DOCUMENT RESUME

ED 324 329 TM 015 516

AUTHOR Paulson, Leon F.; Paulson, Pearl R.

TITLE How Do Portfolios Measure Up? A Cognitive Model for Assessing Portfolios. Revised.

PUB DATE 15 Aug 90

NOTE 32p.; Paper presented at the Annual Meeting of the Northwest Evaluation Association (Union, WA, August 2-4, 1990).

PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Academic Achievement; *Cognitive Processes; *Educational Assessment; Elementary Secondary Education; Models; Outcomes of Education; *Portfolios (Background Materials); Qualitative Research; Statistical Analysis; *Student Developed Materials; Student Evaluation

IDENTIFIERS Aggregation (Data); Stakeholder Evaluation

ABSTRACT

Issues that must be addressed in designing procedures for aggregating portfolio data are considered. These issues have profound implications for what is aggregated and how data are collected, combined, and interpreted. Portfolio assessment occurs at the intersection of instruction and assessment; it requires students to collect and reflect on examples of their work, providing both an important instructional component to the curriculum and offering an opportunity for complex authentic assessments. A multidimensional, cognitive process model of assessment is recommended in this paper. The cognitive model for assessing portfolios is structured much like the program evaluation model developed by R. Stake (1967). The model is designed to be broadly descriptive, yet to provide a framework for presenting high quantitative and qualitative data in a coherent fashion. Its results are judgments taken from the perspectives of multiple parties. The model does not provide a single score; rather, it provides a comprehensive view of complex learning outcomes in a context where instruction and assessment are inseparable. The three-dimensions of the model are the following: (1) activity, including assessment activities intrinsic to both compiling a portfolio and aggregating across portfolios; (2) historical, including changes beginning from conditions at the outset through transactions to outcomes; and (3) stateholder, which reflects viewpoints of stakeholders with an interest in the portfolios. Three sample portfolio projects are outlined: the Pupil Product Portfolio; a writing portfolio; and a metacognitive letter. Five figures and a 65-item list of references are included. (TJH)

# How Do Portfolios Measure Up?

## A Cognitive Model for Assessing Portfolios

by

F. Leon Paulson
&
Pearl R. Paulson

Revised 8/15/90

This paper was presented at a conference *Aggregating Portfolio Data* sponsored by the Northwest Evaluation Association in Union, WA, August 2-4, 1990.

F. Leon Paulson is an Assessment Specialist, Multnomah ESD, Portland OR.
Pearl R. Paulson is a Student Services Coordinator, Beaverton (OR) School District.

# Contents



# List of Figures

*Many millions of years ago a race of hyperintelligent pandimensional beings got so fed up with the constant bickering about the meaning of life that they decided to sit down and solve their problems once and for all. And to this end they built themselves a stupendous super computer which was so amazingly intelligent that even before its entire data banks had been connected up it started from "I think therefore I am" and got as far as deducing the existence of rice pudding and income tax before anyone managed to turn it off.*

[The builders confronted the computer, Deep Thought, with their ultimate question.] *"...the task we have designed for you to perform is this. We want you to tell us the Answer...to Life...The Universe...Everything."*

*"Yes," said Deep Thought, "I can do it."* [To be continued...]

-- Douglas Adams *The hitchhiker's guide to the galaxy*

Parents, school boards, state departments, and federal agencies are calling for accountability in education. Usually, they call for test scores. Are they up? (Up is good!) Down? (Down is bad!)

Soon, the evaluation community will be called on to use portfolios for accountability. The portfolio has the potential to be a powerful educational tool for encouraging students to take charge of their own learning (Swing, Stoiber & Peterson, 1988). Portfolio assessment requires students to collect and reflect on examples of their work, providing both an important instructional component to the curriculum and offering an opportunity for complex, authentic assessments.

Portfolio are an intersection of instruction and assessment -- they are neither instruction nor assessment, they are both instruction and assessment. Portfolio assessment forces us to adopt alternative ways of thinking about what we do and how we measure. To answer questions of accountability, we must aggregate but without corrupting the portfolio's instructional benefits.

We begin our paper with several issues that must be addressed when designing an aggregation strategy. Next, we contrast the unidimensional assessment of achievement tests with multidimensional assessment from a cognitive process perspective. We recommend that portfolio assessment requires a multidimensional, cognitive process model. We propose such a model, CMAP. We conclude with a review of methodologies appropriate to aggregating portfolio data and illustrate the model in use.

## Issues in assessing portfolios

Our paper is not a recipe for aggregating portfolio data. Rather, it explores issues that must be addressed when designing procedures for aggregating data while preserving the integrity of the portfolio itself. These issues have profound implications for what is aggregated and how data are collected, combined, and interpreted.

- **Aggregation.** A dictionary defines an aggregation as "a total or whole: a group of distinct tnings gathered together." If we view aggregation as ways to report about distinct things, we go about our task one way. If we view it as ways of representing commonalities, we go about our task in another. In aggregation, is diversity welcome, or a nuisance?

- **Knowledge.** Education is often thought of as a means through which students acquire knowledge. But what do we really mean by the term *knowledge*? Is it something we *have*, or is it something we *do*? Is it something educators give, or is it something students construct? This epistemological debate between structuralism and constructivism has profound implications for how we go about aggregating portfolio data.

- **Authenticity.** We hear the call for authentic assessment from almost every corner. There is an assumption that authentic assessment techniques like portfolio assessment are going to tell something different from the assessment techniques already in place. What kinds of information does authentic assessment offer not found in traditional techniques?

- **Scope.** Should portfolio assessment try to replace traditional assessment, or should it focus on complex performance not tapped by traditional methods? Is it all things to all people, or is it specialized?

- **Standardized Contents.** As evaluators, we routinely look for ways to standardize the conditions under which we observe in order to reduce error of measurement and produce high reliability and validity coefficients. Portfolios celebrate diversity -- each portfolio is highly personal in nature. Standardizing portfolio contents changes the nature of the portfolio itself. Should we be willing to pay the price of modifying the portfolio as an instructional tool to obtain better statistics?

- **Measurement Standards.** While educators often use the terms reliability and validity in a narrow, quantitative sense, the Joint AERA, APA, NCME Committee for the Review of the Standards for Educational and Psychological Tests (1985, see pages 9 and 19) defines them broadly in relation to purpose and context. In portfolio assessment, should we be guided by standards often applied to multiple choice tests, or should we rethink the ways we accumulate evidence that portfolio assessments are consistent (reliable) and accurate (valid)?

- **Observer Impact.** One of the profound realizations of 20tn century science is that the act of observing changes the thing observed. In education, assessment has major impact on the educational experience itself. Because what is assessed becomes valued, should we look at best performance, typical performance, or deficit performance?

- **Empowerment.** A major instructional goal of portfolios is metacognitive, to help students develop their capacity for self-reflection and making judgments. The portfolio encourages the student to develop a set of values, assess their work according to those values, celebrate when they meet the expectations implied in those values, and develop new directions for themselves. How can we empower both aggregators and students so that the judgments of one do not invalidate the judgments of the other?

## The traditional approach to measuring achievement

Traditionally, educators think of measuring achievement by using norm or criterion referenced achievement tests. As a rule, achievement tests are built on the assumption of unidimensionality, that is, what is measured occurs on a single, underlying dimension. Achievement is the knowledge base growing bigger fact by fact. This assumption has a profound impact on the tests we develop. We use mathematical models like classical test theory and item response theory to refine our tests and control how they work. Achievement becomes defined by what our achievement tests measure. We produce tests that conform to our models, then we use those same models to demonstrate that those tests are reliable and valid. But the nagging question remains -- reliable and valid measures of *what?* Has our ability to derive equations surged ahead of our understanding of learning and cognition? John Tukey (1972), the mathematician, has said,

> It is well to understand what you can do before you learn to measure how well you seem to be able to do it (p. 52).

Our unidimensional theories of testing may be dictating *defacto* theories of learning that we use to drive instruction (see Shepard's, 1990 discussion). These theories assume that students first master basic skills, then go on to higher ones. *Higher order thinking skills*, a designation frequently used to identify worthy instructional goals, implies that thinking is just another skill to be taught. But the dimensional model simply does not fit the complexity of human behavior in context. Resnick and Resnick (1989) observe:

> One of the most important findings of recent research on thinking is that the kinds of mental processes associated with thinking are not restricted to an advanced or "higher order" stage of mental development. Instead, thinking and reasoning are intimately involved in successfully learning even elementary levels of reading, mathematics, and other school subjects. Cognitive research on children's learning of basic skills reveals that reading, writing, and arithmetic -- the three Rs -- involve important components of inference, judgment, and active mental construction. The traditional view that the basics can be taught as routine skills, with thinking and reasoning to follow later, can no longer guide our educational practice. (p. 4)

The fact that the nature of thinking has been misinterpreted as *the sum of its parts* may be the reason that many of us have been frustrated in our attempts to develop paper and pencil tests of thinking. Shavelson, Carey & Webb (1990) put it this way:

> ...the response format of multiple-choice tests, despite its efficiency, actually prevents us from measuring some of the things that we consider most important. For example, educators want to know how well a student formulates problems and develops answers, not simply whether the student selects the correct alternative....
>
> Moreover, our culture has become so accustomed to achievement testing that citizens do not even ask how achievement test scores were derived -- but only who is on top and who is on the bottom. Achievement test scores fit the American belief in a single dimension of ability that distinguishes winners from losers. (p 693)

If thinking does not lend itself to unidimensional measurement, then neither should the evaluation of thinking be constrained by standards derived for unidimensional measurement.

Portfolio assessment occurs at the intersection of instruction and assessment. The relationship is multiplicative: change one and you change the other, set either to zero and both are lost. If achievement test methodology imposes a unidimensional viewpoint on the interaction, then instruction will also be compromised.

## A cognitive approach to defining achievement

One thing handy about a unidimensional learning theory is that there is a limit to the possibilities. Cognitive theory, on the other hand, can be bewildering. Yet there are important themes that recur, ones that we should take into account in the way we instruct and measure.

One theme is that there is more than one dimension to mental processes that must be considered. Howard Gardner (1983) says there is a combination of seven kinds of intelligence relevant to designing educational programs. The combination includes visual-spatial, bodily-kinesthetic, musical-rhythmical, interpersonal, intrapersonal, logical-mathematical, and verbal-linguistic. Gardner makes the case for the multidimensional nature of human ability by noting that

> the "linguistic" sensitivity to the sounds and construction of language is exemplified by the poet, whereas the interpersonal ability to discern and respond to the moods and motivations of other people is represented in the therapist. Other occupations more clearly illustrate the need for a blend of intelligences. For instance, surgeons require both the acuity of spatial intelligence to guide the scalpel and the dexterity of the bodily-kinesthetic intelligence to handle it. (Gardner & Hatch, 1989, p. 5)

A second theme is that so called higher functioning develops parallel with, not out of, lower level functioning. The passage from Resnick & Resnick (*op.cit.*) reflects this view. One attempt to make sense out of how humans think has come from the study of artificial intelligence (AI). AI researchers believe that if we understand how people think, we should be able to program a computer to mimic the process (see McCordick, 1972). They report that the so called higher order thinking skills are surprisingly easy to program, yet the approach is really not very helpful in simulating intelligent behavior (see Rose, 1984; Schank, 1984). Minsky (1984) offers this comment:

> It is interesting to note that some of the earliest [AI] computer programs excelled at what people consider to be "expert" skills A 1956 program solved hard problems in mathematical logic, and a 1964 program solved college-level problems in calculus. Yet not until the 1970s could we construct robot programs that could see and move well enough to arrange children's building blocks into simple towers and playhouses. Why could we program computers to do expert things before we could program to do childish things? The answer may seem paradoxical. much of "expert" adult thinking is actually simpler than what's involved when ordinary children play.
>
> To be considered "expert", one needs a large amount of knowledge of only a relatively few varieties In contrast, an ordinary person's "common sense" involves a much larger variety of different types of knowledge -- and this requires more complicated management systems. (1986 p 72 )

These words should have profound impact on those performing educational evaluations. There is considerable difference between looking for straightforward, bottom-line changes on basic skills measures and dealing with multiple intelligences and cognitive management strategies.

A third theme is that many kinds of processes combine to produce outcomes. In his classic 1963 paper on criterion-referenced testing, Robert Glaser wrote,

> Underlying the concept of achievement measurement is the notion of a *continuum* of knowledge acquisition ranging from no proficiency at all to perfect performance. (1971, p. 7, emphasis added)

Twenty-seven years later, he describes major aspects of competence as

> ...the compiled, automized, functional, and proceduralized knowledge characteristic of a well-developed cognitive skill, the effective use of internalized self-regulatory control strategies for fostering comprehension, and the structuring of knowledge for explanation and problem solving (Glaser, 1990, p 29f)

Instead of a continuum of knowledge acquisition, Glaser's more recent quote talks about metacognition and knowledge-as-process.

Finally, some investigators have begun to examine the highly complex ways in which cognitive processes are influenced by context. Consider the question of Aptitude-Treatment Interactions (ATI). ATI is assumed when testing specialists interpret the meaning of achievement test scores and their implications for classroom instruction. In 1957, Cronbach wrote that a complete science of psychology is possible if researchers looked at interactions. Later, Cronbach and Snow (1977) analyzed hundreds of research and evaluation reports that examined interactions. While concluding that ATIs are indeed very important, they are also very complex, often highly context specific, and difficult to generalize. Cronbach observed that looking at interactions is like entering a "hall of mirrors" that extends to infinity (1975, p. 119). Later (1988), he updated his simile to reflect the language of the emerging science of chaos:

> ...t's like walking through a maze where walls rearrange themselves with every step you take. (1988, p. 47, quoting Gleick, 1987 p. 24.)

Then, putting it into a cognitive framework, he continued

> ...serious cognitive scientists can no longer model the mind as a static structure. They recognize a hierarchy of scales, from neutron upward, providing an opportunity for the interplay of microscale and macroscale so characteristic of fluid turbulence and other complex dynamical processes. (1988, p. 47, quoting Gleick, 1987. p 299)

## A cognitive approach to assessing portfolios

Through assessment we attempt to discover what is really being achieved -- the educational bottom line. By using test theory models, we attempt to improve on our ability to measure what is achieved. But, whenever we go after objectivity we make subjective choices to get there. We accumulate achievement test scores in the hope that they correspond to "objective" reality. However, to know whether our observations correspond to what is real externally, Eisner (1990) tells us that we need two things. First, we need to know what reality actually is, and second, we need to know what we observe. But all we actually know is what we observe; we never know reality directly. And what we observe depends on what we happen to be looking at and why. It is a product of the evaluator's own perspective, and no two people have quite the same perspective. O. J. Hardison (1989) explains the dilema like this:

> The problem of the world seen from different perspectives is both familiar and profound To someone wearing dark glasses the world looks dark, but it is not necessarily dark. Take off the glasses and the world is flooded in intolerable light. One kind of mathematics reveals one kind of pattern: another reveals another The brilliance of the world becomes yellow with yellow lenses, blue with blue lenses. (p.46)

What we see when we evaluate a portfolio is the product of the glasses we wear when we evaluate portfolios. If we wear glasses designed to reveal mathematical thinking as the sum of basic skills, then that may be all we see. And by using these glasses we may never notice those aspects of mathematical thinking that led the French mathematician Henri Poincare to assert that mathematical thinking may have as much to do with aesthetics as formal logic (Pappert, 1980).

Donmeyer (1990) warns that our theories

> ...can easily become reality for those who employ them. When this occurs, other conceptions of reality are not even considered; indeed, the possibility that alternative conceptions of reality exist is normally not even recognized. (p. 182)

## A cognitive model for assessing portfolios

The Cognitive Model for Assessing Portfolios (CMAP) is structured much like the program evaluation model developed by Robert Stake (1967). CMAP's *activity* dimension uses Stake's words *rationale*, *intents*, *contents*, *standards*, and *judgments* and CMAP's *historical* dimension uses Stake's *antecedent*, *transaction*, and *outcome*. Although we use many of the same words, we have redefined the concepts consistent with portfolio rather than program assessment.

The model is designed to be broadly descriptive, yet provide a framework for presenting both quantitative and qualitative data in a coherent fashion. It also provides results that are the judgments from the perspectives of multiple parties. The model does not necessarily produce convergence in a single, bottom-line, snapshot kind of result. Rather, its purpose is to provide a comprehensive view of complex learning outcomes in context. Stake (1975) says

> More ambiguity rather than less may be needed in our reports. Oversimplification obfuscates (p. 23)

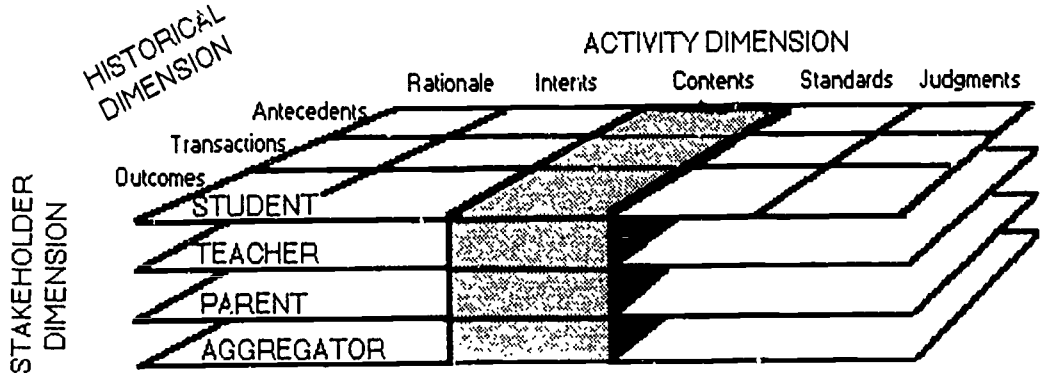


Figure 1. The cognitive model for assessing portfolios showing the activity, historical, and stakeholder dimensions.

The model has three dimensions, each reflecting a major aspect of the portfolio. The first dimension is *activity*. This dimension includes assessment activities intrinsic to both compiling a portfolio and aggregating across portfolios: rationale (what is the purpose for forming the portfolio), intents (what, in general terms, are the goals), contents (the materials found in the portfolio itself), standards (what is good and not so good performance), and judgments (what conclusions we draw). The second dimension is *historical*. The historical dimension is sensitive to changes over time: antecedents (conditions at the outset), transactions (what is occurring over the the time span covered by the portfolio), and outcomes (conditions at present). Finally, the third dimension, *stakeholder*, reflects viewpoints of the various groups with an interest in the portfolio program. The identity of the stakeholders may vary. The student is the central one but stakeholders may also include teachers, parents, district specialists, representatives of the community, the aggregator, and others.

Figure 1 represents all three dimensions of the model and provides a reference as we describe the various rows, columns, and planes within the model.

## The *activity* dimension

The activity dimension relates to the operations involving putting together portfolios.

**Rationale** describes the philosophical basis and operational guidelines for collecting materials and putting them into a portfolio. The rationale varies with the stakeholder in question. The student's rationale may be similar and different from the teacher's (or any other stakeholder's) rationale in interesting and important ways.

**Intents** describe the areas to be represented by the portfolio. These may be the goals and objectives that indicate the curricular scope of the portfolio (*e.g.*, "math and science"); they may be stated operationally (*e.g.*, "demonstrate improvement in..."); they may show what is valued (*e.g.*, "originality", "conceptualization", "accuracy"). Intents may reflect an interest in unintended as well as intended effects; the value of some work is discovered only in retrospect (see Scriven, 1972, on goal-free evaluation). Intents should be stated in *middle distance* (Hardison, 1989) terms so that they are easy to recognize and work with (Valencia. 1990). The statement of portfolio intent plays a key role in portfolio assessment by providing the basic organization and outline for indexing contents.

**Contents** are the exhibits, the actual things found in the portfolio[1]. Contents are primarily students' work. These may include classroom assignments (along with student comments) as well as work students developed especially for their portfolios. The materials that form the portfolio contents tell us much about the level of conceptual understanding (*cognitive fidelity*) and the relevance of the performance to context (*process relevance*) (see Shavelson, Carey & Webb, 1990). A student can demonstrate an understanding of a concept in many ways and the cognitive model permits the student maximum freedom in choosing what they may put into their portfolios.

**Standards** tell how well students should perform; they are the performance criteria and may be relative or absolute. Relative criteria may compare a student's earlier and later work, or an individual student with other students in some appropriate reference group (as with standardized achievement tests). Absolute criteria reflect judgment about what is acceptable performance on an externally established standard (*e.g.*, judgment of Olympic style competition).

---

1 In developing guidelines for assessing portfolios, the Northwest Evaluation Association (1990c) refers to *components* found in a portfolio. CMAP refers to this collection of components as 'contents'.

**Judgment** refers to making statements about how well a program is working. Portfolios are all about judgment. In developing a personal portfolio, students develop the capacity to judge the quality of their work and reflect on growth. In aggregating, the cognitive model also calls for the aggregator's judgment. This is in sharp contrast to assessments using achievement test scores and other "objective", quantitative procedures. Wolf (1975) points out a basic problem with evaluations based on test scores.

> Great collections of numbers, such as those found in children's cumulative files and school or program evaluation studies, tend to blur and obscure rather than sharpen and illuminate the education process. In seeking objectivity, the decision maker may exclude a factor that ought to be of fundamental concern: human judgment. (p. 185)

Eisner (1990) also sees no benefit in letting test scores replace judgment but points out some of the reasons that evaluators find doing so attractive.

> ...the creation of procedures that eliminate judgment are certainly possible. Hermetically sealed saran-wrapped achievement tests whose questions are to be answered by filling in blanks with graphite so that they can be scored by machine untouched by human hands, provide ample testimony to the attractiveness of such procedures. Such tests are not only politically safer than exercising judgment, (Exercising judgment on high stake tests can be dangerous.) they are also very efficient. Yet consensus achieved through procedural objectivity provides no purchase on reality. It merely demonstrates that people can agree: we hope that for good reasons, but what constitutes good reasons...is itself a matter of consensus. That might be all we ever have, but we ought to recognize it for what is. (p. 8f)

The activity dimension of the cognitive model of portfolio assessment is illustrated in Figure 2. It indicates that in assessing portfolios, the student and aggregator divide their attention across five aspects of a portfolio: rationale (why the portfolio is being assessed), intents (what learning, in general terms, the portfolio will demonstrate), contents (the materials that find their way into the portfolio), standards (how to describe good and not so good performance) and judgments (using the standards, what the contents tell with respect to the rationale and intents).

In Figure 2 the contents column is specially shaded as a reminder that contents refers to the tangible materials that make up the portfolio, and that the selection of the contents is tied, explicitly or implicitly, to rationale, intents, standards and judgments.

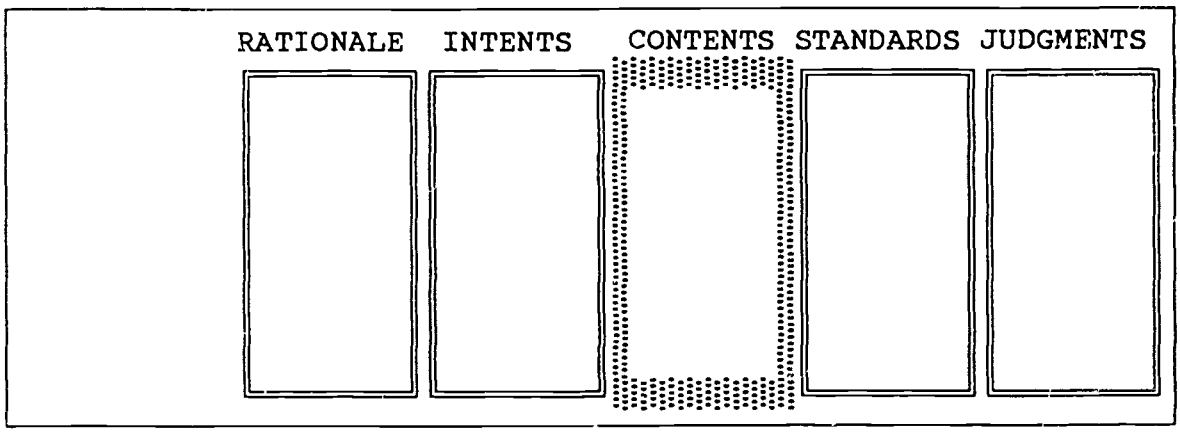


Figure 2. The *activity* dimension of the cognitive model for assessing portfolios.

### The *historical* dimension

The historical dimension puts the portfolio into temporal perspective. This is the dimension that looks at last year's portfolio and this year's portfolio from the perspective of next year's class.

Antecedents refer to baseline performance, the context, and learner characteristics. The antecedents define the starting points and set the stage for the judgments that will be made about student learning and program effectiveness.

Transactions are the countless encounters that occur around the portfolio itself. The portfolio becomes an organized depository for materials that reflect a large number of transactions. Transactions are any instruction, experience or exposure that brings about change. Transactions include encounters between the learner and the thing being learned, and between the learner and a different stakeholder's set of rationales, intents, contents, standards, or judgments. "Transactions" equals knowledge-as-a-verb, the "knowing", the "doing", the "constructing", the "understanding". In the cognitive model, transactions account for the differences between antecedents and outcomes and are of great interest to an aggregator.

Outcomes, traditionally, have been the major interest of most educational assessment. In Glaser's 1963 quote (*op.cit.*), they are the everything. The proposed model places outcomes in balance with the rest of the program. The aggregator looks at outcomes, not just for their own merit, but also to compare them with antecedents and to identify or infer the transactions that took place in between.

Figure 3 relates the historical and activity dimensions. In Figure 3, the student work being "celebrated" in the portfolio is represented by the intersection of contents and outcomes. Earlier work that is exhibited to demonstrate progress is represented by the intersection of contents and antecedents. Early drafts of a work, conference notes, or students' reflections on their own learning are represented by the intersection of contents and transactions. Figure 3 also shows that historical changes in contents are irrevocably tied to historical changes in rationale, intents, standards and judgments.

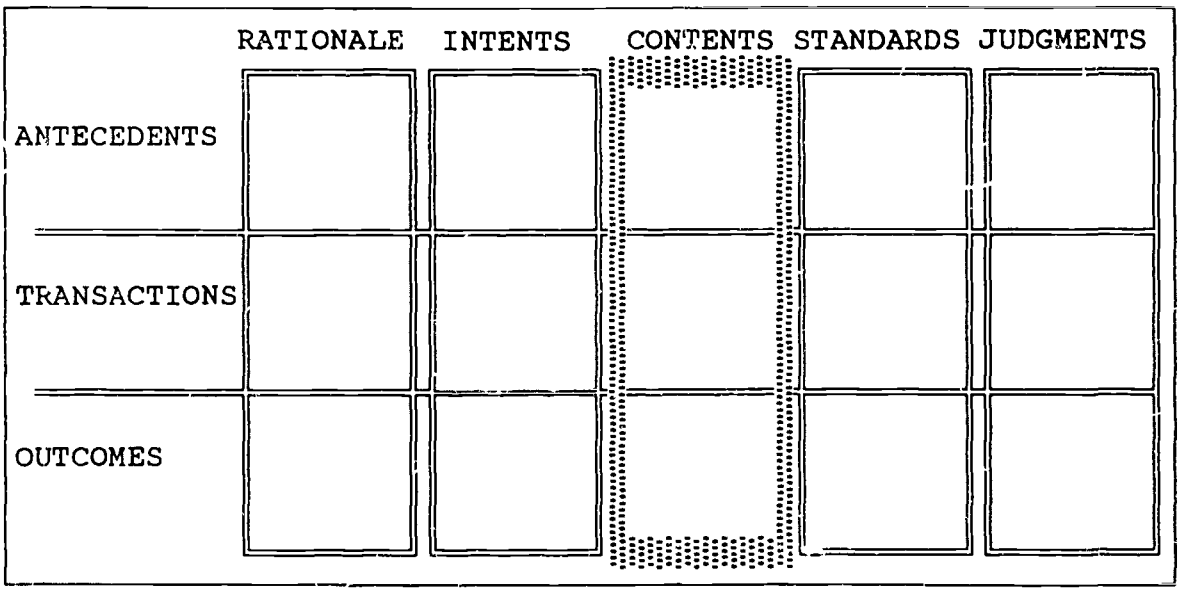


Figure 3. The *activity* and *historical* plane of the cognitive model for assessing portfolios.

An aggregator following a traditional, unidimensional assessment model would take inventory of the outcomes exhibited in a portfolio's contents, or measure the changes between the antecedents and outcomes. The cognitive model of portfolio assessment does indeed allow such aggregation. However, aggregators using the multidimensional cognitive model are also able to infer metacognition any time there is evidence that a student has related outcome contents to *any other cell* in the model. Students show evidence of metacognition any time they contrast recent work with earlier work (outcomes with antecedents); include early versions of finished products or reflect on their learning strategies (transactions); or relate their outcomes to some rationale, intents, or standards. The cognitive model allows aggregators to use portfolios as a unique way to assess metacognition.

The *stakeholder* dimension

Finally, we have the third dimension of the cognitive model. The stakeholder dimension identifies those individuals with an interest in the portfolio. The concept of the stakeholder in evaluation has been fully developed by Guba and Lincoln (1989) in a concept they call *fourth generation evaluation*. In fourth generation evaluation, different stakeholders are provided enough information that they can reach consensus or at least negotiate a compromise. CMAP facilitates such consensus and compromise by having all stakeholders announce their rationale, intents, standards, and judgment processes at the outset and by making timely reports of their transactions and outcomes[2].

In the development of portfolios, stakeholders may include the student, the teacher, the family, the district and the society in general. The aggregator is a stakeholder who may represent an additional perspective or a perspective already represented. We limit the present discussion to only the three: the student, the teacher and the aggregator.

Student. A portfolio developed by a student will probably integrate the rationale, intents, standards, and judgments of the teacher. However, the student as a stakeholder will have personal reasons for developing a portfolio. Students have personal goals and objectives and set standards that pertain to them. Variation will exist across students. For example, it would be expected that a student preparing to enter college to study science would have rationales, intents, contents, standards, and judgments different from a student preparing to enter the personal computer sales force or one planning a career in law enforcement.

Teacher. The portfolio project planned by the teacher would probably integrate the rationale, intents, standards, and judgment processes established by the district in its curriculum, modified by to reflect each teacher's personal philosophy and the perceptions of this year's class.

Aggregator. The aggregator brings an outside viewpoint to the portfolio. The rationales, intents, contents, standards, and judgments will be similar to those of the agency that contracted the aggregation and also be determined by the intended use of the aggregator's report.

Portfolio assessment starts at the student level with metacognition, the student making judgments about the contents that form the portfolio. Portfolio assessment continues at the teacher level with teachers reviewing and reflecting upon the portfolio contents and making judgments. The process is consistent at the two levels and mutually supportive. Each successive stakeholder enters the picture to review, reflect on, and make judgments. Portfolio assessment thus becomes a consistent, supportive process at all levels of aggregation. A worthy object of review is the degree of compatibility or

---

[2] The NWEA (1990c) guidelines for portfolio assessment introduce the concept of a *composite portfolio*. A teacher could put together a portfolio that represents an individual class. This composite would contain all elements of an individual student portfolio including teacher's self-reflection. A composite portfolio might be assembled at any stakeholder level

incompatibility between stakeholders with respect to rationales, intents, and standards. This aggregation plane is represented by Figure 4.

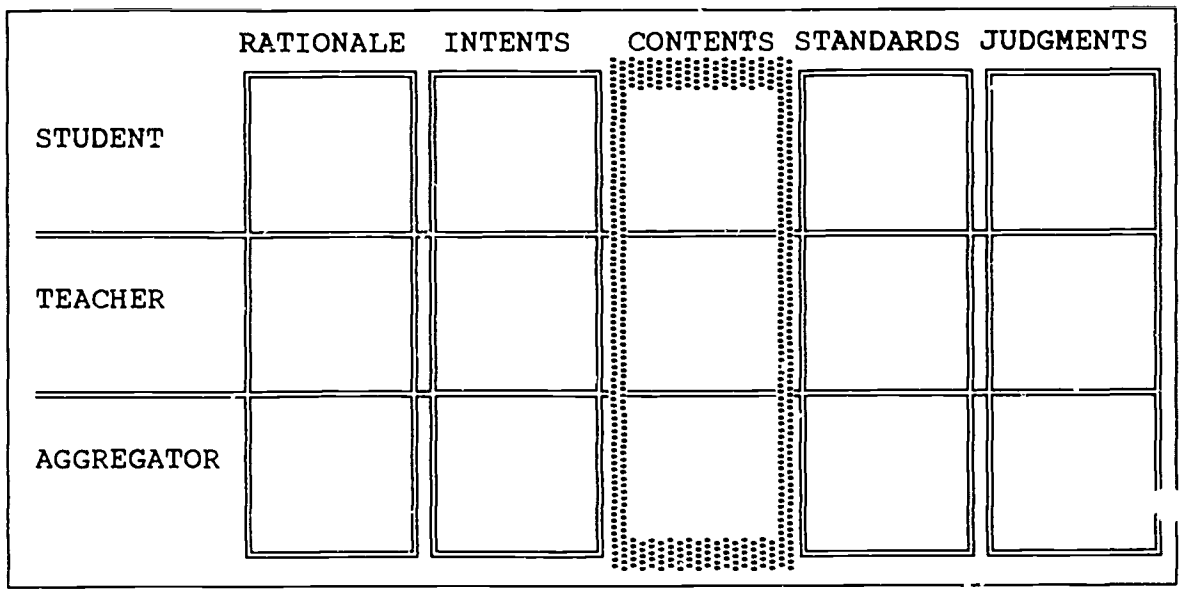


Figure 4. The *activity* and *stakeholder* plane of the cognitive model for assessing portfolios.

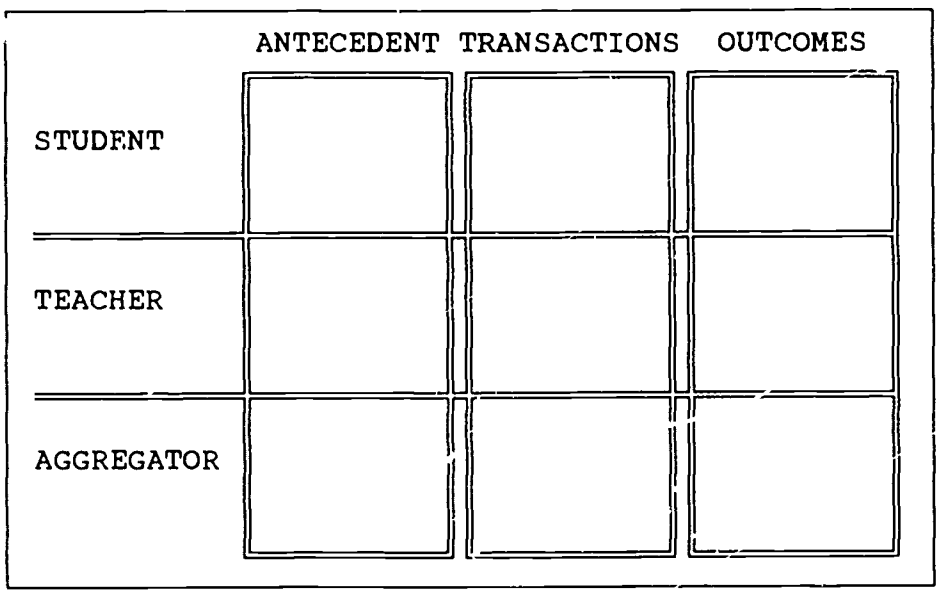


Figure 5. The *historical* and *stakeholder* plane of the cognitive model of portfolio assessment.

The second area of interest is the impact of stakeholders on each other. See Figure 5. Will the students', teachers', and aggregators' knowledge of each other's rationale, intents, standards and judgments have a positive or negative impact? Will any resulting consensus or compromise be educationally desirable or undesirable? These questions are explored by using the plane depicted in

Figure 5 as a way to slice through the activity dimension, examining how stakeholders influence each other's rationales, intents, contents, standards, and judgments.

## The cognitive model: Implications for aggregation

The cognitive model for assessing portfolios makes no direct requirement that certain content be in place. If a teacher intends to teach *X*, and the standards permit judgment of whether or not student has demonstrated *X*, we can permit wide latitude on how the student (or teacher and student together) choose to demonstrate *X*. This may pose problems for aggregators who proceed with traditional expectations that student behavior will occur under controlled and carefully standardized circumstances. In this section we discuss how aggregators can proceed without detracting from the instructional purposes of the portfolio, and even exploit them as evidence of students' capacity to reflect on their own learning.

As they develop portfolios, students evaluate their own work; the aggregators evaluate the work of many students. Such dual evaluation of portfolios is simultaneously a strength and a vulnerability. The selective inclusion and exclusion of materials in a specific portfolio reflect a particular student's rationale, intents, and standards; assessment across portfolios is based on the aggregator's rationale, intents, and standards. The two sets may be different, the cognitive model tolerates such differences, thereby safeguarding portfolios as a context where students develop independence in the processes of self-assessment.

In the event that an evaluation plan requires that the students and aggregators activities be compatible, or even identical, it is the aggregators' activities, not the students', that must be aligned. As a result, the aggregators may not have a choice in whether best work or process is described, whether single or integrated content is examined, or whether judgments are based on relative or absolute standards. The students must not be asked to engage in activities that prevent them from pursuing their own rationale, intents, and standards. A major function of portfolios is to encourage students to become independent in all the activities of self assessment. The students' independence must not be compromised.

There are several ways that aggregation can proceed, even when students are given free rein in determining rationale, intents, and standards. An obvious way is for aggregators to explain their interests *before* students begin their portfolios, asking students to address both their personal and the aggregators interests. Ideally, knowledge of the aggregators' interests will be instructive, providing students an alternative assessment model that makes them aware of the values held by other stakeholders. for example, say the aggregator is looking for the correct use of punctuation. If the student has written a piece using a *stream of consciousness* technique, knowledge of the aggregator's interest in assessing use of *conventional* punctuation will allow the student to call attention to the use of nonconventional punctuation and request that it not be judged using conventional standards. It is important that the aggregators' activities be explained in such a way that they do not devalue the students' personal choice of rationale, intents, and standards. The aggregator might, for example, consider adding categories that record both the accurate use of conventional punctuation and the appropriate use of nonconventional punctuation. Thus, the collected work placed in the portfolios provides a context that may lead to a richer understanding of, or even different conclusion about a student's abilities made with more traditional methods. Conceivably, this approach could be used to provide a deeper appreciation of the performance of students from linguistic or ethnic minority groups.

Portfolios are also repositories of metacognitive data that are unavailable in more traditional sources. Aggregators have the opportunity to report the frequencies with which students' metacognitive reflections fall within the various cells of the historical by activity dimensions (Figure 3). For example, evidence of students' metacognitive activities may include students' description of the experiences that they believe account for differences between early and recent work, or descriptions of their activities as they proceeded to compile their portfolios and evaluate the contents. Students may report how personal standards have changed, how what they once considered a satisfactory level of performance is no longer. Aggregators also have an opportunity to observe the degree to which students satisfy their own goals and the consistency and sophistication with which students make judgments about their work.

Additionally, the portfolio can provide information useful for program evaluation. The program evaluator can report the degree to which there is evidence that student portfolios are consistent with program rationale (philosophy), intents (goals), content (curriculum), and standards. In other words, the aggregator can make statements regarding the match between what the students find valuable and the goals that educational policy makers believe should be valuable.

In summary, the cognitive model is like the traditional models of assessment in that both set specific criteria on what constitutes a portfolio. The traditional model focuses on structure: what specific components must be present to have a proper, scorable portfolio. The components must be standardized. The cognitive model is constructivist: it focuses on the processes that must be present in order for the student to produce a proper, describable portfolio. In assessing, the cognitive model focuses more on process than product. The cognitive model allows students options in how they address their own or the aggregators' rationales, intents, and standards. It is our personal belief that requests for specific exhibits (*e.g.*, outputs, checklists, assignments) can and should be avoided. By doing this, aggregators may be unable to report what students do not know, but have a unique opportunity to discover what they do know and how they choose to put this learning to use.

## Assessment using judgment

Thus far, we have argued that portfolio assessment should focus on multidimensional, cognitive processes rather than unidimensional achievement of facts and skills. We described an assessment model that is descriptive and encourages the collection of complex and varied information. We took the stance that the ultimate outcome, the bottom-line of portfolio assessment, should be determined by human judgment. Finally, we discussed how the model could be used to aggregate information.

There is a fundamental question that we must now confront in proposing assessment techniques that don't just allow diversity, but encourage it. That question is this: Does our view of assessment leave us with a solipsistic world, one without rules in which "anything goes", where "what is the answer?" is replaced by "what do you want the answer to be?" What happens to "objectivity" in the cognitive model of portfolio assessment? D.C. Phillips (1989) provides an answer that works for us. He writes

> It turns out, then, that what is crucial for the objectivity of any inquiry -- whether qualitative or quantitative -- is the critical spirit in which it has been carried out." (p. 35)

Thus, we call for portfolio assessment that respects the human capacity for judgment.

Instead of using test scores *per se* as the outcome indicators, our model uses judgments that have criteria and are open to review. P. Paulson (1972) had judges refer to a taxonomy of possible content

found in student paragraph summaries in order to reliably identify commonalities even though the paragraph summaries had no words in common. F. L. Paulson (1974) used a judgment system to assess the impact of instructional television on the cooperative behavior of preschool children.

Several portfolio projects base outcomes decisions on judgments. The Fort Worth pilot project in writing portfolios (Lewis, 1990) uses structured teacher judgment to determine change both within and between grades. The procedure focuses on variables to which the portfolio is particularly sensitive (*e.g.*, increases in length or complexity of pieces, self-initiated revisions, risk taking, metacognition). Medford (Oregon) High School employs panels of faculty and community members to rate the performance of students participating in the *Senior Project* (Weiss, 1989; *The Senior Project*, undated). Vermont's statewide portfolio program (see Writing Assessment Leadership Committee, undated) asks citizens to participate in portfolio assessment in an effort to encourage discussion and provoke thought about how well the schools are achieving their goals. According to the Commissioner of Education (Mills, 1989), "We are undertaking this massive project because we are interested in *real* student's work, *real* performance, not the proxy delivered by standardized tests." (p. 11) The Stanford University Teacher Assessment Project (a project with a strong portfolio component) is developing judgment based criteria for assessing teacher competence (Hertel, 1990; Schulman, 1987), and Connecticut performance assessment program is developing classroom observation instruments (Connecticut Competency Instrument) that are flexible and context sensitive (Suen & Davey, 1990; Hertel, 1990).

## Methodology: judgment, description, and (yes!) numbers, too

Whatever the approach to making judgments, several research and evaluation methodologies are available that can make the process more rigorous. This section looks at several that are particularly promising. One comes from psychometrics, the study of ways subjective human judgment may be quantified. Another is an extension of classical test theory that recognizes and estimates multiple sources of measurement error allowing for a more multidimensional approach to measurement. Others from ethnographic research provide clues on how to aggregate from case studies and descriptive data. Additionally, we speculate about the implications of chaos theory for educational evaluation.

Psychometrics began with the pioneering work of Francis Galton and was continued by Thorndike (1948) and Stevens (1958). Under pressure of congressional legislation to protect environmental and ecological resources, the U. S. Forest Service asked psychologist Terry C. Daniel (1990) to develop ways to use human judgment to to measure the quality of the natural environment. Daniel's charge was to expand the narrow, bottom line, unidimensional thinking that represented environmental values in economic terms (*e.g.*, representing the value of trees as board feet) to a broader basis that would accommodate ways of representing *natural scenic beauty, wilderness experience, outdoor recreation*, and *visibility values*. In a sense, his problem was to apply a new way of aggregating the value of the Department of the Interior's portfolio.

Daniel had to deal with two constraints that often confront portfolio assessors. First, existing policies and procedures placed a premium on quantified, objective traits. Daniel had to translate the subjective assessment of beauty into a standardized, "objective", quantifiable framework acceptable to his audience.

Psychometric techniques used by Daniel use multiple judges (sometimes representing different points of view) to make independent ratings of stimuli (in this case landscape examples). The procedure uses sums of judgments by all judges across landscape examples and the sum of landscape examples across all judges to develop an overall index of scenic beauty. What goes in are highly "subjective" human judgments of the beauty of different settings, what comes out are surprisingly stable estimates of environmental quality across different reference groups and environmental settings. Daniel's work is particularly appealing for portfolio assessment where students are encouraged to attend to self reflection, feelings, aesthetics, social significance, and other dimensions that are poorly measured using standardized techniques.

A second technique often used in judgment-based studies (see, for example, Paulson, 1972) comes from the use of the analysis of variance, a statistical technique that looks at test reliability from a multidimensional viewpoint (Cronbach, *et.al.*, 1972; Feldt & McKee, 1958; Shavelson, Webb, & Rowley, 1990). Generalizability theory uses the analysis of variance (ANOVA) to attribute error of measurement to multiple sources of variation. In their paper on performance assessment, Shavelson, Carey & Webb (1990) contend that

> [b]ecause the complexities of such [performance assessment] measures will not match the assumptions of current psychometric theories (*e.g.*, item response theory), creativity is...needed to exploit existing psychometric theory (*e.g.*, generalizability theory) and to develop new theories to make the new testing technologies efficient and cost-effective. (p. 697)

Generalizability theory is uniquely qualified to address issues identified by Suen and Davey (1990) in their analyses of Connecticut's performance assessment procedures. One issue is whether different standards for reliability of measurement apply for performance assessment and traditional achievement tests. The more situations are allowed to vary (*i.e.*, remain *non*standardized) *the more the traditional measures of reliability and validity drop.* In ANOVA terms, the more fixed and the fewer random facets (*i.e.*, the more standardized the test), the "better" the reliability and validity. Traditional paper and pencil measures maintain high reliability by fixing conditions, but pay the price in reduced authenticity. The contrast between paper and pencil tests and authentic tests (including portfolio assessment) can be viewed as a continuum of high to low standardization. Generalizability theory is a technique that can yield a more refined analysis throughout this continuum.

Psychometrics and generalizability theory both grow out of the tradition of quantitative research in psychology. Enthnographic research also provides approaches to aggregation used in anthropology and sociology (Schofield, 1990). Ragin (1987) is developing a *qualitative comparison method* that uses boolean algebra to make sense out of complex, multiple patterns found in case studies. His system lends itself to social research because it allows the same cause may have different effects and the same effect may have different causes. Noblit and Hare (1988) while rejecting the concept of aggregation because it loses more than it keeps, offer a technique called *meta-ethnology* designe ' to preserve the interpretive nature of qualitative information across studies. Their technique is designed to develop a synthesis, a translation into a common language permitting comparisons across studies.

Our final example of analytic techniques with potential application to portfolios (at least in the long term) comes from the emerging science of chaos (Gleick, 1987). Chaos theory is the study of the growth of complexity in nature. It is the study of patterns in natural events that had formerly been thought to be random (test theories assume variation away from the unidimensional model is random error). Chaos theory uses the mathematics of nonlinearity in which simple relationships are replaced by complex patterns that accurately describe the behavior of many naturally occurring events. It has already found practical application, for example in medical treatment (see Gleick, 1987 p. 280ff). Techniques with colorful names, such as *bootstrap* and *jackknife*, are already finding their way into the behavioral sciences (Cronbach, 1988, p. 49, Lunneborg, 1988).

The methods reviewed come from both quantitative and qualitative research. Some, like psychometrics, and generalizability have been available to educational researchers for some time. Others like those used in ethnographic and qualitative research are beginning to find acceptance in educational research and evaluation. The *American Educational Research Journal*, for example, has only recently accepted qualitative research manuscripts The more exotic techniques like chaos theory may remain in the wings for now but have a leading role in the next generation of educational research.

Obviously there are a great many techniques that might be used in the cognitive model for assessing portfolios. We urge adoption of the pragmatic position recommended by both Patton (1990) and Stake (1963). Portfolio assessors and aggregators should adopt and adapt freely from all schools of research methodology but carefully remain free from the confines of allegiance to any single approach. In portfolio assessment, diversity in content should beget diversity of analysis.

## Practical applications of the model

To this point, we have explored theoretical issues with profound practical implications for portfolio assessment, and we have outlined a conceptual model for assessing portfolios and aggregating the results. How might such a model look in practice?

Any or all subject areas may be represented. Within subject areas, students and teachers generally have wide latitude regarding material that finds its way into the portfolios. There may be more than one set of rationales, intents, standards, or judgments. Multiple sets of these may represent the aggregator, teachers, students and other stakeholders. The teachers and students may compare their sets to those of the aggregator, but not necessarily with any intention to make them the same.

Teachers as well as students may place material into the portfolios if ground rules are clear at the outset. In principle, students should have complete freedom over what they may put in the portfolio, but they are encouraged to give their reasons for the selection. Teachers provide younger children more models of the decision making process.

The aggregator's initial attention is to the extent to which the district has used a model that preserves the function of portfolios for teaching self-assessment. Aggregators determine the degree to which students and teachers have related contents to district intents and evidence that students and teachers have modified or enhanced goals to reflect individual needs or interests. The aggregator identifies which district goals and priorities are addressed in the portfolios and seeks confirmation that individual students are free to demonstrate competence in whatever ways make the most sense personally. In addition, the aggregator ascertains that the district encourages students to contribute their own personal goals. Finally, the aggregator judges the degree to which students are successful in meeting their own standards.

At first glance, the "contents" of the portfolios are simultaneously an evaluator's nightmare and an anthropologist's bonanza. However, most of the material is carefully indexed and, the indexing provides structure. The indexing may be to district goals, or according to the student's, teacher's, or aggregator's intents, or cross-indexed to several. Indexing to these goals and intents provides the aggregator a clearer pathway through the portfolio contents.

All portfolios also contain specific guidance on how performance is to be judged. For example, a writing sample is accompanied by a handbook on the district's analytical trait writing assessment model. In mathematics, there is a list of criteria applied to problem-solving exercises (For examples, see Beaverton, 1986; California Mathematics Council, 1989, p. 19)

Contents are not standardized. The following examples give a flavor of the diversity of contents that could be indexed to a district mathematics objective; all student portfolios contain a form instructing them on how to present and document evidence that they can perform long division but students have used innovative ways to document competence at long division. One student's portfolio contains a copy of a standardized achievement test report with the number of items answered correctly in long division highlighted by a yellow marker and the district goal number penciled in. This student has written a note that her dad is only interested in the bottom line! Another student's portfolio contains materials developed as a part of a "Business in the Schools Project" sponsored by a local business group. Although there are dozens of cross references to district objectives, the cross reference to long division is on the project's inventory reports; the student attached several examples of long division necessary for completion of these reports. A third student includes a two page written problem analysis of the solution to an arithmetic challenge problem (see California Mathematics Council, 1989, p. 12). An appended note points out how the solution required the accurate calculation of long division. At the end of the year, the portfolio review committee, applying standards stated at the outset, will look at these various kinds of evidence to determine whether students satisfied the goal.

Portfolio contents are unique in their attention to metacognitive processes. Students and teachers don't just put material into the portfolio, they reflect upon the material, what they learned from the exercises, and why they are placing material into the portfolio.

There is an important strategy implicit in this description. The aggregation process is structured in such a way that external impact on the portfolio contents is limited to the statement of instructional goals and standards. The contents are free to vary widely but with the challenge to the student and teacher to relate them to the goals and to discuss the outcomes in terms of the standards. This leaves the program open to multidimensional thinking. One can approach the goals in a variety of ways using a variety of media. There is minimal impact of the assessment on the thing assessed. Diversity is encouraged and the assessment is performed by judges using criteria that adapt to variability.

## Example portfolio projects

The examples to follow illustrate how the cognitive model for assessing portfolios (CMAP) may be used to describe three existing but very different portfolio projects. One project uses portfolios for deciding which students should be admitted into programs for the gifted. We illustrate how CMAP could provide a structure for project evaluation. The second describes a program in which a classroom teacher uses the portfolio to promote student growth in writing. CMAP illustrates a structure for evaluating the instruction students receive in self assessment as well as the quality of their writing. In the third, we show how CMAP may be used to describe an example of metacognitive activity. CMAP provides a structure for evaluating student's self assessment as reported and documented by those students.

### Example 1. The *Pupil Product Portfolio*

The *Pupil Product Portfolio* (PPP) is a system for helping districts identify gifted children (Grades K - 3) and for helping teachers make instructional decisions. It was developed by Bertie Kingore (1990a, 1990b & 1990c), a specialist in gifted children at Hardin Simmons University. The PPP has been tested under her supervision with several hundred children, and reports are in preparation (Kingore, 1990c).

The PPP is a collection of physical evidence of excellence in performance. It is assembled and kept by the teacher. The PPP may contain three categories of materials:

1. **Spontaneous products from school and home** that the child produces spontaneously and not as a part of class assignment. Kingore describes them as things that "just happen."

2. **Periodic Products** collected from classroom work (e.g. writing journals) in order to document growth. (This category is used for instructional purposes only and not used for identifying the gifted.)

3. **Planned experiences** from standardized situations collected on all children at the same time in order to give all children an equal chance to qualify for gifted programs. An example is *Drawing Starts*, ambiguous, partially drawn figures that children complete. The scoring focuses on process, with no right or wrong answers.

To identify gifted primary students, a judgment team reviews six examples of planned experiences and three spontaneous products found in the PPP. Giftedness may be demonstrated in five categories:

1. In-depth understanding

2. Unique or unexpected idea

3. Exceeds grade or age level expectation

4. Advanced, complex organization

5. Resourceful and clever use of materials

Each judgment decision is on a binary scale: the child either does or does not demonstrate performance consistent with one or more of the categories. The total score on PPP is the sum of points awarded for the 6 planned experiences and 3 spontaneous products. Children who receive either 8 or 9 points are classified as gifted. Kingore (1990c) reports relatively high (90% plus) levels of agreement between judges.

An analysis of PPP using CMAP reveals almost immediately that PPP does not focus on the student as a self-assessor. However, the model can be directly applied as a structure for comprehensive *program* evaluation. And while the PPP does not focus on self assessment, CMAP enables the aggregator to discover student metacognition as "serendipitous breakthroughs" within the contents of the portfolios.

Applying CMAP to the PPP project, the active stakeholders clearly are the district and the teacher. The student is a passive stakeholder. Although the student is the person who supplies the contents and

is directly affected by the decisions, the student is not asked to actively participate in any of the other assessment activities: formation of a rationale, intents, standards or judgments.

The activity dimensions are readily identified for the district stakeholder and the teacher stakeholder. The district's rationale is to identify gifted students. The intents are for students to demonstrate giftedness as defined by the five categories. The portfolio contents are to be six planned and three spontaneous student products, selected by the teacher. The standards are binary ones applied to each each of the nine products. The judgment is giftedness, based on a score of 8 or 9 points. An aggregator applying the cognitive model to this dimension would conclude that all the activities of assessment were in place and that there was internal consistency along the activity dimension.

The teacher's rationale has two components, one being to meet a district requirement for selecting gifted students. The teacher would therefore have adopted the district's intents, standards, and judgments as the basis for selecting student work. Program evaluation would look for consistency in the way the teacher went about the district's business. The teacher's other rationale would be to guide instruction. The PPP project does not specify the intents, standards or judgments in relation to instruction. The aggregator would have to determine these through teacher interview and look for evidence of these in the contents of the student portfolios.

Of the three dimensions of the cognitive model, the historical one is least relevant to the PPP project because, by definition, the Pupil *Product* Portfolio is specifically concerned with outcomes, not transactions or antecedents. Transactions that account for these outcomes are not a focus of the district as it goes about selecting gifted students. It is assumed that transactions are of great concern to the teacher who helps students develop planned and periodic products and who uses these (and spontaneous products) to guide future instruction. While perusing the portfolios, the aggregator might find evidence that the students were aware of their transactions (just as there might be evidence that they engaged in activities of self-assessment) and such a discovery would enrich the aggregator's report. The final component of the historical dimension, antecedents, does not appear for any stakeholder or activity.

### Example 2. A Writing Portfolio

A description of a writing portfolio activity in a fourth grade classroom appeared in the NWEA *Portfolio Assessment Newsletter* (NWEA, 1990b). It is an activity conducted by Rhonda Woodruff, a fourth grade teacher in Beaverton, Oregon. The project can be described using the three dimensions of the CMAP.

#### *Stakeholders.*

The student's writing portfolio is used by three stakeholders, students, teachers, and parents. Students take an active role in selecting material and maintaining the portfolio and exhibit a high level of ownership in their portfolios. The teacher uses the portfolio to provide feedback to students, monitor the class, report to parents, and provide information for next year's teacher. The portfolio is available to parents and is a favorite source of information on progress in writing.

*Historical.*

The historical dimension is represented primarily as transactions and outcomes. The transactional dimension is represented in a least three ways. A record of published writing is kept over the entire year. Students are encouraged to include any writing they feel *proud* of. The process of writing is a clear focus of daily instruction and learning. This would support a look at growth. Since the contents are described as finished writing, the primary focus would appear to be on outcomes.

*Activity.*

The teacher described and modeled her rationale, intents, contents, standards, and judgments. She kept a portfolio herself and taught the students how to engage in the same process. Excerpts from the article are accompanied by CMAP analyses to show just how the model can be used as a framework for describing existing program.

*Analysis*

In the analysis, excerpts describing the program appear on the left, the analysis of each excerpt appears on the right.

---

| Program Description | CMAP Analysis |
|---|---|
| *In my fourth grade classroom, students learn writing as a process. The use of writing stages are taught. The writing stages include: pre-writing, drafting, revising, editing, publishing, and sharing.* | Here the teacher conveys to the students her expectations regarding the **transactions** that will occur as a part of the portfolio activity. |
| *In addition, my young writers are instructed in the use of Analytical Trait Scoring. Analytical scoring defines characteristics of writing. It allows us to measure a writer's ability to deal with individual components of writing.* | The teacher's reference to the Analytical Trait Scoring system indicates the **standards** that will be applied to the portfolio. **Transaction** is implied by the teacher's reference to specific instruction in the system. |
| *The six traits we use are: ideas, organization, voice, word choice, sentence structure, and writing conventions* | The listing of the six traits themselves is an indication of **intents**. |
| *Students learn how to assess writing to these traits, recognizing that a paper may be mechanically sound but weak in ideas, or perhaps strong in organization but rather weak in vocabulary. Students also receive instruction on how to revise papers using these traits. They learn how to make weak papers strong. These six traits are my six writing units for the year.* | This statement is **transactional** in which the teacher conveys to the student specific expectations regarding their performance.<br><br>[Not stated in piece: Each student in this district has a copy of the district's writing handbook that lists the **standards** used to judge the district's writing assessment. Thus, the teacher transmits the district's **standards** to the student.] |

| Description (Concluded) | CMAP Analysis (Concluded) |
|---|---|
| *Writing portfolios are essential in helping provide writing success for my children. Portfolios are meaningful to students, teachers, and parents alike as they serve as a place where we, as writers, can store part of ourselves. Our writing successes. Our inspirations. Our hearts.* | Here is a succinct statement of the teacher's **Rationale.** It is stated in a collaborative mode (the reference to students, teachers, and parents) and implies attention to metacognitive processes. |
| *Every writer in my class...has a writing portfolio* [with] ... *(1) a list of books read both in and out of class, (2) a list of writings published or shared during the year, and (3) xerox copies of writing treasures we, as writers, wish to save (the originals are in much demand at home).* | Teacher establishes the general **contents** of the portfolio. Students have wide latitude in terms of the **contents** they select to include. |
| *Writing in the portfolios may come from journals, class assignments, writing composed at home, on trips. Anywhere!* | |
| *At the end of the school year, students choose from their treasures three samples of favorite writings to leave in their portfolios to be passed on to their fifth grade teachers. Students add to the portfolios self-assessments of their progress in writing as well as fall and spring writing assessments completed in class.* | This describes the student as stakeholder. It deals with the **antecedents, transactions,** and **outcomes.** |
| *My wish is to not only send a portfolio on to the next year with every student but to receive one for each new student coming to my room in the fall. I would love to be able to tell every new student the very first day of school, "I may not yet know you by your face, but I know you already by your words Your thoughts Your interests. Your successes. Welcome."* | This year's **outcomes** become next year's antecedents. |

Example 3. The Metacognitive Letter

Jill Marienberg (see NWEA 1990a), a Hillsboro, Oregon, high school teacher, provides examples of letters that students submit that describe their writing portfolios (Marienberg, 1990). These letters, which Marienberg calls *metacognitive letters*, demonstrate that the students have actively engaged in all five activities of self assessment and are aware of many of the transactions critical to their learning.

In the example metacognitive letter that follows, the student states rationale and intents in the first paragraph, lists the contents, then discusses each piece of writing she has chosen to included in the portfolio. In the discussion, the student specifies or implies the standards by which she judged her work.

Apparently, the teacher provided the students with a rationale which might be paraphrased, "The purpose of the portfolio is to allow students to collect samples of their writing during the junior year

that accurately represent their thinking and writing abilities." Students typically provide a version of this statement in their letters, *e.g.*, in the example letter, "...what I consider to be my finest work....an overview of my ability to write and and organize my thoughts."

In examining several metacognitive letters from Marienberg's classes, statements of intents vary from student to student. The example letter lists "expository, descriptive and reflective pieces", as well as evidence of writing and organizing "under pressure." Another student listed intents as demonstrating the ability to produce "expository, informative, and creative pieces".

Contents also vary from student to student. Some students included more poetry than others, a few include research pieces, and most include something written as a part of a major class assignment. Marienberg is clear that which pieces are selected is completely up to the student. The student who wrote the sample metacognitive letter was unique in including an example from the district's direct writing assessment because she wanted to show her ability to write "under pressure". It is interesting that when students were asked to demonstrate how well they can write, the student who choose to include an example taken from the district's direct writing assessment made an explicit disclaimer that it showed the ability to produce under nontypical circumstances.

According to the metacognitive letters, students clearly set standards by which to judge their work. These standards are specific not only to each piece of work but to the interests of each student. Across the class, students' standards vary from holistic ones (e.g., "the 'art' of portraying the usual as unusual") to moderately analytic ones (e.g., "showing voice"), to highly analytic ("improving verb choice"). Statements of standards may include aesthetics considerations (e.g., "startling imagery"), the demonstration of the achievement of traditional goals (e.g. "writing within a certain structure"), or breadth of skill (as in the above example where the student includes an example of writing under pressure to complement the other materials in the portfolio).

There is substantial evidence that students are aware of the transactions that account for their growth during the year. The example letter includes these statements: "I believe that all of my writings have benefited from this assignment early in the year"; "This selection stretched me the most intellectually"; "My knowledge of main characters in *The Scarlet Letter* had to be intimate..."; "My organization and logic skills enable me..."

Finally, the letter ends with an elaboration of the portfolio's rationale: "To assess my growth, competence, and college-level thinking."

A transcript of the metacognitive letter written by student S.D. follows.

---

*To Whom it May Concern*

*Enclosed in my portfolio you will find five samples of what I consider to be my finest written work during my junior year in high school. The portfolio contains writing samples from expository, descriptive, and reflective pieces, to give you an overview of my writing ability. In all but one of these works I was allowed plenty of time for revision and critiques. However one piece which I have included is from a timed in-class assignment. I hope this example will provide you with an idea of my ability to write and organize my thoughts under pressure*

*Contents:*

1. *"House Cleaning"*
2. *Poetry "You and I"*
3. *"My Present"*

4. *Literary Analysis of the Major Characters in The Scarlet Letter.*
5 *Poetry: "Blue"*

*Discussion*

*"House Cleaning" is a short narrative written early in my junior year which focused on the description aspect of writing. I believe that I have captured the "art" of portraying the usual as unusual in this paper. Some of the sentences combined just the right verb choice with the correct description to produce startling imagery. The sentences also pick up a beautiful flowing quality. I believe all my writings have benefited from this assignment early in the year which taught me to be original and leave state-of-being verbs behind*

*The first poetry piece "You and I" is included to show my capability of writing within a certain structure. It uses several figures of speech to depict the same old subject of love in a different way.*

*The third writing which I have included, "My Present" is a piece done in class under a fifty minute time restriction as practice for the Hillsboro District Writing Assessment. Although we were free to write on any subject I feel it is a good example of my ability to work alone under pressure. I would have liked to have had more time to refine and rewrite this piece, however, I feel it has more voice than any of my other works.*

*The fourth piece which I have submitted for you review is "A Literary Analysis of the Major Characters in The Scarlet Letter." This selection stretched me the most intellectually and was the most difficult to write. My knowledge of the main characters in The Scarlet Letter had to be intimate as well as my grasp of Lawrence Kohlberg's theory of moral development. I believe, however, that I rose to the occasion well. My organization and logic skills enables me to put together a smooth-flowing, easy to follow paper, with specific supporting detail throughout. I am pleased with the fluency of the sentences and of the depth of knowledge this paper portrays.*

*The second poetry piece, "Blue" is included because of its startling sensory imagery. Written in free verse I feel it is more me than ony of my other works. This poem contains some intriguing examples of personification and my observation skills are tested. I have given this paper a sense of roundness by bringing "Blue eyes" back in the end.*

*With a wide variety reflected in my five pieces, I hope you will be able to assess my growth, competence, and college-level thinking successfully. I am proud to submit this collection as a reflection of my ability.*

*Respectfully submitted,*

*S.D.*

---

Recap of the three examples

In the *Pupil Product Portfolio*, CMAP clarifies that self-assessment is not a major purpose of this project. In fact, it is not clear whether students are even aware that the assessment is taking place. The historical dimension also plays a minor role. However, the stakeholder dimension is very interesting and an aggregator could employ CMAP to clarify the relationships among the stakeholders and to examine the consistency of their activities.

In the writing portfolio, CMAP reveals that a major purpose of this portfolio project, from the teacher's point of view, is to teach self-assessment. All five activity components of self-assessment are taught, modeled, and practiced in collaboration. The historical dimension is important because the instructional emphasis is on assessment processes. The example lends itself to aggregation in two ways: The aggregator can more easily judge the contents of multiple portfolios at the teacher level because all have identical rationales, intents, and standards. The aggregator can also inspect the

contents for evidence that the students have engaged in self-assessment, and that there have been historical changes in the quality of their self-assessment activities.

In the metacognitive letter, each student provides written self-assessment as part of the portfolio content providing an aggregator with direct evidence of self-assessment. One way to aggregate across portfolios would be to use "evidence of metacognition" as the aggregator's rationale, and to describe the CMAP dimensions represented in the student's self-assessment. By comparing the student's introductory letter to other contents of the portfolio, the aggregator could make judgments regarding the students' ability to self assess and their awareness of growth on the historical (antecedent, outcome) dimension. The aggregator could even look at awareness of the metacognitive transactions that are responsible for this growth.

An aggregator interested in overall results could use one of the judgment-based system considered earlier (page 13ff) to produce information that could be aggregated at the school or district level. For example, in the example of the metacognitive letter, an aggregator might use teams of judges representing various stakeholder groups to score the overall quality of writing in the portfolios. The aggregator could then write a comprehensive report to describe the program from the perspectives of the stakeholders while putting the quantitative results into a context that would make them meaningful.

## Conclusion

There is comfort in test scores. They are units that can be counted and accounted. Test scores fit nicely the American value system of achievement measured by big *bottom line*. You can take them to the bank! The questions this paper raises are not about whether we measure outcomes in quantifiable units -- our "bottom line" oriented culture demands it. Rather, we raise the concern that we examine carefully our currency *before* we start counting. We are a diverse people who think of our society as a great *melting pot* where strength grows out of an amalgam of differences. We are the descendants of pioneers, those *risk takers* who had great ideas and the *initiative* to carry through. And, maybe most American of all, when we are at our best we demonstrate *know-how*, that fabled *Yankee ingenuity* that allows us to carry through. Look again at the multiple choice tests that guide the education of our next generation of leaders. Are they telling us the real bottom line of American education? Are we focusing on developing students with initiative, know-how, and a willingness to take risks, or are we focusing on those who know the "right answer" in its most limited definition?

Portfolios provide a complex and comprehensive view of student performance that encourages us to look at learning as a complex and multidimensional process. They allow us to define achievement in broad, adaptive terms rather than narrow, restrictive ones. Portfolios also provide a complex and comprehensive view in a context where instruction and assessment are inseparable. Because assessing profoundly impacts instruction, we urge that assessments be designed that support the instructional value of the portfolio. We encourage an approach that focuses on the process of assembling and using a portfolio rather than on standardizing what is placed into the portfolio.

The assessment community is at a juncture. We can view portfolio assessment as "business as usual" and impose existing, inadequate, procedures based on traditional models, or we can approach the challenge from new and exciting perspectives. The choice is ours. In advocating the pluralist perspective, we echo Eisner's (1990) assertion that

> ...there is no single, legitimate way to make sense of the world. Different ways of seeing give us different worlds. Different ways of saying allow us to represent different worlds. Helping people participate in a plurality of worlds...is what education ought to try to achieve....We need multiple voices and we need people who can understand them. (p. 11f)

At the start of this paper we noted that the dictionary's definition of *aggregate* read, "a total or whole: a group of distinct things gathered together". The dictionary offers a second definition as well, "the sand and pebbles used in making concrete." Together, these definitions provide a metaphor to summarize our argument. In the attempt to find the aggregate in portfolios, we should seek to uncover the beauty of the structures and the strength of the concrete, not count pebbles and sift sand.

> [Meanwhile...after seven and a half million years, Deep Thought signaled that it had finished computing the answer. The crowd gathered. The tension was unbearable.]
>
> *"do you have..."*
>
> *"an answer for you?" interrupted Deep Thought majestically. "Yes I have. Though I don't think that you are going to like it."*
>
> *"It doesn't matter....We must know it. Now!"*
>
> *"All right....The Answer to the Great Question. Of Life, the Universe, and Everything...is", said Deep Thought, and paused.*
>
> *"Yes..."*
>
> *"Is Forty-two", said Deep Thought with infinite majesty and calm.*
>
> *It was a long time before anyone spoke.... "Forty two...Is that all you've got to show for seven and a half million year's work?"*
>
> *"I checked it quite thoroughly," said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is....Once you know what the question actually is, you'll know what the answer means."*
>
> Douglas Adams, *The hitchhikers guide to the galaxy*

## References

Beaverton School District (1986). *Analytical trait writing: Student Copy.* Beaverton OR: Beaverton School District 48J.

California Mathematics Council (1989). *Assessment alternatives in mathematics,* Berkeley CA: EQUALS, Lawrence Hall of Science.

Cronbach, Lee J. (1957). Two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

Cronbach, Lee J. (1975). Beyond the two disciplines of scientific psychology, *American Psychologist, 30,* 116-127.

Cronbach, Lee J. (1988). Playing with chaos, *Educational Researcher, 17(6)*, 46-49.

Cronbach, Lee J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements. Theory of generalizability of scores and profiles*. New York: Wiley.

Cronbach, Lee J. & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on aptitude-treatment interactions*. New York: Irvington.

Daniel, Terry C. (1990). Measuring the quality of the natural environment: A psychophysical approach. *American Psychologist, 45 (5)*, 633-637.

Daniel, Terry C. & Boster, Ron S. (1976). *Measuring landscape esthetics: The scenic beauty estimation method*. (USDA Forest Service Research Paper RM-167). Ft Collins CO: Rocky Mountain Forest and Range Experimentation Station. (Out of print)

Donmoyer, Robert (1990). Generalizability and the single-case study. In E. Eisner & A. Peshkin (Eds) *Quantitative inquiry in education: The continuing debate*. New York: Teachers College Press.

Eisner, Eliott (1990). Objectivity in educational research. Paper read at the annual meeting of the American Educational Research Association, Boston MA. (The paper is from *The enlightened eye: On doing qualitative inquiry*. New York: Macmillan, in press)

Feldt, L. S. & McKee, M. E. (1958). Estimation of the reliability of skills tests, *The Research Quarterly, 29 (3)*, 279 - 293.

Gardner, Howard (1983). *Frames of mind*. New York: Basic Books.

Gardner, Howard & Hatch, Thomas (1989). Multiple intelligences go to school. *Educational Researcher, 18 (8)*,4-10.

Glaser, Robert (1990). The reemergence of learning theory within instructional research. *American Psychologist, 45(1)*, 29 - 39.

Glaser, Robert (1971). Instructional technology and the measurement of outcomes: Some questions. In W. J. Popham (Ed), *Criterion-referenced measurement: An introduction*. Englewood Cliffs, NJ: 1971. (Originally published in *American Psychologist*,1963, *18*, 519-521.)

Gleick, James (1987). *Chaos: Making a new science*. New York: Viking.

Guba, E. G. & Lincoln, E. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Hardison, O. B., Jr. (1989). *Disappearing through the skylight: Culture and technology in the twentieth century*. New York: Viking.

Hertel, Edward H. (1990). From expert opinions to reliable scores: psychometrics for judgment-based teacher assessment, Draft paper presented at the meeting of the American Educational Research Association, April 1990, Boston.

Joint AERA, APA, NCME Committee for the Review of the Standards for Educational and Psychological (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Kingore, Bertie (1990a). Appropriate programs for gifted young children, *Tempo, (Spring)*, 8 - 9.

Kingore, Bertie (1990b), *The Kingore observation inventory (KOI).* Des Moines, Iowa: Leadership Publishers (P.O.Box 8385, Des Moines IA 50301)

Kingore, Bertie (1990c). Audio tape of teacher workshop *KOI Training* presented at Multnomah ESD Feb 8-9, 1990. Kingore discusses the Pupil Product Portfolio in a forthcoming book.

Lewis, Linda (1990). *Pilot project for portfolio assessment.* Fort Worth Independent School District: Fort Worth, TX.

Lunneborg, Clifford E. (1988). Bootstrap applications for the behavioral sciences, *Applied Psychological Measurement, 12(1)*, 105-106. (Software available in MS-DOS format)

Marienberg, Jill (1990). Unpublished examples of student portfolios. Hillsboro High School, 3285 S.E. Rood Bridge Road, Hillsboro OR 97123.

McCordick, Pamela (1972). *Machines who think.* San Francisco: Freeman.

Mills, Richard P. (1989). Portfolios capture rich array of student performance, *The School Administrator, December*, 8 - 11.

Minski, Marvin (1986). *The society of mind.* New York: Simon and Shuster.

Noblit, G. W. & Hare R. D. (1988). *Meta ethnography: Synthesizing qualitative studies.* Sage University Series Qualitative Research Methods, Vol 11. Beverly Hills: Sage.

Northwest Evaluation Association (1990a). Profile: Jill Marienberg: Language arts in Hillsboro, *NWEA Portfolio Assessment Newsletter, Vol. 1 (1).* 5 Centerpointe Dr. Suite 100, Lake Oswego OR, 97035.

Northwest Evaluation Association (1990b). Profile: Rhonda Woodruff: Fourth grade teacher in Beaverton, *NWEA Portfolio Assessment Newsletter, Vol. 1 (2).* 5 Centerpointe Dr. Suite 100, Lake Oswego OR, 97035.

Northwest Evaluation Association (1990c). *NWEA white paper on aggregating portfolio data: Revised.* (Prepared by Carol Meyer & Steven Schuman). 5 Centerpointe Dr. Suite 100, Lake Oswego OR, 97035.

Papert, Seymour. (1980). *Mindstorms: Children, computers, and powerful ideas.* New York: Basic Books, 190 - 207.

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* Newbury Park (CA): Sage.

Paulson, F. L. (1972). Live versus televised observation of social behavior in preschool children, *Proceedingsk 80th Annual Convention, APA*, 135 - 136.

Paulson, F. L. (1974). Teaching cooperation on television: An evaluation of *Sesame Street* social goals programs, *A. V. Communication Review, 22 (3)*, 229 - 246.

Paulson, Pearl R. (1972). Paragraph summarization by computer: A comparison with summaries made by pupils. Unpublished doctoral disseration, Stanford University.

Phillips, D. C. (1990). Subjectivity and objectivity: An objective inquiry. In E. Eisner & A. Peshkin (Eds), *Qualitative inquiry in education: The continuing debate*. New York: Teachers College Press.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

Resnick, L. B. & Resnick, D. P. (in press). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C O'Connor (Eds), *Future assessments: Changing views of aptitude, achievement, and instruction*. Boston: Kluwer Academic Publishers. (Quoted in L. Sheppard (1990). Psychometricians' beliefs about learning. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 17, 1990.)

Rose, Frank (1984) *Into the heart of the mind: An American quest for artificial intelligence*. New York: Harper and Row.

Rottenberg, C. & Smith, M. L. (1990). Unintended effects of external testing in elementary schools. Paper read at the annual meeting of the American Educational Research Association in Boston MA.

Schank, Roger C. with Childers, Peter G. (1984). *The cognitive computer: On language, learning, and artificial intelligence*. Reading (MA): Addison-Wesley.

Schofield, Janet Ward (1990) Increasing the generalizability of qualitative research. In E. Eisner & A. Peshkin (Eds) *Qualitative inquiry in education: The continuing debate*. New York: Teachers College Press.

Schulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Scriven, Michael (1972). Pros and cons about goal-free evaluation, *Evaluation Comment, 3*, 1-7.

Shavelson, R. J., Carey, Neil B. & Webb, Noreen M (1990). Indicators of science achievement: Options for a powerful policy instrument. *Phi Delta Kappan*, May, 692 - 697.

Shavelson, Richard J., Webb, Noreen M., & Rowley, Glenn L. (1989). Generalizability theory. *American Psychologist, 44 (6)*. 922-932.

Shepard, Lorrie A. (1990) Psychometricians's beliefs about learning. Paper presented at the annual meeting of the American Educational Research Association, April 17, 1990.

Stake, Robert (1975). To evaluate an arts program. In R. Stake (Ed) *Evaluating the arts in education*. Columbus OH: Merrill,13-32.

Stake, Robert (1967). The countenance of educational evaluation. *Teachers College Record, 68 (7)*, 523-540.

Stevens, S. S. (1958). Problems and methods of psychophysics, *Psychological Bulletin, 55(4)*, 177 - 196.

Suen, Hoi K. & Davey, Bruce (1990). Potential theoretical and practical pitfalls and cautions of the performance assessment design. Paper read at the annual meeting of the American Educational Research Association, Boston MA.

Swing, S.R., Stoiber, K.C., & Peterson. P.L. (1988). Thinking skills vs learning time: Effects of alternative classroom-based interventions on students' mathematics problem solving. *Cognition and Instruction, 5 (2)*, 123-191.

The Senior Project (undated). *The senior project.* Information brochure published by Far West Edge, Inc., 1817 Woodlawn Ave., Medford OR 97504.

Thurstone, L.L. (1948) Psychophysical methods. In T. G. Andrews (ed) *Methods in psychology.* New York: Wiley, 124 - 157.

Tukey, John (1972). Data analysis, computation, and mathematics, *Quarterly of Applied Mathematics*, 51-65.

Weislogel, R. L. & Schwartz, P. A. (1955). Some practical and theoretical problems in situational testing. *Educational and Psychological Measurement, 15*, 39-46.

Wiggins, Grant (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, May 1989*, 703-713.

Weiss, Stefanie (1989). No more business as usual. *NEA Today, 7(7)*, 4 - 5.

Wolf, Robert L. (1975). Trial by jury: A new evaluation method, *Phi Delta Kappan, 57(3)*. 185-187.

Writing Assessment Leadership Committee (undated). *Vermont writing assessment program.* Montpelier VT: Vermont Dept. of Education.

Valencia, Shiela (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. *The Reading Teacher* January 1990, 338-340.

Appendix 16

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991