DOCUMENT RESUME

ED 378 223 TM 022 592

AUTHOR Gearhart, Maryl; And Others
TITLE "Whose Work Is It?" A Question for the Validity of Large-Scale Portfolio Assessment.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY Apple Computer, Inc., Cupertino, CA.; Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CSE-TR-363
PUB DATE Jul 93
CONTRACT R117G10027
NOTE 14p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Competence; *Educational Assessment; Elementary Education; Elementary School Students; *Elementary School Teachers; Inferences; *Portfolio Assessment; Portfolios (Background Materials); *Scores; Student Evaluation; Teaching Methods; Test Validity; *Validity; *Writing Evaluation
IDENTIFIERS *Large Scale Programs

ABSTRACT

This study explored the meaningfulness of "student" scores derived from assessment of student portfolios. Nine elementary teachers documented the instructional support they provided for the writing assignments of each of six target students. Support ratings captured dimensions used to assess students' writing progress (Content/Organization, Style, Mechanics) as well as assignment challenge, the extent of copied work, and time required. Teachers' ratings tended to fall within the low to moderate range, varied with student writing competency, and showed marked variation among teachers. The study raises questions concerning validity of inferences about student competence based on portfolio work. Four tables. (Contains 21 references.) (Author)

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

Whose work is it?

A Question for the Validity of Large-Scale Portfolio Assessment

CSE Technical Report 363

Maryl Gearhart, Joan L. Herman
Eva L. Baker, and Andrea K. Whittaker

▸ UCLA Center for the Study of Evaluation

in collaboration with:

▸ University of Colorado
▸ NORC, University of Chicago
▸ LRDC, University of Pittsburgh
▸ The RAND Corporation

Whose work is it?

A Question for the Validity of Large-Scale Portfolio Assessment

CSE Technical Report 363

Maryl Gearhart, Joan L. Herman
Eva L. Baker, and Andrea K. Whittaker

July 1993

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein has received partial support from the Apple Classrooms of Tomorrow^{sm} Project, Advanced Development Group, Apple Computer, Inc., and from the Educational Research and Development Center Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of Apple Computer, the Office of Educational Research and Improvement, or the U.S. Department of Education.

Whose work is it?

A QUESTION FOR THE VALIDITY OF LARGE-SCALE PORTFOLIO ASSESSMENT *

Maryl Gearhart, Joan L. Herman,

Eva L. Baker, and Andrea K. Whittaker

Abstract

This study explored the meaningfulness of "student" scores derived from assessment of student portfolios. Nine elementary teachers documented the instructional support they provided for the writing assignments of each of six target students. Support ratings captured dimensions used to assess students' writing progress (Content/Organization, Style, Mechanics), as well as assignment Challenge, the extent of Copied Work, and Time required. Teachers' ratings tended to fall within the low to moderate range, varied with student writing competency, and showed marked variation among teachers. The study raises questions concerning validity of inferences about student competence based on portfolio work.

Recent debate surrounding writing assessment has addressed the appropriateness and meaningfulness of standardized direct assessments of children's writing. Criticisms of direct writing assessments focus on the limited time to accomplish the writing, the artificiality of the topics and assignments, and the restricted genres assessed (Freedman, 1993). Responses to criticisms have prompted a move toward further authenticity—performance-based assessments which may incorporate shared readings of common background texts, collaborative planning, and opportunities for revision. Portfolio assessment in particular represents the growing commitment to bridge between the worlds of public accountability and private

* Thanks to the teachers who served as our raters, and to John Novak for assistance with data analysis.

classroom, between the worlds of policymaker and child (Calfee & Perfumo, 1992; Camp, 1992, in press; Camp & Levine, 1991; Freedman, 1993; Hiebert & Calfee, 1992; Moss et al., 1991; Murphy & Smith, 1992; Simmons, 1990; Valencia, in press; Wolf, 1989).

But designing and implementing methods of large-scale portfolio assessment is a daunting challenge. Current efforts at the state and district levels are confronting multiple hurdles to implementation and technical quality (Koretz, McCaffrey, Klein, Bell, & Stecher, 1993; Koretz, Stecher, & Diebert, 1992; LeMahieu, 1992; Reidy, 1992). The design of large-scale portfolio assessments requires the development of performance standards, criteria for portfolio inclusions, and methods for scoring the resulting collections (Herman, Aschbacher, & Winters, 1992). There is as yet no consensus on how these goals can be achieved for diverse kinds of student work.

One issue frequently raised but not yet directly investigated concerns the authorship of classroom work (Condon & Hamp-Lyons, 1991; Gearhart, Herman, Baker, & Whittaker, 1992; Herman, Gearhart, & Baker, in press). When raters assess students' portfolios, whose work are they assessing? During classroom assignments, students may work with peers and receive assistance from teachers and parents. From assignment to assignment, the support provided by others will almost certainly vary. In addition, the support provided to particular students may vary—think of the student who always needs special help, or the student whose parent is overzealous in assistance at home. Finally, reflecting teachers' instructional philosophies, support will range from encouragement of student creativity to firm requirements and close monitoring.

For the study reported here, we documented patterns of instructional support across writing assignments, students who vary in grade and ability level, and teachers. Our purpose was to raise technical issues concerning the meaningfulness of "student" scores derived from assessment of student portfolios. If there is substantial variation in instructional support, what do ratings of portfolio contents reflect about *student* competencies?

Our Project

Our work stems from a long-term collaboration between the Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the

teachers of one elementary school to develop coordinated methods of portfolio assessment for uses at the classroom, school, and district levels (Baker, Gearhart, Herman, Tierney, & Whittaker, 1991).

The data reported here were collected from nine teachers spanning Grades 1–6 in the spring of 1991.

Methods

Target Students

In the fall of 1990, nine teachers were asked to designate two students at each of three levels of writing competency (high, medium, and low) and to collect complete portfolios of all of their work. Compliance was excellent, although teachers requested reclassification of a few students in the spring. The dataset for this study consisted of spring 1991 ratings of 228 assignments from a total of 54 students. The number of assignments per student ranged from 1 to 20, with a modal number of 3. (One teacher differed from all others, with 14 to 21 assignments per target student, compared with 1 to 5 for the remaining eight teachers.)

Ratings

Teachers rated the instructional support provided each target student's assignment during the composing and editing phases. Ratings were keyed to the same dimensions we used to assess students' writing progress (Baker, Gearhart, & Herman, 1992): *Content / Organization* (topic/subtopics or theme, and their structure, format, or arrangement); *Style* (elements of text like descriptive language, word choice, sentence choice, tone, mood, voice, and audience); and *Mechanics* (spelling, grammar, punctuation, and other conventions). As shown in Table 1, the scale points were defined along a continuum from 0 (no support) to 3 (teacher has specified the requirement in detail). Additional 0–3 ratings were made of: *Challenge* (the challenge of this assignment for this particular child) and *Copied work* (the extent to which the student's work appeared to be copied from peers or from direct modeling by a teacher or parent). Teachers estimated the *Time* the child spent on the assignment in hours or fractional parts of hours.

For each of the rating dimensions listed in Table 1, a weighted average was computed for each teacher to compensate for variation in the number of

Table 1

Instructional Support Rating Scheme

Support for organization/content: Topic, subtopics, theme, genres and their structure, format, or arrangement

3 Student provided detailed guidelines specifying the content and organization of the project (e.g., an outline showing what sections in what order).

2 Student provided with some T-prepared guidelines which may or may not have been elaborated during the prewriting phases.

1 Student provided with a minimal, brief, but reasonably structured assignment.

0 Student given no guidelines for this piece of writing.

Support for style: Descriptive language, varied word choice, varied sentence choice, tone, mood, voice, and audience

3 Student provided detailed guidelines and feedback on style.

2 Student provided some guidelines and feedback on style.

1 Student provided with general guidelines and reminders of those guidelines, e.g., "Use descriptive language. Don't forget to use dialogue. Show not tell."

0 Student given no guidelines or feedback for this piece of writing.

Support for mechanics: Grammar, spelling, punctuation, capitalization

3 Student provided with very detailed editing of mechanics.
2 Student provided with a moderate amount of editing of mechanics.
1 Student provided with a little editing of mechanics.
0 Student provided with NO editing of mechanics.

Level of challenge: How difficult was this task for this child?

3 Extremely difficult, frustrating
2 Moderately difficult, challenging
1 Not difficult, within the child's current level of competence
0 Extremely easy, no challenge whatsoever

Amount copied: Copying applies when students copy sentences or long phrases; using facts, terms, or words from a resource is not copying.

3 Copied almost everything. Little of the writing is the child's.
2 Copied a fair amount, but some of the writing is the child's.
1 Copied a little. Most of the writing is the child's.
0 Copied nothing, and all of the writing is the child's.
NA Not applicable. There were no opportunities for copying.

Time spent by the child:

Enter an estimate in hours or parts of hours.

---

assignments rated per child and the number of children designated as high, medium, or low in writing ability. Thus, for each student, a teacher's ratings were averaged across the student's assignments, and then a "mean of means" was computed for each teacher.

Results

Teachers' reported levels of instructional support tended to fall within the low to moderate range (Table 2). Teachers tended to provide less instructional support to "high" students than to "medium" and "low" students. Indeed, while support for high students was not likely to be rated at 2 or 3 (Content/ Organization 34%; Style 13%; Mechanics 26%), support for low students was frequently rated at 2 or 3 (Content/Organization 72%; Style 55%; Mechanics 60%).

Table 2

Teachers' Ratings for Students Judged as High, Medium, or Low in Writing Competency: Descriptive Statistics

	High			Medium			Low			Total		
Rating	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Instructional support ratings												
Content/ Organization	1.47 (19)	0.79	0–3	1.88 (17)	0.76	0–3	2.05 (18)	0.60	0–3	1.79 (54)	0.70	0–3
Style	1.01 (19)	0.70	0–3	1.54 (17)	0.57	0–3	1.78 (18)	0.71	0–3	1.43 (54)	0.73	0–3
Mechanics	1.05 (19)	0.72	0–3	1.60 (17)	0.49	1–3	1.94 (18)	0.61	0–3	1.52 (54)	0.72	0–3
Other ratings												
Challenge	1.22 (19)	0.59	0–3	1.68 (17)	0.52	1–3	1.86 (18)	0.64	0–3	1.58 (54)	0.64	0–3
Time (hours)	2.48 (19)	1.59	1–5	2.89 (17)	2.23	1–6	2.26 (18)	1.66	0–6	2.53 (54)	1.82	0–6
Copied work	0.41 (18)	0.50	0–3	0.54 (17)	0.75	0–3	0.75 (17)	0.82	0–3	0.56 (52)	0.70	0–3

Note. Means were computed as the group mean of each student's assignment mean. Numbers in parentheses are total number of students.

Teachers estimated that the assignments reflected low to moderate challenge for most students, that students spent an average of 3 hours on each assignment, and that the work reflected "a little" copying (Table 2). "High" students tended to be perceived as less challenged, as spending less time on their assignments, and as engaging in less copying.

For each of the instructional support variables, there was substantial variability. First, for each dimension of writing competence, teachers varied in their reported levels of support for students of different levels of ability (Table 3). For example, for Content/Organization ratings, Teachers A, D, and I differed little across students' ability levels, while Teachers C, E, and H reported markedly different levels of support.

Second, for each student ability level, teachers varied in the consistency of their support across dimensions of writing competence. For example, while Teachers C, E, and H appeared to provide consistently more assistance in all three categories to Low ability students, Teacher A provided High ability students less assistance on Mechanics, and Teacher B provided High students less assistance with Style.

Third, the patterns of teachers' ratings differed for teachers who varied in their experience with portfolio assessment (Table 4). Those three teachers who had been exploring portfolio assessment for a year and a half reported providing greater assistance than those teachers who had only recently agreed to participate. The difference may reflect the more experienced teachers' emphasis on a writing process approach to writing instruction, an approach which emphasizes teachers' involvement with students as they develop their compositions.

Discussion

Our results revealed variability in the amount of support teachers provide student work, in the time students spend on assignments, and in the extent to which students' work was copied from others. While this study was exploratory, we believe that the general pattern of these results will be confirmed. Future studies should consider larger sample sizes and additional methods of documentation to verify the variety of support provided students' classroom performance.

Table 3

Variation in Teachers' Instructional Support Ratings, Illustrated for Students Judged as High or Low in Writing Competency

		Content/ Organization		Style		Mechanics	
Teacher	Grade	High	Low	High	Low	High	Low
A	1	2.83 (2)	2.88 (2)	2.29 (2)	2.42 (2)	1.69 (2)	2.23 (2)
B	2	1.67 (2)	1.83 (2)	0.67 (2)	1.33 (6)	2.00 (2)	2.17 (2)
C	2	1.75 (2)	2.50 (2)	1.50 (2)	2.50 (2)	1.75 (2)	2.50 (2)
D	3	1.83 (2)	1.67 (2)	1.50 (2)	1.67 (2)	1.67 (2)	1.83 (2)
E	4	0.50 (3)	2.17 (4)	0.17 (3)	1.94 (4)	0.08 (3)	2.61 (4)
F	4	1.67 (2)	1.83 (2)	1.00 (2)	1.42 (1)	1.00 (2)	1.58 (2)
G	5	1.60 (1)	2.00 (2)	1.00 (2)	1.00 (2)	0.45 (2)	1.00 (1)
H	5	1.17 (2)	2.50 (2)	0.83 (2)	2.50 (6)	0.67 (2)	1.50 (2)
I	6	1.00 (2)	1.00 (1)	1.00 (1)	0.75 (2)	1.00 (1)	1.25 (2)

Note. Means were computed as the group mean of each student's assignment mean. Numbers in parentheses are total numbers of students.

Confirmation of our findings would certainly raise questions about the meaning we can ascribe to "student" work contained in portfolio collections. In our study, the quality of work appeared to be a function of substantial and uncontrolled support as well as student competence. Thus the validity of inferences we can draw about student competence based solely on portfolio work becomes suspect. While this is not a grave concern for classroom assessment where teachers can judge performances with knowledge of their context, the problem is troubling indeed for large-scale assessment purposes where comparability of data is an issue.

Table 4

Teachers' Ratings for Students Judged as High, Medium, or Low in Writing Competency: Comparison of Teachers With Greater or Lesser Portfolio Experience

Rating	High		Medium		Low		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Greater portfolio experience[a]								
Content/ Organization	2.05 (6)	0.76	2.13 (6)	0.56	2.18 (6)	0.56	2.12 (18)	.60
Style	1.32 (6)	0.80	1.66 (6)	0.73	1.72 (6)	0.59	1.57 (18)	.69
Mechanics	1.56 (6)	0.46	1.80 (6)	0.65	1.99 (6)	0.34	1.79 (18)	.50
Lesser portfolio experience[a]								
Content/ Organization	1.21 (13)	0.66	1.74 (11)	0.59	1.99 (12)	0.64	1.63 (36)	.70
Style	0.87 (13)	0.64	1.47 (11)	0.49	1.81 (12)	0.79	1.37 (36)	.75
Mechanics	0.80 (13)	0.70	1.48 (11)	0.37	1.92 (12)	0.72	1.38 (36)	.77

Note. Means were computed as the group mean of each student's assignment mean. Numbers in parentheses are total numbers of students.

[a] Greater portfolio experience = 1 1/2 years; Lesser = 1/2 year.

Thus, whose work is classroom work? It seems it depends—on the assignment itself, on the teachers' instructional interactions with particular students, on peer and other resources available within the classroom, on the structure provided in the instructional process. If portfolio assessments are to be used to rank or make serious decisions about students, school, or districts, portfolio ratings could be adjusted to reflect differences in support and assignment difficulty. Whether making such adjustments is feasible, adjustments of some kind will be necessary to assure comparability of results.

References

Baker, E.L., Gearhart, M., & Herman, J.L. (1992). Apple Classrooms of *Tomorrow*^{sm}: *The 1989-1990 evaluation report*. Los Angeles: University of California, Center for Technology Assessment/Center for the Study of Evaluation.

Baker, E.L., Gearhart, M., Herman, J.L., Tierney, R., & Whittaker, A.K. (1991). Stevens Creek portfolio project: Writing assessment in the technology classroom. *Portfolio News*, *2*(3), 7-9.

Calfee, R.C., & Perfumo, P.A. (1992). *A survey of portfolio practices*. Berkeley: University of California, Center for the Study of Writing.

Camp, R. (1992). Changing the model for the direct assessment of writing. In M. Williamson & B. Huot (Eds.), *Holistic scoring: New theoretical foundations and validation research*. Norwood, NJ: Ablex.

Camp, R. (in press). The place of portfolios in our changing views of writing assessment. In R. Bennett & W. Ward (Eds.), *Construction vs. choice in cognitive measurement*. Hillsdale, NJ: Erlbaum.

Camp, R., & Levine, D.S. (1991). Portfolios evolving: Background and variations in sixth- through twelfth-grade classrooms. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and product* (pp. 194-205). Portsmouth, NH: Heinemann.

Condon, W., & Hamp-Lyons, L. (1991). Introducing a portfolio-based writing assessment: Progress through problems. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and product* (pp. 231-247). Portsmouth, NH: Heinemann.

Freedman, S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment*, *1*(1), 27-52.

Gearhart, M., Herman, J.L., Baker, E.L., & Whittaker, A.K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (Tech. Rep. No. 337). Los Angeles: University of California, Center for the Study of Evaluation.

Herman, J. L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Herman, J.L., Gearhart, M., & Baker, E.L. (in press). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*.

Hiebert, F., & Calfee, R.C. (1992). Assessment of literacy: From standardized tests to performances and portfolios. In A. E. Fastrup & S. J. Samuels (Eds.), *What research says about reading instruction* (pp. 70-100). Newark, DE: International Reading Association.

Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont Portfolio Assessment Program* (CSE Tech. Rep. 355). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., & Diebert, E. (1992). *The Vermont Portfolio Assessment Program: Interim report on implementation and impact–1991-92 school year* (CSE Tech. Rep. 350). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

LeMahieu, P. (1992, April). *Perspectives on the assessment value of portfolios: How robust are they?* Presentation at the annual meeting of the American Educational Research Association, San Francisco.

Moss, P.A., Beck, J.S., Ebbs, C., Herter, R., Matson, B., Muchmore, J., Steele, D., & Taylor, C. (1991, April). *Further enlarging the assessment dialogue: Using portfolio to communicate beyond the classroom.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Murphy, S., & Smith, M. (1992). Looking into portfolios. In K.B. Yancey (Ed.), *Portfolios in the writing classroom* (pp. 49-61). Urbana, IL: National Council of Teachers of English.

Reidy, E. (1992, June). *What does a realistic state assessment program look like in the near term?* Paper presented at the annual ECS/CDE meeting, Boulder, Colorado.

Simmons, J. (1990). Portfolios as large-scale assessment. *Language Arts, 67*, 262-268.

Valencia, S.W. (in press). Portfolios: Panacea or Pandora's box? In F. Finch (Ed.), *Educational performance testing*. Chicago, IL: Riverside.

Wolf, D.P. (1989, April). Portfolio assessment: Sampling student work. *Educational Leadership, 46*(7), 4-10.