the reading groups" (*Maintaining* 278). Condon and Hamp-Lyons state: "We felt that if we could increase the reader's investment in the portfolio . . . then we could change reader's behavior so that they would once again attend to the portfolio as a whole" (*Maintaining* 278). This "investment" is what White refers to as "a sense of ownership," which is necessary in

creating a community of readers who are able to "score willingly and reliably" (*Holistic* 97). This type of community is not established in the first study conducted by Condon and Hamp-Lyons. In the second study, readers are specifically trained to read holistically. "The results from the whole-group standardizing session indicate to [Condon and Hamp-Lyons] that readers considered evidence from all pieces in the portfolio" (*Maintaining* 281). In other words, readers had successfully learned how to read portfolios holistically. Condon and Hamp-Lyons conclude that "restructuring . . . the portfolio reading procedures did change reader's behaviors" (*Maintaining* 282), providing them with "a more stable, more reliable, and more thoughtful assessment" (*Maintaining* 278). Both studies emphasize the necessity of developing criteria and standards that are peculiar to a particular purpose, standards that will help readers read portfolios holistically and reliably. Hamp-Lyons and Condon stress this finding by stating that the "superiority of portfolios as an assessment tool is dependent on readers reading, judging and valuing 'all' the texts" (*Questioning* 183).

Laurel Black, Jeffrey Sommers, Donald A. Daiker, and Gail Stygall prove, in their study, that valuing "all" texts, in other words, reading portfolios holistically and reliably, is possible. Black et al. offer students who are participating in their study, two methods of achieving advanced placement: by submitting a portfolio or by writing

an essay in two hours on campus in response to an assigned prompt. Both portfolios and timed essays were scored on a 1-6 scale (9). This study offers the unique opportunity of comparing the reliability of the two assessment methods. Black et al. conclude from this research "that rating portfolios can be as reliable as rating essays" (9). The following data support this finding:

The numbers [from the portfolio and timed essay scores] are remarkably close; 85.5% of the first and second readers of the essays recorded scores no more than one point different from one another, while 85.8% of first and second readers of portfolios recorded scores no more than one point different from one another. In other words, if an essay or portfolio received a score of 3 from a first reader, 85% of the time the second reader recorded a score of 2, 3, or 4. (9)

These findings continue to be consistent when comparing the third reading of discrepancies; in fact, "[where] the percentages shift slightly, the shift favors portfolios ratings" (10). Black et al. conclude from this data that, "[w]hile it may be too optimistic to suggest that reliability may increase by using portfolios, it seems fair to assert that reliability in portfolio rating can compare favorably to reliability in single-setting essay rating" (10). After analyzing "scoring patterns and raters' own opinions," Black et al. attribute their success to adequate rater training (14-15).

Black et al., unlike Hamp-Lyons and Condon, realize from the very beginning of their study that "genuine holistic rating is more likely to take place if the rating

committee and chief reader foreground its importance and its challenges forthrightly in the scoring guide and the training session" (18). One way to "foreground" holistic reading is to develop a scoring guide that has built in holistic criteria. For example, the description of a "6" in the scoring guide developed for Black et al.'s study "reads in part: 'A portfolio that is excellent in overall quality'" (17). This type of criterion keeps the reader from making a judgment based on the first essay because one cannot judge "overall" quality by one piece of writing. Raters are "explicitly reminded not to score individual pieces but rather to withhold judgment until reading all of the written work," and the nature of the scoring guide helps keep them on the right track (17). Black et al. compare reading a portfolio as a whole to reading an essay as a whole:

Just as rating an essay holistically does not mean assigning a score to each paragraph in the piece and averaging those scores to arrive at a final rating, so rating a portfolio holistically ought not mean assigning scores to each piece and then averaging them. (17)

This analogy serves to point out that unreliable readings of portfolios (such as those in the first study by Hamp-Lyons and Condon) are not inherent in portfolio assessment itself, but rather are the result of poor reader training. "It is vital," according to Black et al.,

to stress to raters that they need to be self-aware, monitoring how they are being influenced by their reading . . . [Readers should] make a conscious effort to avoid assigning scores until they have read the entire portfolio . . . (22)

This "self-awareness," this "consciousness," can be developed through considerate and rigorous reader training.

The study by Black, Sommers, Daiker, and Stygall, along with the second study by Hamp-Lyons and Condon, point out that we can indeed achieve reader reliability in portfolio assessment, reliability that is at least as good as that found in holistic essay assessment. While these studies are enlightening and bode well for the future of portfolios as assessment tools, we should not forget that they are each based on their own particular purpose. Each time portfolios are used as an assessment tool, the purpose for the assessment must be assessed as well. Each time portfolios are used as an assessment tool, the reliability of the tool must be judged anew.

The Stepping Stones Portfolio Project has not yet achieved any real measure of reliability. With vague criteria, student portfolios will have no common basis for assessment. Logically, this means that, at this step, even if students' portfolios are assembled and gathered, they cannot be reliably assessed. Does this mean the project should be abandoned? The answer to this, I believe, can be found in the purpose for the assessment. Stepping Stones Portfolios will eventually be used to fulfill less than 1/6 of a state accountability mandate. The state requires multiple assessments, which the Morongo Unified School Districts has divided into three types: State produced common tests such as the Stanford 9, criterion-referenced

tests such as essay exams based on common prompts, and classroom based assessment. Portfolio assessment, along with subject grades, will be used as part of the classroom-based assessment. Because portfolios are only one small part of this accountability process and because portfolios pull into the mix the only opportunity for authentic assessment, I believe that, in this instance, it is appropriate to use this tool, invalid and unreliable though it may be. However, it is my hope that Stepping Stones and MUSD will continue to look at these issues and to refine the portfolio project until one day we can prove the MUSD portfolios to be both valid and reliable.

Even with all of the demands of establishing reliable and valid portfolio readings, Peter Elbow's confirmation in the value of portfolio assessment as "[rewarding] . . . the essential things we try to place at the heart of our writing courses . . . [such as] exploratory writing . . . discussion with peers and with teacher; feedback on drafts from peers and teacher; and extensive, substantive revision . . ."

makes it clear that portfolio assessment is worth the continuing effort of those involved (*Foreword xv*). The benefits of portfolios as an assessment tool are becoming more and more apparent to those in the composition community. These benefits establish portfolio assessment as a worthwhile endeavor for the community of teachers and learners. Most of all, we who are interested in teaching and assessing writing must believe that we can work together

to establish portfolios as a reliable assessment tool. "If the goals are to be reached," says White, "universities and their faculties will need to believe the measurement and be part of it" (*Assessing* 118). Portfolios, when developed as a reliable and valid assessment tool for a particular purpose, can only benefit all involved.

While portfolios can be reliable and valid tools for measuring writing, they are not the only tools that can make this claim. Portfolios are not always the best or most feasible tool for measurement. The validity and reliability of any measurement tool is dependent on purpose. The purpose for each assessment must be considered when selecting a measurement tool. This consideration must also include cost effectiveness. In other words, the question must be asked, "Is the expense of a particular assessment method worth the results?" Portfolio assessment can be costly and therefore is not always a viable alternative. The next chapter will examine the feasibility and cost-effectiveness of portfolio assessment in relation to issues of validity, reliability, administrative needs and the conflicts that emerge when these needs diverge.

CHAPTER 6

FEASIBILITY: Overpasses and Tunnels, or "Getting There"

Issues of validity and reliability in assessment, while extremely important, are not the only determiners of whether or not a particular assessment tool should be used. We, as writing specialists, must determine, for each assessment opportunity, the feasibility of the assessment tool; we must ask ourselves, "Is this assessment method 'doable'?" Feasibility is unique to each assessment; it is an outgrowth dependent on purpose, use, and context. "[No] assessment," according to White, "exists outside of its context, its uses and its effects; no tests or assessment systems have value in themselves" (*Apologia* 33). An assessment method must be chosen, not only for reasons of validity and reliability, but also for reasons of cost and time in relation to value gained.

Assessment is big business. Large scale assessment can be very costly, and we must be certain that money put into assessment is money well spent. This does not mean, however, that we should always use the cheapest assessment tool possible. According to White:

When we deal with the issue of cost, we need to point out that cost by itself is meaningless. What matters is *cost effectiveness*, what we get for the money spent. Something that is cheap and useless is less cost-effective than something that is expensive but highly useful. (*Apologia* 43)

The three main tools for assessing writing: multiple-choice

tests, essay exams, and portfolios, vary in expense:

"[Essay] scoring costs about five times as much as multiple-choice testing and portfolios cost about five times as much as essays" (White *Apologia* 43). However, these tools also vary a great deal in the type of information they supply and in how that information is obtained. Because portfolios can contain a variety of writing modes and are developed through authentic classroom work, many advocates such as Peter Elbow believe that the high cost of portfolio assessment is *always* worth it. Brian A. Huot, however, offers a note of caution:

The need to consider the validity of portfolios for a specific testing situation is necessary to ensure the continued use of portfolios as an assessment instrument. If we overuse portfolios, convincing those who hold the purse strings to invest unnecessarily, we run the risk of having them branded as another educational fad. (*Beyond* 332)

Huot's warning serves to point out that an assessment method does not have value in itself; its value depends on its ability to be the best choice for a particular assessment purpose. The best assessment method may not always be the most costly. White reiterates Huot's concern by stating that "with any complex and high quality operation, the financial and human cost is high enough to make reasonable people ask if every assessment should be a portfolio assessment" (*Apologia* 39). Reasonable people will choose a writing assessment tool by making sure that it not only fulfills all of the requirements for validity and

reliability, but also fulfills all of the criteria for cost-effectiveness.

Whether or not an assessment costs too much depends on the purpose for the assessment. In the case of Stepping Stones, for example, portfolio assessment seems a good choice. MUSD made the decision that their writing assessment tool should not only give them information to report to the state (which is requiring multiple assessments), but should also validate their students' classroom experiences. Since California standards require that students learn writing through a process of revision and editing, it makes sense to support an assessment that measures this construct of learning. It is true that portfolio assessment will cost more money. Teachers must be trained, and they must be given time to participate in scoring sessions, which will incur costs as well. It remains to be seen whether or not MUSD continues to support portfolio assessment as the issues become more complex and more expensive. While MUSD, along with other K-12 districts, struggles with the challenges of portfolio assessment, universities are also struggling with similar challenges in looking for ways to positively utilize this type of assessment.

"Most writing assessment in higher education," says Peter Elbow, "is for placement . . ." (*Virtues* 51).

Universities, when determining which assessment device to use for placement exams, must find a balance between costs

and results (portfolio scoring sessions can be particularly expensive). Elbow asks the following question concerning the expense of portfolio assessment in relation to placement exams: "Is it really useful to spend such extraordinary amounts of time and money in order to move some students into a remedial course or to exemption?" (*Virtues* 51)?.

(Elbow does not view placement exams as "significant" enough to merit the use of portfolio.) Portfolios take more time to score and assess than do essay exams; time invariably equals money. White also questions the cost-effectiveness of scoring writing portfolios, "[Portfolios] need to demonstrate that they can be assembled and scored at reasonable cost in time and money before they can command the respect of the assessment community or of higher education generally" (*Portfolios* 38). Elbow's and White's comments invite caution; they encourage us not to jump on what Edward Kearns calls "the running board of the portfolio bandwagon" (50), without first asking ourselves whether or not portfolio assessment is the best, or as Elbow puts it, the most *useful* tool for each particular assessment.

I have discussed how the study by Black, Sommers, Daiker, and Stygall helps prove that portfolios can be assessed as reliably as essay exams (*Chapter Four*); however, this study does not necessarily prove that portfolios are the best tool for the job because it does not address issues of cost-effectiveness. In fact, the study

finds that the correlation between essay reliability and portfolio reliability is more or less equal. Does that not mean, then, that essay exams are just as effective as portfolios--maybe even more cost-effective because they are cheaper to score? The answer to this lies not only in the purpose for the assessment, which is to award students advanced placement, but also in the university's definition of assessment. According to Ryan and Miyasaka:

Assessment is increasingly seen as an integral part of the teaching-learning process, not merely as an activity used for accountability purposes. Viewed from this perspective, assessment is not seen as a decontextualized, objective process from which the influence of teachers should be removed. Rather, teachers are seen as the cerebral and most important assessors in the lives of students, and assessment is seen as one of the important tools teachers use to facilitate learning. (10)

In the study by Black et al., students in the portfolio group worked along with their teachers in developing their writings for assessment; students in the essay group did not. If students in the portfolio group gained more from the study (for example, becoming metacognitive about their processes of writing and their learning styles) than did the students in the essay group, does this mean that portfolio assessment, in this instance, was more cost-effective (more "bang for the buck") than essay exams? In truth, the answer can be either "yes" or "no" depending on the consensus of the assessors. However, the point remains that cost-effectiveness, or feasibility, has only partly to do with actual costs; it has very much to do with the university's

definition of assessment and its commitment to this definition.

Ryan and Miyasaka's statement concerning assessment and the "teaching-learning" continuum reiterates the argument in Chapter Three concerning consequential validity: Should assessment be a natural outgrowth of teaching or should it be a separate entity? Should an assessment produce positive consequences for a community of teachers and learners or does it exist simply to supply data for instructors, administrators and accountability groups? These are questions that must be asked, along with all of the previously implied questions concerning validity, reliability etc., when determining the cost-effectiveness of an assessment tool. Perhaps Brian Huot sums it up best in his discussion of institutional validity:

Ultimately, the decision about which method of assessment is best in a given situation can only come from within an institution. A measurement should have institutional validity in that it should be sensitive to the needs of particular students, teachers and educational programs that are part of the teaching and learning environment of a particular institution. (*Beyond* 332)

If we agree with Huot, then we must add "sensitivity" to our list of considerations--cost, results, usefulness, validity, and reliability--for a feasible assessment tool. We must, in determining the *best* tool for the job, decide which of these considerations (some or all) are important, and we must do this at each assessment opportunity.

While Elbow and White both caution that portfolio

assessment can be very costly and perhaps not always worth the expense, it is worthwhile to keep in mind that Elbow also says that portfolios can reward "the essential things we try to place at the heart of our writing courses . . . " (*Introduction* xv), and White says that portfolios "offer to the world of assessment a view of student learning that is active, engaged, and dynamic . . ." (*Portfolios* 27). These statements do not necessarily clash with their statements of concern. They simply serve to point out the elusive nature of cost-effectiveness and assessment. We cannot make blanket decisions concerning any aspect of assessment. There is no one *right* or *wrong* tool for assessment; there is only the most feasible assessment tool for a particular assessment. And those who care about teaching, learning, assessment, and students, need to take the time to learn about and understand the issues involved in *all* types of writing assessment.

"Portfolio assessment," according to Belanoff, "brings people together to create a literate environment" (21). Students can see the connection between what they are taught about writing and the way they are being assessed. They are afforded the opportunity to become active participants in their own assessment, and, from this opportunity, they can learn how to assess themselves. Teachers can see the connection between what they are teaching and how their students are being assessed:

[Portfolio] assessment is attractive to teachers because it "rewards" rather than punishes the essential things we try to place at the heart of our writing courses: exploratory writing, in which the writer questions deeply and gets lost; discussion with peers and with teacher; feedback on drafts from peer and teacher; and extensive, substantive revision." (Elbow *Introduction* xv)

Portfolio assessment offers rewards that go beyond an external measurement. Portfolio assessment can bring together writing specialists, teachers, and students so that the three are part of a continuum, a continuum that incorporates assessment, teaching, and discovery into opportunities which can have positive lifelong learning implications for teachers and students alike.

CONCLUSION: The Light at the End

The Stepping Stones' narrative discussed in Chapter One embodies the process of the genesis of a portfolio assessment plan. While the goals of the project have yet to be fully realized, the process has resulted in a district-wide conversation about how writing is taught, assessed, and the importance of accountability. District administrators and teachers have worked in consort to develop a broad consensus for Stepping Stones' first steps. Many of Stepping Stones' participants serve also on key committees throughout the district. Site principals and faculty are involved in six separate Program Quality Reviews that are focused on language arts, assessment, and accountability, all of which involve Stepping Stones' coordinators and participants. Clearly, while the destination of large scale portfolio assessment has yet to be achieved in the Morongo Unified School District, the journey has had a revitalizing effect on teaching and learning in all content areas. At the same time, by encouraging teachers and administrators to collaborate at all levels, at their sites and across the district, a more closely knit community of educators is emerging, teacher professionals who share a voice in the operation of their district, a voice that is less isolated, and more accountable. Whether or not the Stepping Stones' portfolio plan is fully implemented, those of us involved believe that the experience thus far has certainly been

worth the investment of time and money.

Hopefully, the excitement and interest shared by Stepping Stones' participants can be communicated to other teachers, administrators, parents and students at individual sites and throughout our communities. With continued support and continued effort, the Stepping Stones' Project can grow into a viable assessment plan, a plan which has been developed and implemented by those involved in the teaching of writing. Thus, teachers and students, rather than being on the periphery of an assessment which makes determinations about their fates, can become, instead, participants in their own assessments and determiners of their own fates.

Works Cited

- Belanoff, Pat. "Portfolios and Literacy: Why?" New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, and Gail Stygall. Portsmouth: Boynton/ Cook Inc., 1994. 13-24.
- Belanoff, Pat and Marcia Dickson. "Introduction." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. xix-xxiv.
- Belanoff, Pat and Peter Elbow. "Using Portfolios to Increase Collaboration and Community in a Writing Program." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 17-36.
- Black, Laurel, Sommers, Jeffrey, Donald A. Daiker, and Gail Stygall. "The Challenges of Rating Portfolios: What WPAs can Expect." Writing Program Administration 17 (F/W 1993): 7-29.
- Bruffee, Kenneth A. "Collaborative Learning and the 'Conversation of Mankind'." College English 46 (Nov. 1984): 635-652.
- Caccaveri, Peter. "Collaborative Teaching as a Model for Student Writing." The Compost Heap 2 (Spring 1992): 49-53.
- Camp, Roberta. "Changing the Model for the Direct Assessment of Writing." Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations. Ed. Michael M. Williamson and Brian A. Huot. Cresskill: Hampton Press, 1993. 45-78.
- Camp, Roberta and Denise Levine. "Portfolios Evolving: Background Variations in Sixth- Through Twelfth-Grade Classrooms." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 194-205.
- Cherry, Roger D. and Paul R. Meyer. "Reliability Issues in Holistic Assessment." Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations. Ed. Michael M Williamson and Brian A. Huot. Cresskill: Hampton Press, 1993. 109-141.

Claggett, Fran. A Measure of Success: From Assignment to Assessment in English Language Arts. Portsmouth: Boynton Cook Inc., 1996.

Condon, William and Liz Hamp-Lyons. "Maintaining a Portfolio-Based Writing Assessment: Research That Informs Program Development." New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring. Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, Gail Stygall. Portsmouth: Boynton/Cook Inc., 1993. 277-285.

Cooper, Charles R. "Holistic Evaluation of Writing." Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Teachers of English (NCTE), 1974. 3-31.

Daiker, Donald A., Jeffrey Sommers, and Gail Stygall. "The Pedagogical Implication of a College Placement Portfolio." The Politics and Policies of Assessment in Writing. Ed. Sandra Kamasukiri, William Lutz, and Edward M. White. New York: Modern Language Association. 1996.

Elbow, Peter. "Foreword." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. ix-xvi.

---. "Ranking, Evaluating, and Liking: Sorting out three Forms of Judgment." College English 55 (Feb. 1993): 187-206.

---. "Will the Virtues of Portfolios Blind Us to Their Potential Dangers?" New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring. Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, and Gail Stygall. Portsmouth: Boynton/Cook Inc., 1994. 40-55.

Elbow, Peter and Pat Belanoff. "State University at Stony Brook Portfolio-based Evaluation Program." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 3-16.

Hamp-Lyons, Liz and William Condon. "Questioning Assumptions about Portfolio-Based Assessment." College Composition and Communication (CCC) 44 (May 1993): 176-190.

Gere, Anne. Writing Groups: History, Theory, and

- Implications. Carbondale: Southern Illinois University Press, 1987.
- Huot, Brian. "Beyond the Classroom: Using Portfolios to Assess Writing." New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring. Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, and Gail Stygall. Portsmouth: Boynton/Cook Inc., 1991. 325-333.
- . "Reliability, Validity and Holistic Scoring." CCC 41 (May 1990): 201-213.
- Kearns, Edward. "On the Running Board of the Portfolio Bandwagon." WPA: Writing Program Administration 16 (Spring 1993): 50-58.
- Larson, Richard L. "Portfolios in the Assessment of Writing: A Political Perspective." Assessment of Writing: Politics, Policies, Practices. Ed. Edward M. White, William D. Lutz, & Sandra Kamusikiri. New York: MLA., 1996. 271-283.
- Messick, Samuel. "Standards of Validity and the Validity of Standards in Performance Assessment." Educational Measurement: Issues and Practices (Winter 1995): 5-8.
- Mills-Courts, Karen and Minda Amiran. "Metacognition and the Use of Portfolios." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 101-112.
- Murphy, Sandra and Mary Ann Smith. "Looking into Portfolios." Portfolios in the Classroom. Ed. Kathleen Blake Yancey. Urbana, Ill.: NCTE, 1992. 49-60.
- North, Stephen M. The Making of Knowledge in Composition: Portrait of an Emerging Field. Portsmouth: Boynton/Cook Inc., 1987.
- Popham, James W. "Circumventing the High Costs of Authentic Assessment." Phi Delta Kappan 74 (Feb. 1993): 470-473.
- Roemer, Marjorie, Lucille M. Schultz and Russel K. Durst. "Portfolios and the Process of Change." CCC 42 (Dec. 1991): 455-469.
- Ryan, Joseph M. and Jeanne R. Miyasaka. "Current Practices in Testing and Assessment." Bulletin: The National Association of Secondary School Principals (Oct. 1995): 1-10.

- Sawyer, Mary H. "The Portfolio Research Project: A Successful School/University Collaboration." English Journal 88 (September 1998): 64-67.
- Smith, Cherryl Armstrong. "Writing Without Testing." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 279-291.
- Stygall, Gail, Laurel Black, Donald A. Daiker, and Jeffrey Sommers. "Introduction." New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring. Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, and Gail Stygall. Portsmouth: Boynton/Cook Inc., 1994. 1-9.
- Trimbur, John. "Response: Why Do We Test Writing?" Assessment of Writing: Politics, Policies, Practices. Ed. Edward M. White, William D. Lutz, & Sandra Kamusikiri. New York: MLA., 1996. 45-48.
- Wauters, Joan K. "Evaluation for Empowerment: A Portfolio Proposal for Alaska." Portfolios: Process and Product. Ed. Pat Belanoff and Marcia Dickson. Portsmouth: Boynton/Cook Inc., 1991. 57-68.
- White, Edward M. "Apologia for the Timed Impromptu Essay Test." CCC 46 (Feb. 1995): 30-45.
- . "Assessing Higher-Order Thinking Skills in College Graduates Through Writing." The Journal of General Education 42 (1993): 105-122.
- . Assigning, Responding, Evaluating: A Writing Teacher's Guide. New York: St. Martin's Press, 1992.
- . "Holistic Scoring: Past Triumphs, Future Challenges." Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations. Ed. Michael M. Williamson and Brian A. Huot. Cresskill: Hampton Press Inc., 1993. 79-108.
- . "Portfolios as an Assessment Concept." New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large Scale Scoring. Ed. Laurel Black, Donald A. Daiker, Jeffrey Sommers, and Gail Stygall. Portsmouth: Boynton/Cook Inc., 1994. 25-39.
- . "Power and Agenda Setting in Writing Assessment." Assessment of Writing: Politics, Policies, Practices. Ed. Edward M. White, William D. Lutz, & Sandra Kamusikiri. New York: MLA., 1996. 9-24.

- . Rev. of Portfolios: Process and Product, ed., Pat Belanoff and Marcia Dickson. CCC 43 (Dec. 1992): 537-539.
- . Teaching and Assessing Writing: Recent Advances in Understanding, Evaluation, and Improving Student Performance. San Francisco: Jossey-Bass, 1985.
- . Writing Assessment Seminar, Dept. of English. California State University, San Bernardino, 24 Jan. 1994.

Williamson, Michael M. "An Introduction to Holistic Scoring: The Social, Historical and Theoretical Context for Writing Assessment." Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations. Ed. Michael M. Williamson and Brian A. Huot. Cresskill: Hampton Press, 1993. 1-43.