DOCUMENT RESUME

ED 376 209 TM 022 325

AUTHOR Paulson, F. Leon; Paulson, Pearl R.
TITLE Assessing Portfolios Using the Constructivist Paradigm.
PUB DATE Apr 94
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cognitive Psychology; *Constructivism (Learning); Educational Assessment; Educational Change; Elementary Secondary Education; Models; *Portfolio Assessment; Portfolios (Background Materials); Program Evaluation; *Psychometrics; *Test Use
IDENTIFIERS Atlanta Public Schools GA; Authentic Assessment; *Performance Based Evaluation; *Positivism; Wyoming School District MI

ABSTRACT

In a previous paper the authors presented a model of assessing portfolios, the Cognitive Model for Assessing Portfolios (CMAP), as a lens to view, think about, and make decisions about portfolio projects. Two projects in which the CMAP has been used are described. One is an authentic assessment project in mathematics and science by the Educational Testing Service in Atlanta (Georgia), and the other is the reform of assessments in the Wyoming (Michigan) school district. The clear focus of the Atlanta project is on student outcomes. The entire project is systematically designed to produce outcomes that can be summarized across individuals and groups. The focus of the Michigan project is far more diverse, with generally defined outcomes and a more general assessment design. Atlanta represents a positivist and psychometric model of assessment while Wyoming (Michigan) represents a constructivist point of view, with the portfolio seen as a learning environment. Portfolio assessment is better suited to the constructivist model, in that efforts to apply the positivist model to portfolios invariably impose meanings that are different from those created by the learner. One figure and one table illustrate the comparison. (Contains 27 references.) (SLD)

Assessing Portfolios Using the Constructivist Paradigm

F. Leon Paulson
Multnomah ESD

Pearl R. Paulson
Beaverton SD

In 1990, we participated in a conference on portfolio assessment sponsored by the Northwest Evaluation Association. The topic of the conference was aggregating portfolio data. By way of preparation, we wrote a paper called *How do Portfolios Measure Up? A Cognitive Model for Assessing Portfolios* [CMAP] (Paulson and Paulson, 1990). In our paper, we offered an assessment model loosely based on Robert Stake's (1967) evaluation model, one that could report descriptive or numerical material in ways that allowed educational activities to be assessed in context (see also P. Paulson and L. Paulson, 1991; and Paulson, Paulson, and Frazier, in press).

There were two reasons for creating this model. One was that we wanted an assessment model that would be comprehensive. Portfolios are by nature complex and holistic pictures of a child's learning and we wanted to provide an analysis model that preserved as much complexity as possible. Second, we were concerned about the aggregation of portfolio data. Efforts to aggregate data usually involve standardizing the events being aggregated. Our concern was that attempts to aggregate might change the very thing being aggregated. Thus, the second concern of our model was to be able to document the impact of the aggregation itself.

We intended CMAP as a means of describing portfolios and portfolio projects in context rather than as a cookbook for conducting portfolio assessments. Thus, CMAP was designed to be a lens to view, think about, and make decisions about portfolio projects.

Paper presented at the American Educational Research Association meeting in New Orleans, April 1994.

CMAP views a portfolio simultaneously from three perspectives or dimensions — (1) people influencing or being influenced by the portfolio (**stakeholders**), (2) activities involved in constructing the portfolio (**process**), and (3) the record of change it presents over time (**history**). These relationships are represented visually in Figure 1.

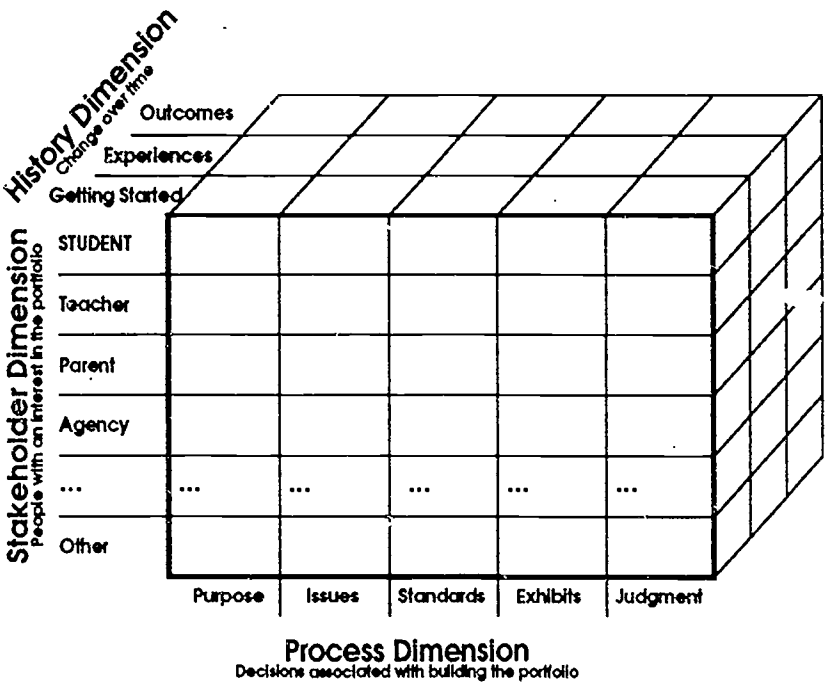

Figure 1.
The Cognitive Model for Assessing Portfolios showing the Stakeholder, Process, and History Dimensions

Two educators representing very different perspectives contacted us to say they were using CMAP in designing portfolio projects. Margaret Jorgensen, a researcher associated with the Atlanta Office of the Educational Testing Service, told us that she was using a somewhat modified version of CMAP to design an authentic assessment project in math and science that she was proposing for the National Science Foundation. She invited us to be consultants to the project. Anita Rutlin, an associate superintendent of the Wyoming, Michigan School District told us that the Wyoming district was making substantial changes in the way they conducted assessments and that she was using CMAP for guidance.

We tracked the development of these two projects with interest and, in the process, have become aware that they offer an interesting and instructive contrast. Both projects involve a multidimensional approach implied in the model. Yet, the differences in the way the model is interpreted are sharp. The Atlanta model (Jorgensen, 1993, 1994) is probably more familiar to those in educational research and development. It comes out of a long tradition of research and development: goals are established; work schedules set; and activities are carefully managed, monitored and reported. In contrast, the Wyoming model (Rutlin, personal communication) is probably more famil-

iar to practitioners concerned with ongoing educational programs. Like the Atlanta model, the Wyoming model is driven by a clear philosophic direction, but unlike the Atlanta model, management is decentralized and there is an expectation that the territory is unmapped and the route indirect. Let us take a look at some of the contrasts between the two projects.

A Comparison of Contrasting Portfolio Projects

Overall design

Both projects were concerned that assessment be more authentic even when large numbers of students were involved. Both were guided by a specific philosophy. While Atlanta was concerned with supporting teacher–assessment of many students across districts, Wyoming, Michigan, was concerned with supporting self–assessment by many students within its district. Atlanta was guided by a philosophy that placed priority on providing good assessment; Wyoming, Michigan, was guided by a philosophy that placed priority on supporting classroom learning.

For the Atlanta project, ETS developed an assessment plan, obtained necessary letters of support from local school districts, and submitted a proposal describing the project and how it would be implemented. The Atlanta model was designed to encourage input from stakeholders. The project called for assembling teams of teachers from the six participating districts. Its kickoff was a workshop during which teams selected curriculum goals that were common to all districts. The teams then developed definitions that became the design specifications for performance tasks, also developed by teachers. The majority of the assessment tasks would allow teachers to use content from their local district curriculum or even materials specially tailored to student needs or interests. Each task would yield concrete material suitable for inclusion in portfolios. Students would develop portfolios by completing a prescribed set of performance assessments. As the project proceeded, project staff sought input from stakeholders in the field of science and education, and took steps to make parents more fully aware of the project.

The Wyoming, Michigan, project began with a set of *General Learner Outcomes* adopted by the district's school board. These were general philosophical statements of what constitutes an educated student, which in turn identified curriculum areas that students would be expected to master during their years in school. While the general outcomes were set by the school board, there was a clear philosophy that the process of reaching those goals would be as decentralized as possible. In addition to expecting its students to demonstrate that they have achieved each outcome, Wyoming expected them to interpret the outcomes in ways that are personally meaningful. The teachers provide a curriculum that offers information and experiences that allow students to interpret the outcomes in a knowledgeable and socially responsible manner, and help them gather evidence for their portfolios. Michigan, like most states, publishes a large curriculum document describing goals

and objectives for students at each grade level and holds districts legally accountable for their attainment. In the Wyoming district, copies of this document are placed in each classroom as a resource document for *both* teachers and students. Teachers refer to this document as they prepare lesson plans and students refer to as they make decisions about the competencies they document through their portfolios. Table 1. summarizes the similarities and contrasts between the two programs.

	Atlanta, Georgia	Wyoming, Michigan
General Approach	To achieve consensus on general and specific outcomes to guide instruction and assessment.	To provide a philosophical context to guide instruction and assessment.
General Goals	Example Goal: *To develop students who are* *— effective collaborators* *— creative and strategic* *— reflective thinkers and self-evaluators* *— self-motivated learners* *— effective communicators* *— responsible global citizens*	General Learner Outcomes: *Effective communicator* *—Complex thinker* *— Creative thinker* *— Problem solver* *Self-directed learner* *Contributor to well being* *Cooperative worker* *Effective citizen*
Specific Goals and Standards	Project staff and teachers further define each goal, seeking an operational level, e.g., "collaborators:" *— recognizes self worth and that of others* *— believes that the collaborative result will be better than any single effort* *— demonstrates respect for self and others by accepting responsibility for collaborative participation* *— recognizes the right of all members to participate and have a vote*	Goals supplied to teachers and students who further define and operationalize. (District provides training and support.)
Evidence of attaining goal in relation to standard	Group creates rubric to score examples of student work. Examples come from tasks designed by teams of teachers at the district level. The portfolio is a collection of the tasks. Students may or may or may not select the contents of their portfolios.	The General Learner Outcomes are used in the classroom as teachers and students develop rubrics and set standards. Students in collaboration with teachers select evidence of the achievement of the goals. Students select the items that go into the portfolios.
Judgment	Trained teams of teachers score products on a individual basis. Psychometric standards apply to ensuring the validity of the process.	A complex judgment is made that includes the student as a self-assessor, peers, teacher, parents, and eventually people from the community who judge the students' portfolios.

Table 1.
A comparison of the two projects.

Stakeholders

An interesting question concerns what one means by stakeholder. One way of using the term would be as a way to refer to those who are interested and who the project must address as an *audience*. Another way of looking at stakeholders is to consider them as individuals who play an active role as *participants* in shaping the project. Let us look more closely at the contrasts between the Atlanta and the Wyoming approaches. Both projects focused on CMAP's stakeholder dimension as a place from which to start. The Atlanta project saw stakeholders (teachers, students, parents) as relatively discrete groups that districts could involve independently at many points in the process. For example, they decided to involve parents during the second year.

Wyoming, in contrast, tended to involve all stakeholders from the beginning. Parents were often included in the earliest discussions and some parts of the program were suggested by parents. They began trained individual teachers on portfolio processes and ideas but left the implementation of those ideas to the classroom. Gradually, through a process that was more evolutionary than anything else, certain practices emerged as more viable than others and were adopted in more and more classrooms.

Setting specific instructional goals

In the Atlanta project, each performance assessment produced information reflecting the curricular goal decided on by the centralized agency, a consensual decision by representatives of six school districts. The teams of teachers from participating districts assumed the challenge of defining the outcome goals in more and more specific terms and specifying processes and activities that would be used in all districts. These test–like situations sometimes used prescribed math and science content and sometimes allowed teachers to insert content from their own curriculum.[1] For example, an assessment task might call for scientific observation, but just what the student was observing might vary from classroom to classroom. In all cases, the activities prescribed what materials students would generate for their portfolios. The process was reminiscent of developing test items, only the items were performance assessments that could take place in the classroom.

In contrast, Wyoming provided its teachers with the most general statements of outcomes and invited them to translate them into classroom practice. No specific curriculum and no specific test–like situations were implied or described. Rather, teachers were encouraged to share the general outcomes with their students, and students working with teachers were encouraged to come up with definitions that were meaningful to the individuals involved. For example, one of the general learner outcomes was "cooperation." The district said only that students were required to learn to become cooperative citizens. It was the responsibility of the teacher and the

[1]Trevisan, Paulson, and Weaver (in preparation) used one of the activities, a rating of students' ability to make scientific observations, to compare two contrasting third grade science curricula and found it an efficient and sound performance assessment.

students to interpret the outcome and decide what evidence might be presented in order to show that this outcome was attained. Part of the learning for the student was to interpret the goal and to make it operational.

Project management

Project management also provides an interesting contrast. The Atlanta project was carefully managed and executed. Everything in the project was either planned in advance or modified from an existing plan. The Wyoming project, in contrast, seems to have "just grown" around a central philosophy of student ownership. The Wyoming approach became self-perpetuating in an intriguing way. The State of Michigan offered a series of small grants to support innovative educational programs. Anita Rutlin posted the notice of this grant on the district high school's bulletin board with a note saying that maybe these funds could be used to develop a handbook to help students put together portfolios. She received a response not from teachers but from students. The students, who thought having such a handbook would be valuable, wrote the proposal with Rutlin's encouragement and help. The proposal eventually become the first that the State of Michigan had awarded to a student initiated project. The students have completed the high school handbook and it is in use throughout the district. Parents requested a similar handbook for their elementary age children. A committee of parents, elementary teachers, and high school students prepared an elementary version based on the one developed by the high school students (Wyoming, 1993).

Performance Standards

A major contrast between the two projects involves the setting of standards. In the Atlanta project, standards were set by the teams of teachers who developed consensual rubrics that permitted student work to be judged in a reliable fashion. One of the chief concerns of the project was to develop assessment tasks that could be used over a variety of instructional activities and classrooms. The same teachers who developed activities also developed the rubrics. To refine the rubrics, they repeatedly collected and judged student responses. The rubrics and the training procedures were refined until all teachers and project consultants felt that they had achieved reliable measurement (defined as rater agreement). The rubrics were then field tested with different teachers and in other districts. Often the tasks were used across different instructional settings. Again, the characteristics were refined until a satisfactory level of rater agreement was observed. At that point the rubrics were offered for general use throughout the six participating districts and will be published for adoption anywhere.

Wyoming also encouraged the use of rubrics, but left design of the rubrics to the people in the classroom. For example, a state outcome has to do with cooperation (see Table 1). One teacher had her class brainstorm the characteristics of what they thought a cooperative person was. Another teacher taught a short story unit by first distributing examples of short stories (found in literature as well as collected from the previous year's students) and had

the students analyze the stories and generate statements characterizing good short stories. Each student then wrote an individual rubric to be used to judge their own writing during the remainder of the short story unit. In both cases, the rubrics were used for self–assessment, peer–assessment, and teacher–assessment in a way reminiscent of the *quality–control circle* found in Deming's writings (Deming, 1986) in which students were participants both in setting and applying the standards. Students revised their rubrics throughout the year as their understanding grew, and as their performance improved so did their standards. Thus, standards here are (1) individual to each student, (2) appropriate to the level of performance of each student, and (3) moving upward. The process is consistent with Deming's philosophy of quality control in which quality is controlled locally, not by centralized authority (See Deming, 1986; Gitlow and Gitlow, 1987; Paulson, 1993).

Toward a Constructivist Model for Assessing Portfolios

What do we learn from examining these two projects? First, the challenge of describing them is quite different. The Atlanta project's clear focus is on student outcomes of instruction and the entire project is systematically designed to produce student outcomes that lend themselves to summarization across individuals and across groups. Standard evaluation techniques found in most textbooks are well suited to analyzing this project. The Wyoming project's focus was far more diverse. There were outcomes, but they were very generally defined — certainly not in terms that lend themselves to measurement as we usually think of it. Much of the approach was designed by students and teachers as they went along, guided less by a specific assessment design than a clear statement of philosophy regarding instruction and learning.

Contrasting paradigms: Positivism and constructivism

The contrast here is between two models of assessment: the one (represented by Atlanta) we call positivist, the other (represented by Wyoming) we call constructivist (see Lincoln and Guba, 1985, p 14ff for a discussion contrasting these two approaches). Let us briefly define these two paradigms.

Positivist

> The purpose of the portfolio is to assess learning outcomes and those outcomes are, generally, defined externally. Positivism assumes that meaning is constant across users, contexts, and purposes (making it reasonable, for example, to think about national and even "world–class" standards). The portfolio is a receptacle for examples of student work used to infer what and how much learning has occurred.

Constructivist

> The portfolio is a learning environment in which the learner constructs meaning. It assumes that meaning varies across individuals,

over time, and with purpose. The portfolio represents process, a record of the processes associated with the learning itself and that a summation of individual portfolios would be too complex for normative description.

The two paradigms produce portfolio activities that are entirely different. Hansen (1993), in his anthropological analysis of testing, defines *test* as "a representational technique applied by an agency to an individual with the intention of gathering information" (p. 19). The positivist view of portfolios is consistent with this definition of assessment, the constructivist view is not. Hence, tensions develop around aggregation and other high stakes uses of portfolios.

- The positivist approach puts a premium on the selection of items that reflect outside standards and interests. Thus it is appropriate to include tests or test–like representational situations designed by others. Because outside interests and standards are applied, psychometric standards of reliability (especially inter–rater agreement) are emphasized in the judgments made about the products.
- The constructivist approach puts a premium on the selection of items that reflect learning from the student's perspective. Thus it is not appropriate to require tests or test–like representational situations. Because idiosyncratic standards play an important role, less emphasis is placed on consistency of judgments made about the products and more emphasis is placed on the perspectives represented by the judges.

Pamela Moss recently addressed the issues of how these two paradigms differ within the context of educational assessment.

> There are certain intellectual activities that standardized assessments can neither document nor promote; these include encouraging students to find their own purposes for reading and writing, encouraging teachers to make informed instructional decisions consistent with the needs of individual students, and encouraging teachers to collaborate in developing criteria and standards to evaluate their work. (1994, p. 6)

> ...most hermeneutic philosophers[2] share a holistic and integrative approach to interpretation of human phenomena that seeks to understand the whole in light of its parts, repeatedly testing interpretation against the available evidence until each of the parts can be accounted for in a coherent interpretation of the whole. (1994, p. 7)

The challenge posed by the new assessments

Portfolios, prominent in the movement to find alternative means for assessing students (see, for example, Wolf, Bixby, Glenn, and Gardner, 1992) do not fit easily into traditional concepts of how to go about assessing learn-

[2]We apply her discussion of hermeneutic (interpretive) philosophy to "constructivism."

ing. First, portfolios change the classroom environment (e.g., see Koretz, Stecher, & Diebert, 1992; Viechnicki, Rohrer, Ambrose, & Barbour, 1992), second, they stimulate and support integrative and reflective processes during the assessment (see Moss, 1994, Paulson & Paulson, in press) thereby changing the learning being assessed. Although portfolios have the potential of providing more authentic information on student performance than other, more contrived procedures associated with testing, assessment specialists find it difficult to apply rigorous standards associated with the psychometric paradigm (see Calfee and Perfumo, 1993; Koretz, Stecher, and Diebert, 1992).

Psychometric standards such as reliability require consistency to be defined quantitatively. This consistency increases generalizability across persons and situations, thereby supporting aggregation and other high stakes uses of assessment information. However, the less standardized forms of assessment give students considerable latitude in the interpretation, response to, and (especially in the case of many portfolios) to actually create the tasks. By doing so, they produce formidable challenges to reliability (See Koretz et. al 1993; Shavelson, R. J., Gao, X. and Baxter, G., 1993; and Suen and Davey, 1990).

Moss (1994) summarizes the general finding concerning reliability and the less standardized forms of assessment:

- Reliability defined as rater agreement on a single sample produces "acceptable" levels when rater agreement is calculated and the raters are acceptably trained.
- Reliability defined as rater agreement across tasks is much more difficult to estimate.[3]

Typically, the positivist, psychometric approach aspires to reliability by scoring each piece of the portfolio much as one would a single sample. Each example of student work is scored by readers with no outside knowledge about the learner or of other readers' judgments. It also assumes that each item stands alone and is meaningful in and of itself. Inferences are based on composite scores. The interpretability of these scores rests on previous research. The scores are provided to users with guidelines for interpretation. Users typically consider the scores in light of additional information about the individual, "although mainstream validity theory provides little guidance about how to combine such information to reach a well–warranted conclusion" (Moss, 1994, p 7). Often, concern for traditional measurement stan-

[3]Shavelson, et. al., (1993) conducted a generalizability study looking at sources of error in many kinds of performance assessments. They discovered that inter–rater agreement tended to be a negligible source of error but that error associated with task was considerable in itself and when it interacted with persons (students). They conclude, "In the end, task sampling variability appears to be a fact, not an artifact" (p. 23), underlining the blurring distinction between reliability and validity in performance assessment. The implications when the psychometric model is used can be summed as, "Regardless of subject matter (math or science), domain (education or job performance), or level of analysis (individual or school), large numbers of tasks are needed to get a generalizable measure of performance." (p. 22)

dards leads to pressures to make portfolios more like tests and less like comprehensive learning environments (see, for example, Koretz, McCaffrey, Klein, Bell, and Stecher, 1993). Constructivist (hermeneutic) analysis, however, seems better suited to large scale portfolio assessment even though rater agreement across tasks is more difficult to estimate.

The hermeneutic approach uses holistic interpretations of collected performances that seek to understand the whole in light of its parts. Items have meaning only in context and the context itself changes as an integral part of the process of making judgments. The interpretive approach benefits readers familiar with the context which is why the hermeneutic approach often includes detailed descriptions of the project. This grounds interpretation in context and encourages conversation (occasionally debate) among a community of interpreters. Thus low inter–rater agreement may actually signal a more insightful assessment especially if it leads to consensus, negotiation, or compromise (see F. L. Paulson and P. R. Paulson, 1991, Wolf, et. al., 1991).

There are many examples of this kind of assessment in practice. Moss (1994) uses the example of the way hiring decisions are made in higher education. Candidates for positions often assemble collections of their work that they think best document what they can bring to a position. Search committee members are selected not because they share a common, well trained perspective, but because they represent a broad area of expertise and interest. They do not agree to a common set of criteria or standards. Rather, they represent a collective expertise. Each member examines the full set of materials and together they make an integrative judgment about the candidate following negotiation and compromise. Moss suggests it would be unfair to seek only the same materials from each candidate and set the same judgment criteria to cover all, concluding "...permitting those assessed to choose products that best represent their strengths and interests may, in some circumstances, enhance not only validity but also fairness." (Moss, 1994, p. 8)

Ways portfolio assessment differs from other kinds of assessment

When we began working with portfolios, we were initially attracted to them as an assessment alternative to standardized tests. Like many other educators, we were concerned with the contrived and overly simplified way that standardized achievement tests defined and quantified learning. Our concerns were less with the assessment aspects of achievement testing than with the effects that achievement testing have on the quality of instruction. Resnick and Resnick (1992) have documented the affects of high stakes achievement testing on what and how things are taught in the classroom. Mary Lee Smith (1991) observed evidence of achievement testing leading to what she calls *multiple choice teaching*. Moss points a finger more directly at the assessment model used in constructing tests, noting that, "Current conceptions of reliability and validity in educational measurement con-

strain the kinds of assessment practices that are likely to find favor, and these in turn constrain educational opportunities for teachers and students" (1994, p 10). Even performance tests contain elements that impose the test writer's constructs on the student. Elsewhere (Paulson and Paulson, 1991), we have expressed concern that efforts to attain reliability through high inter-rater agreement may actually degrade the quality of assessment (see also, Wolf, et. al., 1991).

As soon as we began to think about portfolios, however, it became clear that they were much more than merely an assessment device. The strength in the portfolio concept was that it was a total learning environment that included assessment as one, but only one feature, albeit a central one. Within the context of a student portfolio, instruction and assessment coexisted in an extremely compatible manner.

Briefly put, we saw the educational power of portfolios in their traditional context of the arts as opposed to the traditional context of educational assessment. Educational assessment, invariably involved testing students and having other people make meaning of student performance. Remember Hansen's (1993) definition, tests are representational techniques that are externally imposed and constrained. But the kind of assessment observed in the context of portfolios involved creating, reflecting on, and evaluating meaning — where all stakeholders make meaning and the student changes the meaning while it's in the making!

While many educators would agree that it is important to teach students to be their own assessors, to us portfolio assessment offered an avenue to bring students into the very center of the assessment process. Thus, portfolio assessment became the model for a new kind of classroom assessment in which the student became a full stakeholder in the assessment process itself. In this model, the student assumes a leading role in the categories of activity that Robert Stake (1967) listed in his assessment model: setting the purpose, choosing goals, setting standards, collecting data, and interpreting the results. In *Sarah's Portfolio* (Paulson, Paulson, and Frazier, in press), we describe a portfolio used in this way.

Conclusions

How, the positivist tradition asks, can you have a highly individual document that will provide information that can be aggregated? Their answer is that you cannot, that is necessary to impose standardization (i.e., make portfolios representational and imposed — a *test* according to Hansen's definition). Testing is 'top–down' — something done to someone by another. But portfolios are 'bottom–up' — reflecting the desire of someone to communicate documentable information. By imposing standardization, you destroy the individuality of the portfolio thus subverting the very process you are trying to promote. While the standardization may have created something interesting and even useful for the purposes of assessment — that thing is

not a portfolio and cannot be expected to yield the benefits of the portfolio as a method of encouraging self–directed learning (Paulson, Paulson, and Meyer, 1991). At best, you produce only a complex performance assessment.

But in creating a complex performance assessment, you strip the portfolio of one of its major instructional benefits. The portfolio is a way of including students in the assessment process. It is a place where it is perfectly legitimate for the student to deliberately try to influence others' beliefs in what they know. The portfolio is a way of changing the relationship between the student as the assessment process itself — to turn it upside down to make the student a full and active partner in his or her own learning and the assessment thereof, including the design of the assessments that determine the standards and judgments that are reached.

Portfolio assessment is much better suited to the constructivist (hermeneutic) than it is the positivist (psychometric) model. The positivist model requires outcomes to be specified in advance while in portfolio assessment, outcomes, while guided by general purposes, both emerge and are redefined during the instructional process. Just as there are no two students the same, neither are there two portfolios the same, or outcomes the same. A strength of the hermeneutic model is its ability to support synthesis, to reassemble what analysis takes apart. The positivist model supports analysis much better than it handles synthesis. Portfolios are, by their nature, holistic. They are more than the sum of the parts. Efforts to apply the positivist model to portfolios invariably impose meanings on the portfolio that are different from those created by the learner.

As we followed the two projects, we see portfolios in the two projects as fitting two very different paradigms. The Atlanta, Georgia, project is concerned with large scale evaluation and became more and more like externally administrated collections of performance assessments. Yet portfolios from that project are also more amenable to psychometric analysis in the positivist paradigm. At the same time, we see the portfolios from Wyoming, Michigan, becoming more student–directed collections of authentic learning. They are more amenable to hermeneutic assessment consistent with the constructivist paradigm but have not been used this way in a large–scale assessment project. While several large scale assessment projects employ methods similar to the Atlanta project (e.g., projects in Vermont and Oregon), we have yet to see large–scale projects in which individual portfolios are evaluated using constructivist assumptions.

References

Calfee, R. and Perfumo, P. (1993). Student portfolios: Opportunities for a revolution in assessment. *Journal of Reading*, 36(7) , 532–537.

Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: MIT, Center for Advanced Engineering Study.

Gitlow, H. S., & Gitlow, S. J. (1987). *The Deming guide to quality and competitive position*. Englewood Cliffs, NJ: Prentice Hall.

Guba, E. and Lincoln, Y. (1989) *Fourth generation evaluation*. Newbury Park: Sage.

Hansen, F. A. (1993). *Testing Testing: Social Consequences of an Examined Life* . Berkeley: University of California Press.

Jorgensen, M. (1994). Alternative assessment from the perspective of six different school systems — an experiment in consensus building. Paper presented at a meeting of the American Educational Research Association in New Orleans.

Jorgensen, M. (1993). Authentic assessment for multiple users. Paper presented to the National Association for Research in Science Teaching and available from Educational Testing Service, 1979 Lakeside Pkwy, Suite 400, Atlanta, GA 30084.

Koretz, D., Stecher, B. and Diebert, E. (1992). *The Vermont portfolio assessment program: Interim report on the implementation and impact, 1991–92 school year* (CSE Technical Report 350). Los Angeles: UCLA, Center for the Study of Evaluation.

Koretz, D., McCaffery, D., Klein, S., Bell, R., and Stecher, B. (1993). *The reliability of scores from the 1992 Vermont portfolio assessment program* (CSE Technical Report 355). Los Angeles: UCLA, Center for the Study of Evaluation.

Lincoln, Y. E. and Guba, E G. *Naturalistic inquiry*. Beverly Hills: Sage.

Moss, Pamela A. (1994). Can there be validity without reliability? *Educational Researcher, 23(2),* 5-12.

Paulson, F. L. (1993). Bridging the gap: Applying Deming's 'Quality by design' to portfolios. Keynote address Portfolio Conference in the Heartland, sponsored by Heartland AEA, Johnston, Iowa, April 1993. (Available from ERIC Document Reproduction Service).

Paulson, F. L. and Paulson, P. R. (in press). Four Varieties of Self–Reflection. In K. Yancey (Ed). *Portfolios: Voices from the field:* Urbana, IL: NCTE.

Paulson, F. L. & Paulson, Pearl R. (1991, April). *The ins and outs of using portfolios to assess performance* (Revised). In *Measurement Issues in Performance Assessment*, a symposium conducted by the National Council on Measurement in Education, Chicago. (ERIC Document Reproduction Service No. ED 334 250).

Paulson, F. L., and Paulson P. R. (1990, August) How do portfolios measure up? A cognitive model for assessing portfolios . Paper presented at a conference Aggregating Portfolio Data sponsored by the Northwest Evaluation Association, Union, WA. (ERIC Document Reproduction Service ED 324 329)

Paulson, F. L., Paulson Pearl R., and Meyer C. A. (1991, February). What makes a portfolio a portfolio. *Educational Leadership, 46(6),* 60–63.

Paulson, P. R., Paulson F. L., and Frazier, D. F. (in press). Sarah's portfolio. In K. Yancey (Ed), *Portfolios: Voices from the field:* Urbana, IL: NCTE.

Paulson, P. R. & Paulson, F. L. (1991). Portfolios: Stories of knowing. In P. H. Dreyer (Ed), *Claremont reading conference 55th yearbook 1991. Knowing: The power of stories* (pp. 294-303). Claremont, CA: Center for Developmental Studies of the Claremont Graduate School.

Resnick L. B. and Resnick, D. P. (1992) Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford and M. C. O'Connor *Changing assessments: Alternative views of aptitude, achievement and instruction*. Boston: Klewer.

Rutlin, Anita (personal communication). For information, contact Wyoming Public Schools, 3575 Gladiola St. S. W., Wyoming MI 49504.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers *Educational Researcher, 20(5)*, 8–11.

Suen, H. K. and Davey, B. (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.

Trevisan. M., Paulson, F. L., & Weaver, K. (in preparation). An evaluation of a laser–disk science curriculum. Gresham School District, Gresham OR.

Viechnicki, K. J., Rohrer, J. , Ambrose, R. , and Barbour, N., (1992, April). The impact of portfolio assessment on teacher's classroom activities. Paper presented at the annual meeting of the American Educational Research Association in San Francisco.

Weiss, C. (1973). Where politics and evaluation research meet. *Evaluation,* 1 ,37–45.

Wolf, D., Bixby, J., Glenn, J., and Gardner, H (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed). *Review of research in education*. Vol. 17. Washington DC: American Educational Research Association.

Wyoming Public Schools (undated, ca. 1993). PEP Talk: Portfolio exhibit planner: Secondary edition. and PEP Talk: Portfolio exhibit planner: Primary edition. Author, 3575 Gladiola St. SW., Wyoming MI 49509.