CRESST REPORT 736

Maryl Gearhart
Ellen Osmundson

ASSESSMENT PORTFOLIOS AS OPPORTUNITIES FOR TEACHER LEARNING

JUNE, 2008

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

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

Assessment Portfolios as Opportunities for Teacher Learning

CRESST Report 736

Maryl Gearhart
CAESL/University of California, Berkeley

Ellen Osmundson
CRESST/University of California, Los Angeles

June 2008

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 2008 The Regents of the University of California

This study was supported by the National Science Foundation under a grant to WestEd for the Center for Assessment and Evaluation of Student Learning (CAESL), award number #01093264.

The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the positions or policies of the National Science Foundation, or WestEd for the Center for Assessment and Evaluation of Student Learning (CAESL).

ASSESSMENT PORTFOLIOS AS OPPORTUNITIES FOR TEACHER LEARNING[1]

Maryl Gearhart
CAESL/University of California, Berkeley

Ellen Osmundson
CRESST/University of California, Los Angeles

Abstract

This report is an analysis of the role of assessment portfolios in teacher learning. Over 18 months, 19 experienced science teachers worked in grade-level teams to design, implement, and evaluate assessments to track student learning throughout a curriculum unit, supported by semi-structured tasks and resources in assessment portfolios. Teachers had the opportunity to complete three assessment portfolios for two or three curriculum units. Evidence of teacher learning included (a) changes over time in the contents of 10 teachers' portfolios spanning Grades 1–9 and (b) the full cohort's self-reported learning in surveys and focus groups. Findings revealed that Academy teachers developed greater understanding of assessment planning, quality assessments and scoring guides, strategies for analysis of student understanding, and use of evidence to guide instruction. Evidence of broad impact on teacher learning was balanced by evidence of uneven growth, particularly with more advanced assessment concepts such as reliability and fairness as well as curriculum-specific methods for developing and using assessments and scoring guides. The findings point to a need for further research on ways to balance general approaches to professional development with content specific strategies to deepen teacher skill and knowledge.

Introduction

In this report we examine ways that assessment portfolios can support experienced science teachers in their efforts to build assessment expertise. Portfolios are widely used in both preservice and professional development contexts to support teachers' reflection on their instructional practices (Mansvelder-Longayroux, Beijard, Verloop, & Vermunt, 2007; Zeichner & Wray, 2001), but, with the exception of their use in one preservice classroom assessment course (Taylor, 1997; Taylor & Nolen, 1996a), portfolios have been rarely used as the principal resource for learning about classroom assessment. In the professional development program we investigated called the Assessment Leadership Academy, assessment portfolios provided science teachers opportunities to learn new assessment

[1]We are grateful to the participating teachers, the professional development team, and our research team for their contributions to the findings reported here. The professional development team was co-directed by Kathy Diranna (WestEd) and Craig Strang (Lawrence Hall of Science), and the team consisted of Diane Carnahan, Karen Cerwin, and Jo Topps of WestEd, and Lynn Barakos of Lawrence Hall of Science. Researchers (in addition to those listed as authors) included: Shaunna Clark, Joan L. Herman, Sam Nagashima, and Terry Vendlinski from UCLA, and Diana Bernbaum, Jennifer Pfotenhauer, and Cheryl Schwab from U.C. Berkeley. Joan L. Herman provided invaluable feedback on this paper.

concepts and practices, and apply their learning to the design and implementation of assessment plans for curriculum units. We investigated the ways in which the Academy assessment portfolio supported teacher learning about assessment.

Our report is organized in three sections:

1. The Introduction provides a description of the Academy portfolio, its conceptual framework, and the professional development strategies designed to support teachers' uses of the portfolios.
2. We also review prior studies of similar professional development approaches to set our investigation in the context of what is already known about assessment-focused professional development. The Findings report evidence of teacher learning from analyses of the portfolios as well as teachers' self-reports in surveys and focus groups.
3. We conclude with reflection on the opportunities and constraints of a portfolio-based program for supporting the growth of teachers' assessment expertise.

Background

The Assessment Leadership Academy was an 18-month program in 2003–05 that engaged 23 experienced science teachers in the construction of assessment portfolios for their curriculum unit.[2] The Academy portfolio was designed as a "learning portfolio" (Mansvelder-Longayroux et al., 2007; Wolf & Dietz, 1998) rather than an evaluative portfolio for monitoring teacher performance. Organized as a semi-structured series of tasks and resources, the portfolio supported teachers in the design, implementation, and evaluation of assessments to track student learning and progress throughout a curriculum unit. Teachers used the portfolio process to develop and implement assessments for curriculum units of their choosing. Integrated with teachers' practice, the portfolio was the context for teachers to apply and reflect on what they were learning about quality assessment in the Academy institutes.

The Academy's choice of curriculum unit portfolios (rather than lesson or module portfolios) to document teacher thinking and learning about assessment was strategic, a decision that was grounded in three tenets of the Academy's theory of action. The first tenet was the critical function of *formative assessment* in effective teaching (Atkin, Coffey, Moorthy, Sato & Thibeault, 2005; Bell & Cowie, 2001; Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998; Wiliam, Lee, Harrison, & Black, 2004). With the support of Academy portfolios, teachers gathered and used ongoing information on student learning, including the alternative conceptions that students construct as they build understanding of complex science ideas (National Research Council [NRC], 2001a, 2001b).

[2] Five districts sent K-12 district teams consisting of several teachers and one administrator (typically a district science or assessment specialist). Our research focused only on teachers.

The second tenet was that teachers need to abandon their delivery orientation to curriculum and *take ownership of their materials* (Diranna et al., 2007), including the embedded assessments. As Academy teachers revised lessons and assessments over several portfolio cycles, they came to view revisions as appropriate and necessary for particular instructional or assessment purposes. The final tenet was that ongoing *reflective practice* (e.g., Schön, 1983, 1987) is essential to the professional work of teaching. Opportunities for reflection were embedded throughout the Academy portfolio cycle of unit planning, implementation, evaluation, and refinement.

Conceptual framework. The Academy choice of a portfolio strategy for professional development shaped the Academy's goals for teacher learning and its conceptual framework. Because the portfolio contained documentation of written assessments (more easily archived, analyzed, and transported in hard copy than audio or video data), the focus was on the assessment expertise that is important for designing and implementing written assessments.

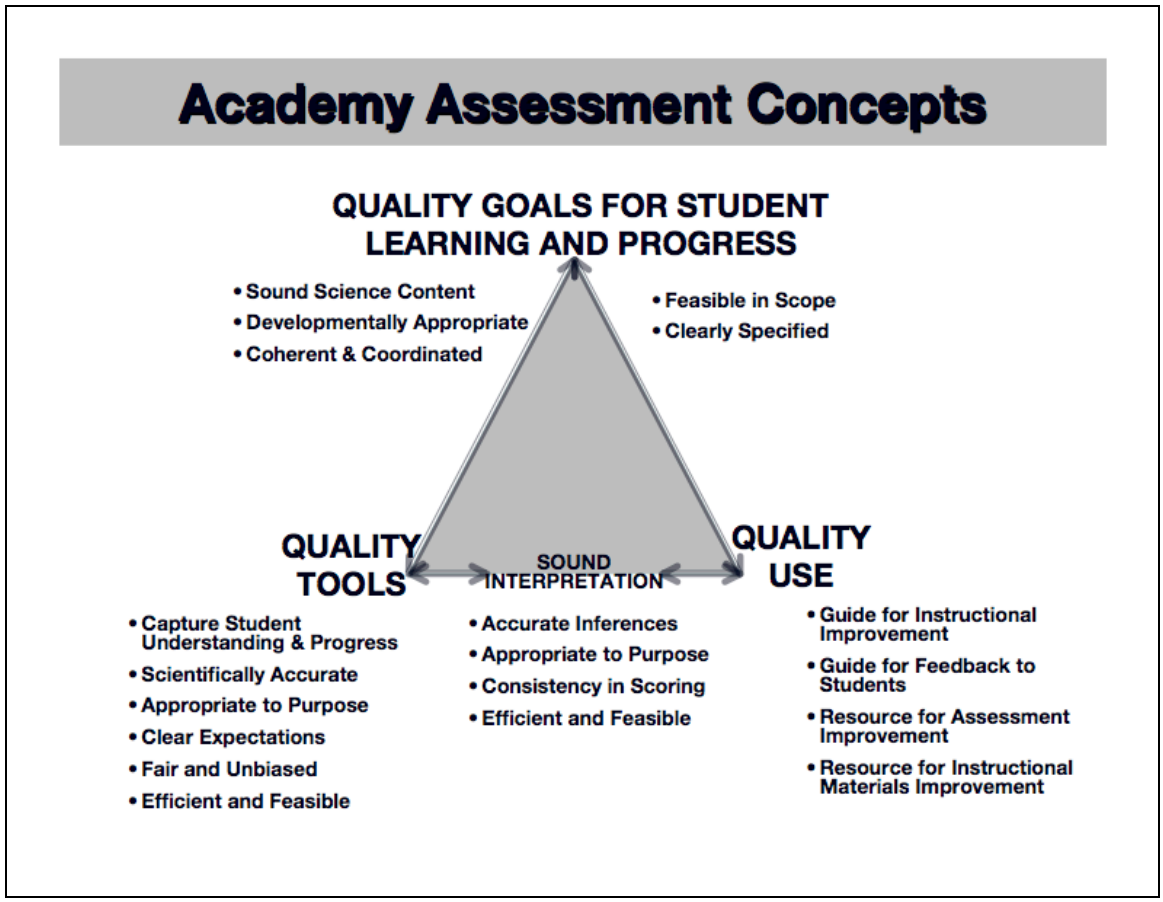

Figure 1. Academy framework for important classroom assessment concepts.

The Academy portfolio framework was based on theory and research from both the psychometric (American Educational Research Association, American Psychological

Association, and National Council on Measurement and Education 1999; Brookhart, 2003; Popham, 2004; Shepard, 2001; Stiggins, 2005; Taylor & Nolen, 1996b; Wilson & Sloane, 2000) and practitioner traditions (Atkins et al., 2005; Black et al., 2003; Black & Wiliam, 1998; NRC, 2001a; Watson, 2000). It was designed to capture relationships between teachers' *understanding of assessment concepts* (Figure 1) and their *skill with assessment practices* (Figure 2).

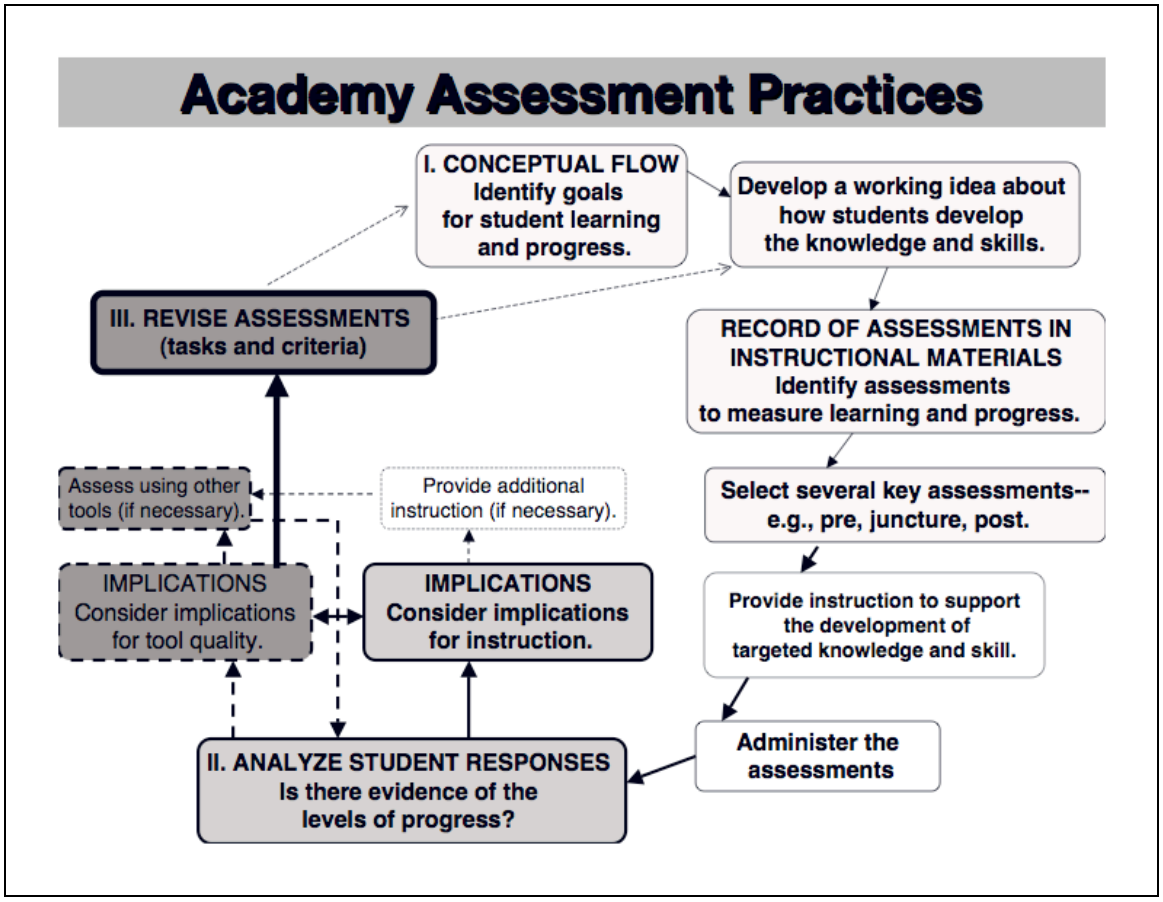

Figure 2. Academy framework for classroom assessment practices integrated with instructional practices.

The network of interconnected assessment concepts in Figure 1 was closely informed by the assessment triangle in *Knowing What Students Know* (NRC, 2001b), but modified to place emphasis on use of assessment information to guide instruction. The core idea was that quality classroom assessment requires a coordination of: (a) clear and valued goals for student learning (NRC, 1996), (b) quality tools for gathering evidence of student learning, sound interpretations of the evidence, and (c) quality uses of the information to guide

instruction and provide students useful feedback.[3] Sub-concepts associated with these major ideas were depicted, and double arrows represented the importance of alignment among all components of assessment.[4] Figure 2 represents classroom assessment practices embedded in a cycle of continuous instructional improvement. Planning begins (at the top of the figure) when Academy teachers identify their learning goals for a science unit and develop an integrated instruction and assessment plan (cf. Wiggins & McTighe, 2005). Implementation entails: repeated cycles of instruction, assessment using a variety of assessment strategies (Stiggins, 2005), interpretation of evidence, and use of information to guide teaching, learning, and further assessment. The bidirectional arrows indicate ongoing formative assessment and instructional improvement throughout the unit.

Portfolio design. The Academy portfolio[5] contained three sections corresponding to the phases in Figure 2, and, within each phase, teachers were asked to reflect on relevant assessment concepts in Figure 1. Table 1 outlines the portfolio sections, tasks, and key assessment concepts in the first three columns.[6] Below we describe the purpose, tasks, and organization of each of the portfolio sections.

[3] The ideas in the framework are somewhat simplified in relation to more comprehensive treatments of classroom assessment (e.g., Stiggins, 2005; Taylor & Nolen, 2004). Omitted or backgrounded are certain technical ideas, students' roles in assessment, and assessment systems that coordinate formative and summative assessments. On the other hand, the idea of 'developmentally sound content' was more emphasized than in other assessment projects, because the Academy was invested in helping teachers interpret student progress along a developmental continuum of understanding (Herman, 2005). For example, during the planning phase when Academy teachers were evaluating the quality of potential assessments, teachers drafted a range of 'expected student responses' to evaluate the capacity of the assessment to provide information on the developmental range of understanding, while, in other settings, teachers are often advised just to write out the correct answers when evaluating assessment items (Taylor & Nolen, 2004).

[4] The figures merge several versions shared with teachers over 18 months as the framework evolved in part through teacher input. Herman (2005) provides a detailed exposition of one version of the framework, and DiRanna et al. (in press) introduces a modified version.

[5] The Academy assessment portfolio differed from the preservice model developed by Taylor and Nolen in two ways (Taylor, 1997; Taylor & Nolen, 1996a). First, it was not a context for feedback by the professional development team; the Academy goal was to promote professional reflection and collaboration, and the team wanted to minimize concerns about evaluation. Second, it was a more ambitious undertaking than Taylor and Nolen could accomplish within a 10-week academic term: The Academy portfolio documented the design of unit assessments, implementation of assessments, and evaluation/refinement of assessments, while Taylor & Nolen's preservice portfolio contained just a unit plan (although the plan was in some ways more comprehensive than the Academy's).

[6] The portfolio forms and tasks were modified twice over the 18-month Academy program, and Table 1 represents an amalgam of the three versions. Information on the evolution of the portfolio is available from the authors. DiRanna et al. (in press) introduces a further evolution of the portfolio.

Table 1

Academy portfolio tasks, assessment concepts, and support for teacher learning.

Portfolio tasks and concepts			Sources of support		
Section	Tasks	Assessment concepts	Portfolio	Institute & facilitator	Other resources
I PLANNING 1 to 3 days in Academy institutes	**Establish unit learning goals** Construct a conceptual flow (CF) to represent learning goals **Select assessment tools** Develop a "record of assessments in instructional materials" (RAIM) to evaluate and select (or design) assessments aligned with learning goals **Draft criteria** Prethink scoring criteria by drafting expected student responses (ESRs) to assessments	**Quality goals** • Coherent and coordinated • Developmentally appropriate • Sound science content • Clearly specified alignment • Tools aligned with learning goals **Quality tools** • Capture the full range of student understanding and progress • Appropriate to purpose • Clear expectations for quality response • Fair and unbiased	**Moderate portfolio support:** • Portfolio prompts keyed to Academy assessment concepts to guide each step of tool selection and refinement	**Extensive facilitator and collegial interaction that varied in content support[a]:** • Academy facilitators[a] with curricular and content expertise • Colleagues[a] in grade level collaborative teams planning the same curriculum units	**Uneven support from other resources:** • Some models of conceptual flows • Teachers' instructional materials that may include an overview of unit goals and embedded assessment tasks and criteria • Science standards and benchmarks • Books such as National Research Council's book on classroom assessment (NRC, 2001a)

[a]Facilitators and teachers varied in levels of content and assessment expertise.

(table continues)

Table 1 (continued)

Portfolio tasks and concepts			Sources of support		
Section	Tasks	Assessment concepts	Portfolio	Institute & facilitator	Other resources
II IMPLEMEN-TATION Generally independent work in classrooms	**Develop criteria from ESRs based on student responses:** • Sort papers into H, M, L. • Examine patterns in relation to ESRs • Revise criteria Score student work Record results in matrix, using/adapting illustrated techniques Interpret patterns and trends Use evidence to provide students feedback and improve instruction	**Quality tools (criteria)** • Capture the full range of student understanding and progress • Appropriate to purpose • Clear expectations for quality response • Fair and unbiased **Sound Interpretation** • Appropriate to purpose • Consistency in scoring • Accurate & unbiased inferences **Quality use** • Guide for instructional improvement • Timely, relevant feedback • Resource for assessment improvement	**Portfolio support greater for interpretation than use:** • Models of criteria, records, and analytic/interpretive techniques tied to Academy assessment concepts[b] • Queries about teachers' feedback to student and uses of evidence to improve instruction	**Limited facilitator & collegial interaction[c]:** • Brief opportunities to discuss techniques for interpreting student work and findings • Brief opportunities to discuss feedback and instructional improvement	**Uneven support from other resources** (depending on relevance and accessibility): • Teachers' instructional materials that may include criteria • One session of on-site facilitator coaching for interpretation of student work (not available to all teachers) • Some teachers shared findings with colleagues on site or in district meetings. • Handout with list of suggested strategies for instructional follow-up • Books such as National Research Council's book on classroom assessment (NRC, 2001a)

[b]Support was strengthened in later versions of the portfolio.

[c]Focus of most institutes was either Planning (I) or Tool Revision (III).

(table continues)

Table 1 (continued)

Portfolio tasks and concepts			Sources of support		
Section	Tasks	Assessment concepts	Portfolio	Institute & facilitator	Other resources
III TOOL REVISION 1 to 1/½ days in Academy institutes	**Revise tools based on students' responses:** • Review student work • Analyze tasks • Revise tasks • Revise scoring criteria	**Quality Tools** • Capture the full range of student understanding and progress • Appropriate to purpose • Clear expectations for quality response • Fair and unbiased alignment • Tools aligned with learning goals	**Moderate portfolio support[d]:** • Forms to evaluate the quality of the tools, and guide tool revision based on Academy assessment concepts	**Extensive facilitator and collegial interaction that varied in content support[e]:** • Academy facilitators with curricular and content expertise[e] • Colleagues[e] in grade level collaborative teams who brought tasks and student work from the same curriculum units	**Limited support from other resources (depending on relevance and accessibility):** • Some models of quality tasks and criteria • Books such as National Research Council's book on classroom assessment (NRC, 2001a)

[d]Support was strengthened in later versions of the portfolio.
[e]Facilitators and teachers varied in levels of content and assessment expertise.

Section I contained the unit plan. At the beginning of each portfolio cycle, Academy teachers were organized as cross-district, grade-level teams to plan learning goals and assessments for a curriculum unit for their grade level. Time provided for unit planning varied from 1 to 3 days. Facilitated by Academy staff or occasionally one of the researchers, each team specified learning goals and represented the goals as a "conceptual flow." Then, using the Record of Assessments in Instructional Materials (RAIM) forms, teachers located possible paper-pencil assessments in their units and selected a series of assessments aligned with key unit goals to track student progress from a pre-assessment, through interim juncture assessments, to a final post-assessment. Guided by RAIM prompts linked to Academy concepts, teachers evaluated the quality of their selected assessments, and a key step in that process was drafting Expected Student Responses (ESRs) to 'prethink' student responses and gauge the likelihood that the assessment would elicit and measure the full range of student understanding. If teams had concerns about the available tasks or criteria, they refined them or designed their own assessments. The resulting assessment plans incorporated both formative and summative as key components of a quality system (cf. Stiggins' [2005] notion of "balanced" assessments "for learning" and "of learning"). Each teacher filed a copy of the team's collaboratively-constructed plan in his or her individual portfolio.

Section II was devoted to interpretation of student work and use of information to guide instruction. Teachers returned to their classrooms to implement the assessments, and most of their work in Section II was completed independently, although a member of the professional development team visited some teachers once for on-site coaching. Time was occasionally provided for discussion of student work in district or institute meetings. The portfolio provided teachers with strategies for interpreting student work: how to construct criteria by modifying expected student responses based on patterns in the student work; procedures for scoring responses; ways to record scores and qualitative notes; and methods of analyzing patterns and trends. Portfolio prompts reminded teachers to document their strategies for interpreting student responses, their inferences, and the ways they used the information to give students feedback and revise instruction. Teachers archived the assessments and their reflections as well as copies of the student work in their portfolios.

Section III contained revisions of the assessments. After teachers implemented their units, grade level teams reconvened at institutes, and facilitators guided teams through a 1- to 2-day process of evaluating and revising their assessments based on students' responses to the assessments. Reflective prompts helped teachers evaluate and then strengthen the quality of their assessments in light of the key assessment concepts in Figure 1. Teams documented

assessment revisions in their portfolios, and each teacher filed a copy of the team's joint work in his or her portfolio.

As outlined in the last two columns of Table 1, teachers completed their portfolios with support from the portfolio forms (including models and reflective prompts), interactions with facilitators and colleagues, and supplemental resources. The extent and nature of Academy support varied for different sections of the portfolio. Section I forms were skeletal as most of the work of assessment planning was facilitated. Interpretation of student responses in Section II was scaffolded by detailed forms that outlined step-by-step methods for developing criteria and whole class analysis, while support for Use in Section II was limited to open-ended queries about teachers' uses. Section III provided teachers with a detailed tool for evaluating the quality of assessments and making appropriate revisions. Table 1 also outlines how teachers' opportunities for learning from the portfolio were coordinated with supports from other resources, including team members, facilitators, and the instructional materials themselves. Teachers completed Section II independently in between Academy institutes, while Sections I and III were completed collaboratively at the institutes with the support of a facilitator. Neither the portfolio nor the Academy provided targeted support for science content knowledge and pedagogical content knowledge, although Sections I and III were contexts for intensive institute discussions of the science and the ways students learn as teams identified unit learning goals, and analyzed and revised assessments.

Prior Research on Professional Development: Setting the Academy Portfolio Strategy in Context

While the Academy assessment portfolio was an innovation, many other features of the Academy program were based on best practices culled from existing research on professional development (Birman, Desimone, Garet, Porter, & Yoon, 2001; Garet, Porter, Desimone, Birman, & Yoon, 2001; Guskey, 2003; Hawley & Valli, 1999; Laguarda & Anderson, 1998; Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003; Wilson & Berne, 1999). First, teachers' opportunities to learn were collaborative and sustained; for 2 years, the Academy supported professional communities both within the Academy and the participating school districts, and the portfolio served as a critical resource that traveled from one professional context to another, supporting different kinds of teacher interaction and work. Second, teacher reflection on practice was embedded throughout the portfolio and institute activities. Third, opportunities for teacher learning were a balance of expert guidance and teacher autonomy; during the institutes, facilitators guided collaborative work on the portfolios, but teachers were individually responsible for implementing the assessments, constructing criteria, analyzing student responses, and documenting their analysis in their portfolios. The

Academy design was, however, weakly aligned with current recommendations to build content knowledge for teaching (Ball, Hill, & Bass, 2005; Hill & Ball, 2004; Weiss & Miller, 2006). Academy teachers certainly engaged in content-rich reflection on learning goals, assessments, and student work as they worked on their portfolios. But teachers were working on assessment portfolios for a wide variety of curriculum units at any given time, and therefore the Academy was unable to organize systematic and targeted, unit-specific experiences for teachers to build knowledge of science, the ways that students learn specific science concepts and processes, and ways to assess based on a developmental continuum of understanding.

Assessment-focused professional development. Prior studies of assessment-focused professional development have shown that teachers can gain assessment expertise through the activities like those embedded in the Academy portfolio, including clarifying learning goals, developing assessment tools, and interpreting and utilizing evidence. The Academy portfolio's particular focus on paper-pencil assessment tasks built on research during the performance assessment movement in the 1990s when teachers collaborated to refine benchmark performance tasks and scoring guides prior to implementation, and again to score student work and consider the implications for instructional improvement (e. g., Falk & Ort, 1998; Sheingold, Heller, & Paulukonis, 1995). Studies in this era reported generally positive impact of project participation on teachers' assessment and instructional practices.

However, there were evident barriers to teacher learning, especially the weak alignment of large-scale performance assessments with classroom curriculum (Aschbacher, 1999; Borko, Mayfield, Marion, Flexer & Cumbo, 1997; Falk & Ort, 1998; Gearhart & Saxe, 2004; Goldberg & Roswell, 1999-2000; Laguarda & Anderson, 1998). The Academy addressed the alignment issue by creating a portfolio tool that engaged teachers in the design and use of assessments for their own curriculum units. In this regard, the Academy portfolio's emphasis on the deep integration of assessment and curriculum was consistent with recent efforts to embed quality assessment systems in science units to help teachers track student progress and support student learning (Aschbacher & Alonzo, 2006; Herman, Osmundson, Ayala, Schneider & Timms, 2005; S. M. Wilson, 2004; M. Wilson & Sloane, 2000). The Academy portfolio, however, was a generic learning tool for assessment development and use, while curriculum-embedded assessments provide teachers with a complete assessment system.

When we consider the Academy in relation to the projects just cited, the Academy's mission appears very ambitious. In other projects, teachers generally focused on developing assessment knowledge and expertise for a limited number of tools, while the Academy's goal was to engage teachers in developing and implementing coherent assessment plans for entire

curriculum units through the construction of unit assessment portfolios. The Academy team was well aware that teachers have limited experience evaluating, refining, and using quality assessments, but they argued that, because most science units lack quality assessments, teachers need to build the expertise to strengthen the assessments in their instructional materials. The intended outcomes of the Academy portfolio strategy were to strengthen teachers' assessment expertise, produce portfolio archives of the process and the products of assessment design and implementation, and support the emergence of professional communities committed to the improvement of classroom assessment.

Study Purpose and Analytic Approach

This report is an analysis of what the cohort of Academy teachers learned about classroom assessment from their work with their portfolios in the institutes and in their classrooms. The findings are organized in sections aligned with the portfolio: Learning about assessment tools (Section I and III), and learning about interpreting and using evidence (Section II). In each findings section, we coordinate two strands of evidence—portfolios and teacher self-report.

Portfolio analysis focused on changes over time in the portfolios of teachers who completed at least two portfolios, and ten teachers (representing Grades 1–9) met this criterion. Analyses of teachers' self-reported learning in surveys and focus groups serve as triangulation of our analysis of the portfolio evidence as well as enriched information about teacher knowledge and application of specific portfolio concepts and methods.

In concert, the portfolios, surveys, and focus groups enabled us to construct a profile of what Academy teachers learned from constructing a series of assessment portfolios for different curriculum units. We conclude the report with reflections on the findings and particular attention to the opportunities and limitations of a generic assessment portfolio for teacher learning.

Method

Participants

Nineteen experienced science teachers from Grades 1–10 participated in the Academy. Based on responses to our initial survey (N = 19), the cohort's mean years of teaching experience was 14.7 (SD = 12.68). The majority had completed coursework beyond their B.A., and half had earned their M.A. Most teachers had participated in professional development programs, and more than half had attended or presented at meetings of the National Science Teachers Association. Teachers generally perceived themselves as

instructional experts. On a scale of 1 (*weak*) to 5 (*very strong*), teachers rated themselves as strong in: confidence in teaching science ($M = 4.58$, $SD = 0.88$), knowledge or understanding of grade level science ($M = 4.41$, $S = 0.83$), and knowledge or understanding of grade-level science standards ($M = 4.46$, $SD = .66$). Even the ratings for "knowledge of a wide variety of assessment strategies and techniques" were fairly high ($M = 4.19$, $SD = .73$).

Data

Portfolios contained evidence of changes in teachers' understandings and practices, while surveys and focus groups provided teachers' perceptions of their learning and the factors that contributed supports and barriers to their learning.

Portfolios. We examined growth over time for those teachers who turned in a series of two or three portfolios each judged either 'Complete' or 'Partially Complete.'[7] To identify these portfolios, two researchers rated the portfolios for completeness, and rare disagreements were resolved through discussion. A *Complete* portfolio contained material for each of the three major sections: I. Learning Goals and Assessment Plan, II. Interpretation of Student Responses and Use of Evidence, and III. Assessment Revision. A *Partially Complete* portfolio contained material for II, the section that a teacher completed independently, as well as material for *either* I or III, the sections completed by collaborative teams. We identified 10 Academy teachers who submitted a series of two or three complete or partially complete portfolios. Portfolios spanned elementary through high school: elementary (Grades 1, 2, 3, 4), middle school (Grades 6, 8), and high school (Grade 9). We consider changes over time in the portfolios of these 10 teachers to be a reasonable estimate of growth in the cohort's understandings and uses of assessment, for two reasons: First, descriptives for the teachers in the portfolio sample were similar to descriptives for the remaining teachers in the cohort. As shown in Table 2, the two groups were similar in distribution of gender and ethnicity; while the non-portfolio group included a greater proportion of teachers with Master's degrees and greater teaching experience, there was a substantial range of education and experience in both groups. Second, the portfolios from the 10 teachers in our sample contained collaborative work representing the contributions of teachers who were not included in our sample.

[7] Criteria for completeness were generous given the challenges facing overworked teachers.

Table 2

Descriptives for teachers in the portfolio sample and remaining teachers in the cohort.

Items	Portfolio sample	Other teachers
N	10	9
Gender (distribution)		
Male	1	3
Female	9	6
Ethnicity (distribution)		
White	6	7
Non-white	4	2
Education (distribution)		
Bachelor's	5	2
Bachelor's + units beyond	0	1
Master's	5	6
Master's + units beyond	0	3
Teaching experience (*Mean, SD*)		
Mean	13.5	18.5
SD	8.2	7.1

Note. 1 = (*not at all*), 3 = (*moderate extent*), 5 = (*great extent*).

We used qualitative methods of analysis to examine growth over time in the quality of the assessment practices documented in each teacher's series of portfolios. For each series, one researcher documented patterns of program impact in a detailed matrix; additional researchers reviewed the same portfolio series, and the matrix was then revised based on intensive discussion of the patterns observed. We then identified patterns of change (or stasis) that were evident in the portfolio series of at least 5 of the 10 teachers. (The modest criterion of 5 out of 10 teachers was a reasonable decision given challenges that teachers often faced applying what they had learned from one portfolio to the next, because curriculum units varied strikingly in content, pedagogy, and quality of the embedded assessments.)

Surveys. We administered two survey instruments. One instrument focused on *classroom assessment practices* to provide baseline information in August 2003 as well as evidence of interim program impact in May 2004 after 9 months of participation. Teachers rated the extent to which they implemented various assessment practices on a scale from 1 (*very limited extent*) to 5 (*great extent*); 19 (of 23) teachers from Grades 1–9 completed the survey on both occasions. The second instrument, an exit survey, focused on the *understanding of Academy assessment strategies*, and it was administered at the conclusion

of the Academy in December 2004. Teachers rated their understanding of Academy strategies from 1 (*none*) to 5 (*full*), and 21 (of 23) teachers responded. The survey items from both instruments are included in the relevant Findings tables.

Both surveys included open-ended items. For the survey on assessment practices administered twice, we analyzed the May 2004 comments to help us clarify trends from August 2003 to May 2004. However, for the exit survey administered in December 2004, written comments were combined with the transcripts from the exit focus group before analyzing exit themes, because these data collection activities were conducted on the same day.

Focus groups. Exit focus groups were conducted in December 2004 on the same day that teachers completed the exit surveys. Members of grade-level and district teams were distributed across five groups to encourage fresh perspectives and clear communication. To structure discussion, participants were shown figures of the program framework (Figures 1 and 2) and asked to circle assessment practices or concepts that they had strengthened as well as those they needed to strengthen, and after each set of choices, teachers explained their selections. Teachers were then invited to identify strengths and weaknesses in the Academy portfolio and the Academy program; and recommend revisions in program goals, the portfolio, and strategies for supporting teacher learning. The groups were lively, and comments were extensive. Recordings were transcribed without identifying names, and the five transcripts (combined with comments from the exit surveys) were used as collective evidence of cohort exit views. Two researchers used descriptive codes to locate topically similar talk and then code each topic to capture themes, and disagreements were resolved through discussion. Themes were further refined based on feedback from other members of the research team and the professional development team.

Findings

Patterns of cohort learning are reported in two parts that are aligned with the sections of the portfolio. We focus first on *learning about assessment tools* and present findings on teachers' progress with planning coherent assessment systems and designing appropriate assessments. We then report evidence of teachers' progress with *interpreting student responses and using evidence to guide instruction.* We begin each section with an overview of the portfolio tasks that provided teachers opportunities to learn and a summary of evidence of teacher learning in the portfolios. Finally, we validate and contextualize the portfolio findings with teachers' self-reported learning on surveys and in focus groups.

Learning about Assessment Tools: Planning a Coherent Assessment System and Refining Specific Assessments

The Academy portfolio provided teachers multiple opportunities to learn how to plan a coherent assessment system and refine specific assessment tools. In Section I, teachers identified and organized their learning goals for their units in a Conceptual Flow, and then used the RAIM to select and refine a series of coordinated assessments to monitor student progress toward the learning goals. As teachers worked to identify the 'expected student responses' to tasks, that process almost always prompted teachers to strengthen the quality of their selected assessment tools. During implementation in Section II (where the focus was on interpretation and use of evidence), teachers refined their assessment tools when they constructed scoring criteria to capture the range of performance in their students' responses. After completing their units, teachers critiqued and revised both tasks and criteria in Section III of the portfolio. As we report next, our analyses of the portfolios and teachers' self reports show that teachers were intensively engaged in developing, evaluating, and revising unit learning goals and assessments to create a coherent assessment system for their curriculum units. However, teachers made more progress learning to establish coordinated learning goals than they did with the selection, development, or refinement of assessment tools.

Planning a coherent assessment system: Establishing goals and selecting assessments. The evidence of planning in the portfolios included the conceptual flow and the RAIM forms. Given the wide range of grade levels and units in our portfolio sample, it was not possible for us to evaluate the quality or clarity of each learning goal in the conceptual flows, nor the capacity of each assessment to measure students' progress toward a given goal. However, we could make observations about shifts over time in the *organization* of learning goals and assessment plans.

One shift in the conceptual flows was toward a *greater focus on big ideas* by removing, adding, or reorganizing unit goals to focus on what was most important for students to learn. For example, a first grade team discarded an item evaluating children's understanding that water in a glass is always parallel to the horizon regardless of how the container is turned, because the task was irrelevant to the learning goals in a unit on solids and liquids. A middle-school team added the concept of density to a unit on plate tectonics, because they knew that an understanding of how matter in the earth's crust shifts is based on student understanding of the concept of density. Another middle school team reorganized their unit on heredity by introducing 'pre-learning' opportunities for students to learn scientific terminology after noticing that their English Language Learner (ELL) students could often correctly identify inherited characteristics, but their descriptions lacked specificity and clarity. These

organizational shifts in conceptual flows toward a clearer focus on big ideas were more evident in the third portfolios when teachers revised an assessment plan they had previously constructed for an earlier portfolio.

Another shift in the Conceptual Flows was toward *more coordinated relationships among big ideas and smaller supporting concepts*. Teams increasingly represented relationships as a concept map rather than listing lesson topics sequentially or linearly. For example, in their first flow for a unit on homeostasis, a high school team depicted regulatory systems as distinct systems in the body; in the team's third conceptual flow for a repeated unit, they highlighted the interconnected relationships between homeostasis and regulatory mechanisms in the body to emphasize the importance of the concepts.

Consistent with increasing clarity and coordination of learning goals, the assessment plans in teachers' later portfolios were more coherently organized. Plans shifted from long lists of possible assessments *toward judicious selection of a few key assessments* for tracking student progress—a pre-assessment, one or more "juncture" assessments, and a post-assessment.[8] The addition of a pretest in many of the assessment plans in the later portfolios was a particularly noteworthy innovation in unit assessment design, since very few units contained pretests. Teachers also worked on strengthening alignment among the pre-, juncture, and post-assessments in order to *track student progress* with a key concept or process. For example, when a middle-school team discovered that students were challenged by the graphing requirements of their unit on density, they added graphing items to each of their assessments to allow them to analyze how student understanding of graphing was developing in addition to students' understandings of density. Some teams repositioned assessments that were targeted and easily analyzed for use as formative juncture assessments, and, likewise, moved their more comprehensive assessments for use as summative tools. Finally, teachers' later portfolios depicted *relationships between learning goals and assessments* in one document rather than separate flows and RAIMs. However, as we will report in the next section on assessment tools, these improvements in the coordination and coherence of unit assessment plans were not necessarily balanced with improvements in the ways teams revised *specific* assessments during the planning phase.

In surveys and focus groups, teachers' self-reports of what they were learning about assessment planning were consistent with the evidence in the portfolios. As shown in Table 3

[8] In the first portfolio, teachers were asked to list all possible assessments before making selections for their assessment plan; in later portfolios, that task was revised to focus teachers more directly on selection of targeted assessments. Thus this pattern of change from comprehensive lists to targeted selection mirrors revision of the portfolio tasks—but that revision was prompted by teachers' requests for a more strategic approach to assessment planning guided by the conceptual flow of learning goals.

for repeated survey items on assessment systems, after the first 9 months of the Academy, teachers reported more extensive efforts to set learning goals, align assessments with goals, and include assessments of prior knowledge, although these trends were not significant. Teachers' survey comments in May 2004 reflected the trends in Table 3. Teachers praised the benefits of a portfolio process that engaged them in planning, implementation, reflection, and revision. For example, one teacher viewed the process as "a mind-opener!"; another teacher commented, "Before I would have believed that I was very good at evaluating the alignment of assessments with assessment targets; it wasn't until I saw my results from the pretest (or lack of results) that I realized I wasn't as good at this as I originally thought."

Table 3

Planning goals and assessments: Means and standard deviations for survey administered August 2003 and May 2004 ($N = 19$).

	Mean *SD*	
Items	August 2003	May 2004
To what extent do you:		
- Set specific goals for student progress?	3.84	4.16
	.69	.77
- Align your assessments with your learning goals?	4.00	4.37
	.75	.76
- Assess students' prior knowledge?	3.84	4.16
	.83	.83

Note. 1 = (*not at all*), 3 = (*moderate extent*), 5 = (*great extent*).

Table 4

Planning goals and assessments: Means and standard deviations for exit survey administered December 2004 ($N = 19$).

Items	*Mean* *SD*
To what extent do you feel you understand the following Academy strategies?	
- Creating conceptual flows.	4.67
	.66
- Using conceptual flows to guide assessment decisions.	4.57
	.60
- Selecting critical junctures as important concepts to be assessed.	4.29
	.72
- Preparing the RAIM plan.	3.79
	.96
- Using the RAIM plan to guide assessment decisions.	3.95
	.86

Note. Scale: 1 = (*poor understanding*), 3 = (*moderate understanding*), 5 = (*excellent understanding*). RAIM = Record of Assessments in Instructional Materials.

Teachers' responses to exit survey items on assessment planning (Table 4) indicated generally strong understandings of the key steps in the Academy portfolio planning process —how to create conceptual flows of learning goals, use the conceptual flow to guide assessment decisions, and select the juncture assessments. The mean ratings for use of the portfolio RAIM form to guide specific assessment decisions and assemble the assessment plan were lower. However, a trend that is consistent with findings we report in the next section is that teachers felt they had gaps in their understanding of the process of developing assessment tools.

In their comments on the exit survey and in exit focus groups, teachers described what they had learned about assessment planning from the portfolio, and their needs for further support. One theme was a recognition of the importance of well specified and sequentially coordinated goals: "I now focus on what students need to know in conjunction with the conceptual flow and not just what I need to cover in the unit." Another theme was a shift from a focus on summative assessment toward the integration of formative assessments: "Before … I was doing backwards design … making my summative assessment ahead of time, but I wasn't planning the formative assessment ahead of time; in making the RAIM, I've already got all the formative assessments identified." But as we discuss next, many

teachers felt they needed better understanding of techniques for strengthening specific assessments.

Learning how to refine specific assessments. We analyzed the evidence of teachers' efforts to improve the quality of their assessment tools from all three sections of the portfolios—revising selected assessments for assessment plans in Section I, developing criteria by revising the Expected Student Responses (ESR) in Section II, and revising assessments after completing the unit in Section III. Comparisons of these sources of evidence in successive portfolios showed mixed patterns of improvement. Note that, while we can describe the ways that teachers revised assessments, we cannot evaluate whether revisions strengthened the quality of the assessments, because teachers had limited opportunity to re-implement their assessments.

All teams made minor revisions to improve clarity of task expectations—modifying the size of figures, leaving more space for students to answer, refining directions and response choices for clarity, and so on. A more substantive endeavor was to revise to strengthen the alignment of assessment tasks with learning goals. For example, one elementary team revised instructions for a performance item on pitch and volume, because students were interpreting the investigation instructions incorrectly, and therefore their conclusions about sound were not relevant to the targeted concepts. AR, an elementary teacher on another Academy team, revised the mineral samples for an assessment of the characteristics of rocks and minerals, when she discovered the students' "kit misconception" that "all minerals are white" because all minerals in the instructional kit were white! YJ, a middle-school teacher, replaced an open-ended essay task assessing students' understanding of plate tectonics with a set of short-answer items that provided more targeted evidence about student understanding. RA, a high school teacher, added an explanation question to his multiple-choice test on the periodic table to provide additional information on student understanding of how the periodic table is organized and how it is useful for predicting the nature of elements. These revisions provide evidence of teachers' deepening understanding of the importance of aligning assessment tasks with learning goals.

Teachers also strengthened the quality of assessment criteria, revising them to differentiate more levels of understanding and dimensions of performance. A few teachers endeavored to capture levels of understanding in ways that could provide students targeted feedback and guide instructional improvement. For example, in CM's third portfolio, she transformed the publisher's three-level holistic scoring guidelines ("complete response," "partially complete response," and "no responses or a response that doesn't make sense") into a four-level scoring guide containing two distinct conceptual dimensions: (a) can

accurately identify and use tests to distinguish different types of minerals, and (b) minerals are the basic elements that make up rocks and have properties that can be described. To guide her feedback to students, her scoring guide specified what additional information and concepts students needed to learn to move to the next level of understanding.

But in some of the other portfolio series, criteria revisions were more in form than in function. An eighth-grade team, for example, added a third level to their two-level holistic rubric for a performance item on the properties of matter, and removed some extraneous criteria to strengthen alignment of the levels. Despite these improvements, the lowest level response in the properties of matter rubric still focused on what was incorrect or incomplete ("inaccurate observations, no or wrong data, 1 weak observation, not specific, unrelated, no idea what student means") rather than students' alternative conceptions. The new middle level was a mix of correct and incorrect features not fully aligned with low or high levels ("accurate, no data, 1 or 2 valid but incomplete observations, not all items included, sloppy but legible"). So while all teachers made steps toward strengthening criteria used to evaluate student work, progress in some of the portfolio series was more limited.

In surveys and focus groups, teachers expressed new insights about assessment tools along with continued uncertainties about how to revise or develop quality tools. After the first 9 months and completion of two portfolios, as shown in Table 5, teachers reported a decline in the extent to which their assessments were of high quality; decreases in ratings for two items (*valid for reason you are using them* and *designed to accommodate learner*s) were statistically significant ($p < .05$), and the trend of decline was the same for items on reliability and fairness. This pattern suggests that what teachers were learning about quality assessment tasks and criteria was making them more critical of the tools they were using, and teachers' survey comments supported our interpretation of the trends.

Table 5

Assessment tools: Means and standard deviations for survey administered August 2003 and May 2004 (N =19).

Items	*Mean* *SD*	
	August 2003	May 2004
To what extent are your tools:		
- Based on strong science content?	3.89	3.95
	.57	1.03
- Valid for the reason you are using them (measures what you thought it would)?	3.72	3.11*
	.58	1.15
- Reliable and accurate?	3.68	3.26
	.58	1.20
- Designed to accommodate learners with various needs?	3.68	2.79*
	.75	.98
- Fair?	3.89	3.42
	.68	1.07

Note. 1 = (*not at all*), 3 = (*moderate extent*), 5 = (*great extent*).
*$p < .05$

On the one hand, teachers reported that they were learning about assessment tools. For example, teachers described new ways that they were using assessments to provide formative information: "I have never given a pretest before … it helped me to see that my students are learning"; "assessing students' prior knowledge has become a more formal process;" "[Now I am] assessing what students know at critical junctures"; "I have been using more pre/post testing than ever before." Teachers also reported new insights about ways to strengthen the quality of assessment tools: "Evaluating the developmental appropriateness of an assessment is something I have tried to learn more about"; "I have made an effort to be sure that the assessments I give measure what I intend them to;" "[from] going through the [portfolio] process, I have gained insight into how to adjust my assessment tools to better understand the students' learning."

On the other hand, teachers reported concerns about the quality of their assessment tools, and these comments mirrored the decrease in ratings of assessment quality in Table 5. One theme was the discovery of *weak alignment of their assessments with learning goals* in the conceptual flows: "[I've seen that] developer-created assessments must be checked and analyzed to determine if their questions are assessing the same objectives you are looking for"; "[the portfolio] made me aware that—even in reform units with embedded

assessments—the assessments did not always assess the concepts we wanted to asses;" "[publisher's] tools of assessment are not always aligned with those strong science concepts." A related discovery was the *weak alignment between assessment and instruction*. Teachers commented, for example, that, "my pre and post tests didn't match the instruction, so next time I will revise them and revise the instruction," and "the juncture assessment asked questions about content the students haven't learned yet, and we need to revise it." Additional concerns were raised about the quality of *item types* ("my pre/post test was a multiple-choice test that told me nothing") and the *fairness* of items ("I am more aware of student accessibility to the question—the language, the vocabulary …").

Given these concerns, teachers requested opportunities to learn more about tool refinement: alignment ("reviewing assessment questions to target what we really want to know, because curriculum relevance and tight alignment are the foundation of effective assessment practice"); item types ("performance based assessments that go with the conceptual flow," "multiple-choice items that accurately demonstrate conceptual understanding," and "interview strategies"); ways to "assess growth over time" through pre-post tests "that will really assess student learning;" and the "process of validating and checking for reliability across assessments." Thus, as teachers worked on specific assessments for their curriculum units, they discovered they needed additional, targeted assessment expertise linked to the assessments in their instructional materials. These needs could not be readily supported by a generic assessment portfolio.

Table 6

Assessment tools: Means and standard deviations for exit survey administered December 2004 ($N = 19$).

Items	*Mean* *SD*
To what extent do you feel you understand the following Academy strategies?	
- Clarifying the concepts assessed for each assessment.	4.10 .83
- Clarifying the Expected Student Responses (ESRs) for each assessment.	4.14 .66
- Developing or revising a pre-assessment.	4.35 .74
- Developing or revising an assessment for the first critical juncture.	4.19 .68
- Developing or revising a post-assessment.	4.35 .67

Note. Scale: 1 = (*poor understanding*), 3 = (*moderate understanding*), 5 = (*excellent understanding*).

Exit findings were similar to mid-program results. As shown in Table 6, teachers expressed moderate to high confidence in their understandings of Academy portfolio strategies, with slightly lower ratings for the detailed work of clarifying concepts and expected student responses as well as the challenging task of developing a juncture assessment. (Developing a juncture assessment requires teachers to specify how students' scientific ideas are developing within the unit, and what tool will capture student progress along the developmental trajectory.) Teachers' lower ratings for understanding the technical aspects of assessment refinement—a pattern we previously identified in May 2004—was supported in teachers' exit comments on the survey and in focus groups.

On the one hand, Academy teachers reported learning the big idea that assessments do more than identify "what [students] got right or wrong"—quality tools provide information on "what students think" and "what the students learned … specifically what the students don't understand." Teachers now understand that tasks or criteria may "need to be tweaked and fixed," and therefore teachers appreciated the portfolio "RAIM process that helps me look for quality assessments" and determine "what does this assessment tell me about what students know or understand." A salient idea for Academy teachers was that evaluating assessment tasks requires examining evidence in student responses: "[What the Academy portfolio] added was refining tasks based on evidence, not on how I feel" and "[I've realized from the portfolio that when] the question isn't really written correctly … you're not going to

get expected answers that you need from the students." A parallel insight was that developing assessment criteria begins with brainstorming expected student responses (ESRs). "[Before the Academy portfolio,] I knew what the right answer should be but I didn't think about, you know, the criteria and how to map it out" and then revising these ESRs with student work in hand: "We really need to look at student work to refine the criteria so the criteria represent an accurate assessment of student learning." Furthermore, criteria may need still further revision to ensure reliable scoring: "[I'm asking myself,] do I really understand what the rubric says or do I need to refine it more to make it more clear so that my interpretation isn't different from the teacher next door?"

On the other hand, teachers expressed needs for further learning that we heard first in May 2004—specific information or resources that the assessment portfolio as a generic tool was not designed to provide. For assessment tasks, teachers asked for help with: item types ("multiple choice questions that accurately demonstrate conceptual understanding"), comparable measures of progress ("pretests and posttests that really assess student learning"), targeted formative tools ("quality juncture assessments"), tools for particular science concepts ("tools for assessing inquiry-based science work"), and tools that minimize bias ("help me to know how to be fair and unbiased in our question—how to word it, how to present it"). For assessment criteria, teachers requested assistance with developmentally appropriate criteria: "I need more work on ESRs … I've been off-base a lot;" "[My criteria need to] go a little deeper than 'my high achieving vs. my low achieving';" "[I wish the Academy] had spent more time on … 'why are they making the incorrect answers—is it because of a misconception? is it because of language?'" Teachers' desires to learn more about assessment development were balanced with desires for already-developed resources—higher quality assessments and criteria embedded in their instructional materials. As one teacher explained, "It's hard to write good assessments—field testing shows unexpected results; it's an iterative cycle, and it's time intensive—if assessment writers were more careful, our jobs would be easier." Another teacher commented frankly, "[If I had] set criteria [in the materials], it would make my life easier."

In sum, after completing two or three portfolios, Academy teachers exited the program with a commitment to formative assessment. They reported progress in their approaches to assessment planning and in their knowledge and understanding of quality tools, while simultaneously identifying their needs for further opportunity to learn how to select and refine tools and for higher quality tools. Many of their needs required resources beyond those provided by the Academy portfolio.

Learning to Interpret Student Responses and Use Evidence to Guide Instruction

Section II of the portfolio provided teachers the opportunity to learn how to interpret student work and use the evidence to improve instruction and provide students feedback. Because teachers completed Section II independently in their classrooms, the portfolio forms were designed to replace the facilitator and scaffold teachers step-by-step through each portfolio task. After a teacher implemented a given assessment, the initial steps were to sort student work into three piles to provide initial information on levels of performance, compare patterns in the piles of student work with the "expected student responses" drafted in Section I, and construct scoring criteria (usually a rubric) through an iterative process of refinement. The portfolio contained models of holistic and analytic rubrics that varied in the number of performance levels to guide teachers as they developed their own scoring guides. Scoring came next, and the portfolio encouraged teachers to record their scores in an "assessment record," a matrix that could include additional qualitative comments on students' responses if teachers felt those would be useful. The portfolio forms then suggested ways to analyze patterns and trends in the assessment record. Teachers were provided space to describe and record their approach to criteria development, their methods of analysis, their findings, and the ways they used the evidence to provide students feedback and guide instruction.

Below we report findings on patterns of teacher learning, first for interpretation of student work and then for use of evidence to guide instruction and student feedback. Our analyses of the portfolios and teachers' self-reports show that teachers were learning new strategies for interpretation of student work, and they were developing more targeted strategies for instructional improvement and feedback. However, teachers made progress in different ways and to different extents in their efforts to strengthen their interpretations of student learning and learn to use assessment data to guide instruction.

Learning to interpret student work. In the first portfolios, we found that most student work was either graded or simply collected, and teachers' inferences about student learning were based on unsystematic reviews of the student papers or on other sources, such as class discussion and informal observation. In the second portfolios, many teachers scored with rubrics,[9] though only some teachers charted scores and analyzed patterns and trends. When asked to report patterns of student learning, many teachers essentially restated the criteria content. For example, one teacher wrote, "Some students correctly predicted how changing the angle of the plane would impact the speed of descent of the water drop" which was a

[9] Unfortunately we could not trace teachers' growth with scoring techniques such as benchmarking or double-scoring, because the portfolio did not ask teachers to document the scoring process.

restatement of her criterion for a "high" score. By the second or third portfolio, all 10 teachers in our sample had used or adapted Academy models to score student work, chart results, and analyze whole class patterns and trends. Some of their assessment records were quantitative (scores), some qualitative (content analyses of responses), and some a hybrid (when teachers supplemented their scores with qualitative notes on the content of student responses).[10] There was variation in the quality and sophistication of teachers' methods of interpretation.

Some portfolio series showed considerable progress in methods of interpretation in the second or third portfolios. For example, JR, a middle school teacher, developed an innovative use of the Academy "hybrid" record when she focused her qualitative notes on the low and medium responses to help her identify needs for further instruction. CM developed a four-level scoring rubric to determine what students understood about specific concepts, and she used several approaches to analyze class patterns: pre-post test comparisons based on the number correct and change in score; item or concept correlation by clustering items related to each concept; and identification of concepts associated with the most frequently missed items.

Some Academy teachers' progress with analysis and interpretation of student ideas was less substantial. For example, some of the whole class analyses in the later portfolios were limited to recording class distributions of total scores or whole class averages. Other analyses were inefficient; for example, one teacher listed the items that each student answered correctly (e.g., "Jenn: 2, 4, 7, 8, 9; Santiago: 1, 2, 4, 7, 8, 9"), but this format was not well suited for interpretation of patterns of student understanding or item by student interactions. Some methods teachers employed to interpret student progress were problematic—for example, when teachers compared class means (or distributions) on identical assessments with no further analysis or compared students' performance on assessments that were not comparable; for example, JR analyzed shifts in students' L-M-H levels of performance over a series of assessments even though the tasks and criteria differed for each assessment.

Overall, the portfolios revealed that teachers were intensively engaged with the analysis and interpretation of student work, and many portfolios contained evidence of shifts in awareness and levels of sophistication of various analytic techniques. However, teachers'

[10] We cannot determine whether the hybrid records reflected limitations in teachers' capacities to construct scoring guides or their growing insight that mixed methods can be efficient and targeted. Shepard (2001), for example, argues that qualitative analysis of the responses scored at medium and lower levels is a flexible and feasible strategy for classroom assessment.

methods of analysis in their final portfolios were uneven in quality, raising questions about teachers' need for additional opportunities to learn targeted to specific curriculum units.

Table 7

Interpretation of student work: Means and standard deviations for survey administered August 2003 and May 2004 ($N = 19$).

Items	*Mean* *SD*	
	August 2003	May 2004
To what extent are your tools:		
- Are you using your assessments to make sound interpretations?	3.68	4.16[†]
	.75	.83
- Do you analyze individual work and responses for specific student understandings?	4.22	4.32
	1.06	.67
- Do you evaluate students' ideas based on a developmental framework of science understanding?	3.84	3.74
	.76	1.24

Note. 1 = (*not at all*), 3 = (*moderate extent*), 5 = (*great extent*).
[†]$p < .07$

Survey and focus group findings were consistent with the portfolio patterns. As shown in Table 7, after the first 9 months, teachers' ratings for their use of "sound interpretations" increased. The higher and stable ratings for "analyze individual student understandings" in contrast to the lower and stable ratings for "according to a developmental framework" suggest that teachers viewed themselves as consistently engaged with student work but not necessarily equipped to interpret conceptual development, and teachers' survey comments support our inference. On the one hand, teachers reported that, through their portfolio work, they had shifted from just grading student work toward "analyzing individual student work" and "analyzing test results from the perspective of student understandings." They also reported learning techniques to assist with interpretation of student responses; teachers' comments such as "use of the matrix-grid where I lay out each student's response for every question, improving my ability to see the weaknesses in student understanding" and "how to quantify and qualify my decisions based on student responses" illustrate teachers' progress in learning these techniques. But teachers also expressed a "need for more help on evaluating students' ideas based on a developmental framework of science understandings;" "(I need help) creating rubrics that go beyond 'got it,' 'mostly got it,' 'most didn't get it,' 'moron,' etc.—give me examples of rubrics and what they are showing!" They also wanted to learn

more about ways to produce information that could inform instruction in explicit ways: "making matrices designed to pinpoint student weaknesses to inform instruction," "identifying trends and misconceptions," "qualitative instead of quantitative" information.

Table 8

Interpretation of student work: Means and standard deviations for exit survey administered December 2004 ($N = 19$)

Items	*Mean* *SD*
To what extent do you feel you understand the following Academy strategies?	
- Analyzing whole class sets of student work.	4.29 .56
- Comparing student performance on pre- and post-assessments.	4.48 .81
- Comparing student performance on pre- and juncture-assessments.	4.52 .60
- Comparing student performance on juncture- and post-assessments.	4.52 .60

Note. Scale: 1 = (*poor understanding*), 3 = (*moderate understanding*), 5 = (*excellent understanding*).

On the exit survey (Table 8), teachers reported moderate to full understanding of Academy portfolio strategies for interpretation, and their comments on the survey and in focus groups captured their new commitment to careful interpretation of student work and their progress with some specific techniques. Teachers' new investment in understanding student thinking was clear: "I really care about what each student is saying, about what each group is thinking about an idea"; "I'm now looking at everything, and it's not that A, B, C, D grade—I'm not even putting grades on anymore;" "I look more carefully at what it is that [students] don't know, and not so much, 'oh they got it, they didn't get it.'"

Teachers reported great appreciation for Academy portfolio methods of recording scores and analyzing patterns: "Having to make a chart of that and analyze that, that really really helped me—I wouldn't have ever thought of that;" "One of the things that I think I came away with was really analyzing student work … and seeing what trends were in the class—you know, if a whole bunch of kids missed that, but they answered it this way, you know, you look at the breakdown;" "Before it would be looking at the individual student and giving them a grade and not really looking at the trends across my class and … individual concepts they might be lacking …. It's great." But Academy teachers also expressed needs

for further learning about aspects of assessment that were included in the Academy assessment framework but unsupported by the portfolio provided. Teachers were concerned, for example, about the concept of fairness: "I wonder to myself, 'Am I being fair to the test or fair to the rubric?' or 'am I being fair to the student?';" "Am I really interpreting this correctly, the way it should be?;" "I have a variety of students from English language learners to Gifted and Talented Education (GATE) students, and I'm uncertain how to be fair and unbiased in my questions." Reliability was another concern. For many teachers, "reliability" remained a highly technical notion; as one teacher expressed it, "I don't have that academic or research background to do that."

The final theme grew out of the experience of interpreting student work independently without the support of team colleagues and a facilitator. Some teachers felt they had learned less than they might have if they had worked collaboratively: "I don't think analyzing student work comes with an individual teacher sitting alone at night;" "I want the opportunity to talk to people, to work on our units together, and [access] to a facilitator to help me with [analyzing student work]." These comments highlight the benefits of collaboration in the context of the realities of project funding; the Academy did not have the resources to bring teachers together across the state during unit implementation.

Learning to use evidence to guide instructional improvement and provide students feedback. Although our evidence of what Academy teachers were learning about using assessment to improve student learning was limited to their responses to two open-ended portfolio prompts, there was a clear pattern of progress. In early portfolios, teachers described generic strategies for follow-up such as giving students the correct answers, reteaching, reviewing vocabulary, or modeling test-taking skills. In later portfolios, teachers began to report lesson-specific follow-up activities that merged instruction with feedback, and challenged students' understandings of core concepts. For example, teachers engaged students in scoring and revising their work, conducting revised investigations, or discussing diverse responses to the assessment in small groups. Some teachers implemented instructional strategies matched to students' needs—for example, more didactic instruction for students with the least understanding, and, for other students, opportunities to discuss their responses in small groups. A few teachers began to plan in advance for the range of student understanding. CM, for example, developed a scoring guide that integrated scoring with her strategies for follow-up and feedback. However, even in later portfolios, some teachers continued to report some limited uses of evidence—for example, using juncture information simply for reteaching or correcting errors, or using pre-assessment results solely

as a baseline measure for pre-post comparisons, neglecting their value for instructional planning.

Table 9

Use of information to guide instruction: Means and standard deviations for survey administered August 2003 and May 2004 ($N = 19$)

Items	*Mean* *SD* August 2003	May 2004
To what extent are your tools:		
- Are you using your assessments to guide instructional improvement?	4.21	4.58*
	.63	.51
- Are you using your assessments to provide communication and feedback to students regarding their performance?	3.84	3.44
	.83	1.19

Note. Scale: 1 = (*not at all*), 3 = (*moderate extent*), 5 = (*great extent*).
*$p \leq .05$

Teachers' self-reported learning supported the portfolio pattern of increasing use of assessment information to guide instruction. After 9 months of Academy participation and two portfolios (Table 9), teachers reported more frequent 'use of assessments to guide instruction,' and, in their survey comments, teachers described using both generic approaches to follow-up ("give students more practice," "reteach the concept," "give students a chance to redo the assessment after a review to clarify and better their understanding," "ask students to reflect on their learning") as well as techniques that can target students' understandings of specific concepts ("more differentiated instruction" and "scaffold instruction based on student needs"). However, some teachers expressed uncertainties about uses of assessment information that suggested they were experiencing some specific dilemmas in the context of their units: "I've made changes in instruction based on student evidence but I am not sure they are the best changes—I at least try;" "now that I know where they are, what do I do?"; "[how do I] make changes in instruction based on assessment results." Table 9 also shows a trend of decreased feedback to students about their performance on assessments, but we hesitate to give an interpretation for the decrease, because there were only two comments regarding difficulties with feedback: "how [should I] give students feedback;" "[I'd like to] hear more about providing effective feedback to students."

Table 10

Use of information to guide instruction: Means and standard deviations for exit survey administered December 2004 ($N = 19$).

Items	*M* *SD*
To what extent do you feel you understand the following Academy strategies?	
- Using the pre-assessment as a key source of information to guide my instruction.	4.38 .80
- Revising instruction based on whole class assessment information.[a]	4.50 .51
- Revising instructional materials based on assessment information.[a]	4.25 .79
- Using the junctures as key sources of information to guide my instruction.	4.40 .75
- Differentiating instruction based on assessment information.	3.86 .73
- Providing feedback based on whole class assessment information.	4.24 .70

Note. Scale: 1 = (*poor understanding*), 3 = (*moderate understanding*), 5 = (*excellent understanding*).
[a]$N = 20$ for these items.

On the exit survey (Table 10), teachers reported moderate to high understanding of uses of assessment information, with the exception of moderate ratings for 'differentiating instruction,' which may reflect teachers' awareness that differentiation requires accurate information on each student's understanding. In their survey and focus group comments, teachers recognized follow-up as a key component of formative assessment: "(now I) really care about what each student is saying, about what each group is thinking about an idea, and how to address that idea;" "I had a more fine tuned lens on what I was looking for, and I would observe my class more (and) there were things that I would kind of check really quickly so that I could move on to address those misunderstandings right away;" "I haven't thought about that before … (i. e.,)'Okay, I teach it, I assess it, we move on; those kids that didn't catch it, okay, I'll catch up with them later' (but) that's not how it works." Teachers identified sources of assessment information that they found particularly helpful for providing feedback to students. Some mentioned pretests: "I never used to do pretests, but now I look for patterns on the pretest to get patterns and change my instruction; I get a current idea of the unit based on student understanding rather than what the teacher thinks the unit should be." Other teachers mentioned the value of close analysis of whole class patterns

and trends on assessments: “Looking at trends, really analyzing your students’ work, and, based on that, ‘what do I do next?’ That, I thought, was really really important;” “The whole class analysis … is a true guide for instruction—it’s a guide for differentiation, it’s a guide for grouping, it’s a guide for a number of classroom issues.” Teachers also commented on their growing use of feedback: “Just last night I was grading the lab reports, and I found out all I’m doing is writing questions to all the kids!” “I have students work together evaluating each other’s assessments and give feedback to one another.”

However, some teachers’ views of instructional follow-up were still limited to reteaching (“analyzing helped me to see what I needed to re-teach”) or reporting assessment results (“Okay, these are the patterns I noticed in the tests”). Some teachers knew they had more to learn about how to use assessment results to guide instruction for specific units. For example, one teacher was concerned about targeted instructional strategies or specific learning needs: “What do I do [when] ten groups of kids are in different spots—[I need] differentiation methods for dealing with the results of the assessments.” Another teacher felt stuck when writing feedback: “I understand completely that it’s important to give descriptive vs. numeric feedback, but, when I sit down to write the feedback, I don’t know what to write—I think it’s because I don’t have clear in my head what the quality criteria are.” The dilemmas these teachers were confronting emerged when they were implementing specific assessments, and the generic Academy portfolio was not designed to provide unit-specific support.

In sum, through their work on Academy portfolios, teachers’ practices of merely collecting or grading student work were replaced with a range of strategies for analysis of student understanding. In concert, teachers’ reported uses of evidence to provide feedback and guide instruction shifted from generic statements about the need to reteach toward more targeted strategies for feedback and instructional improvement. These patterns of broad impact on teacher learning were balanced by teachers’ awareness of specific learning needs. When analyzing student work, teachers discovered a need to understand additional aspects of assessment, such as fairness of interpretation and reliability of scoring, and, when using evidence, teachers experienced uncertainties about specific ways to use information to support student learning.

Discussion

The Academy assessment portfolio was an innovative professional development tool designed to guide science teachers toward deeper understandings of classroom assessment and support implementation of more effective assessment practices. Through a series of three

portfolios, Academy teachers gained experience with a process for designing assessment plans for curriculum units, gathering and analyzing evidence of student understanding, and using the information for instructional improvement. Evidence in teachers' portfolios and in their self-reports in surveys and focus groups revealed that teachers learned a great deal from their portfolio work. Their growth in expertise, however, was uneven, with greater and more widespread impact on a big picture view of curriculum-integrated formative assessment, and lesser and uneven impact on some of the more technical aspects of assessment as well as curriculum-specific methods of assessing student understanding and using information to improve instruction. We close this report with a summary of patterns of teacher learning, and a reflection on factors that supported and limited teachers' growth.

Summary of Findings

One strand of our analysis focused on what the cohort of Academy teachers was learning about assessment planning. From experiences constructing 'conceptual flows' of unit learning goals and assessment plans, teachers came to understand the important roles of big ideas in a unit, coherent lesson sequences, and assessments aligned with learning goals and integrated with instruction. When we compared each teacher's portfolio over time, we found that later conceptual flows were clearer depictions of big ideas and supporting concepts; key assessment points were clearly identified on the conceptual flow representations; assessment systems were explicitly organized to track student progress through a sequence of pre-assessments, formative assessments, and post-assessments. Teachers were also learning a great deal about the quality of specific assessment tools. Through experiences evaluating and revising the quality of tools based on evidence in student responses, teachers discovered problematic alignment of assessments with learning goals or instruction, as well as weaknesses in the quality of tasks and criteria; teachers then revised tasks and criteria in efforts to tighten alignment, clarify response expectations, and capture the full range of student understanding. However, the generic portfolio strategy—like any professional development strategy—had its limitations, and teachers exited the Academy program recognizing a need for further learning, particularly about assessment concepts and strategies targeted to particular curriculum units. Aware that some of their assessments still needed further revision, teachers raised questions about the appropriate role for teachers in assessment design and expressed a desire for higher quality assessments embedded in their instructional materials.

Our second set of findings addressed teachers' growth with interpretation of student responses and use of the information to guide instruction and provide student feedback. Academy portfolios provided teachers repeated opportunities to learn to refine criteria, score

student responses, record and analyze whole class patterns and trends, and use the information to guide instruction and provide students feedback. Our comparisons of each teacher's portfolios over time showed that teachers gradually replaced their practices of merely collecting student work or grading papers with strategies for analysis of student understanding and instructional follow-up. Teachers refined criteria based on patterns in the student work, scored responses, organized scores in a variety of records, and analyzed whole class patterns and trends. In concert, teachers' uses of the evidence shifted from generic statements about the need to reteach toward more targeted lesson-specific strategies for feedback and instructional improvement. But many teachers exited the program with methods of interpreting student work that missed informative patterns or were, in some cases, problematic, such as analysis of progress based on a sequence of assessments that were not comparable. The gaps in teachers' exiting expertise again reflected inevitable limitations of the generic portfolio. Teachers requested further opportunities to learn how to develop criteria that capture a developmental range for a particular concept, ways of gauging student progress in a particular unit, and ways of strengthening the reliability of their scoring for particular scoring guides. They also asked for guidance with more targeted and effective ways of differentiating instruction and providing feedback for specific curriculum units.

Through their portfolio work, Academy teachers adopted a new professional stance toward instructional materials; they recognized that materials are not inflexible scripts to be followed verbatim but revisable resources for teaching, learning, and assessment. They practiced methods for designing and implementing assessments, and they deepened their appreciation for systematic and ongoing formative assessment integrated with instruction. Few teachers, however, felt fully competent with all components of the Academy's vision for quality classroom assessment.

Interpretations and Reflections

In this section, we consider the contributions and limitations of the Academy portfolio strategy. We discuss the role of the portfolio in the growth of assessment expertise within the Academy community as well as dilemmas raised by patterns of uneven growth in teacher learning. We then consider directions for further research, focusing on the need to identify strategies that balance teachers' opportunities to learn broad assessment principles with targeted support for assessments for particular curriculum units.

Supporting the growth of professional community around assessment. The Academy was a statewide effort to build district and state capacity; five district K–12 teams convened three times a year to work as cross-district grade-level teams on unit portfolios. In

this context, a notable strength of the Academy's generic assessment portfolio was its role in establishing a flexible and sustained professional community committed to the improvement of assessment. All teachers were introduced to the fundamental assessment principles represented in the Academy framework, and then grade level teams within the larger community grappled to apply those principles to assessments for their curriculum units. Through both whole-group sessions and team portfolio work, district K–12 teams developed shared knowledge of major assessment concepts, a shared technical language, and portable portfolio examples that could serve as resources to sustain the ideas and the work with non-Academy colleagues back in their districts. There were, of course, inevitable limitations to the assessment knowledge the community could share. The Academy portfolio's principles and practices were not applicable in the same ways to different curriculum units, so, while teachers could share broad insights about assessment with all of their Academy colleagues, they were less well positioned to share or benefit from each team's specific learning.

Uneven opportunities for learning targeted assessment concepts and methods. Opportunities for learning varied among the Academy teachers in relation to the types of assessments they were developing and using, the quality of the assessments available in their instructional materials, the expertise of team members and the facilitator, and team decisions about their learning priorities. *Assessments:* At any given time, one team might be revising multiple-choice items while another team was revising performance tasks. *Available tools:* Even if two teams were both revising performance assessments, there were often marked differences in the content and quality of the assessments included in the instructional materials. *Expertise:* The expertise of teams differed from one team to the next and from one unit to the next; team members varied in their prior experience developing assessments and analyzing student work, in their science backgrounds, and in their familiarity with particular curriculum units. *Priorities for learning:* Finally, teams made different professional decisions about their priorities for learning. While all teams completed the portfolio work of unit goal planning, tool development, and interpretation and use of evidence, some teams took the initiative to work on additional aspects of assessment such as methods of strengthening scoring reliability, strategies for reducing bias in item design and scoring, and the design or interpretation of sets of items in one instrument. The consequence of these variations in opportunities to learn and professional initiative was that growth across the cohort was uneven. The Academy produced a cadre of professionals who had differing areas of strength in their repertoire of assessment knowledge and strategies (Gearhart et al, 2006).

Implications for Research: The Roles of Generic vs. Content-Specific Programs that Support Teachers' Growth with Classroom Assessment

We close our discussion with one central question that emerged from our findings: What is the role of a generic assessment portfolio in relation to other strategies that could provide more targeted support for specific assessments in curriculum units? Research is needed to identify ways to preserve the generic opportunities provided by the Academy assessment portfolio while ensuring teachers opportunities to build assessment expertise appropriate to their curriculum.

It is a reasonable conjecture that teachers' progress with certain aspects of classroom assessment will eventually require deep engagement with content, an opportunity not easily afforded by a generic portfolio approach. Strengthening the validity of an assessment, for example, requires teachers to analyze the soundness of the science content as well as the capacity of students' responses to the task and the scoring criteria to capture the range of student performance and understanding in the domain. Designing strategies for instructional follow-up and feedback also requires that teachers understand which specific strategies are appropriate to students' conceptual challenges. As teachers build assessment expertise, we believe they eventually need content-specific and targeted support integrating quality assessments for each curriculum unit.

Our findings from the generic Academy portfolio program provide a useful contrast with research on teachers' uses of curriculum-embedded assessments. In these projects, assessment developers collaborate with curriculum developers to embed a series of assessments designed to track student progress based on research on conceptual development (e.g., Shavelson, 2005). The assessments and scoring guides serve as an integrated assessment system that provides teachers ways to anticipate patterns of student learning and track progress systematically. There is emerging evidence of the promise of these systems for teacher learning, classroom practice, and student learning (Herman, et al., 2005; Kennedy, Brown, Draney, & Wilson, 2005; Wilson & Sloane, 2000). Curriculum-embedded assessments could address Academy teachers' requests for higher quality assessment tools, tools that shift teachers' attention from assessment development toward interpretation of students' responses and design of instructional follow-up. Professional development could then target the content-specific knowledge and skills teachers need to implement the assessments effectively. The trade-off, however, might be less opportunity to learn the underlying assessment principles and generalizable strategies for strengthening and using assessments. Teachers might not "take ownership" of the assessments in their instructional

materials and might not be challenged to reflect deeply about the relationship between learning goals, assessment tools and the uses of assessment results.

Research is needed on ways to balance or merge the Academy generic portfolio and curriculum-embedded assessments as complementary opportunities for teacher learning. The generic Academy portfolio approach embraced a diverse community of participants who worked collaboratively to build assessment expertise, and that process fostered a new professional stance toward instructional materials—teachers came to own their materials and the assessments, just as they owned the assessment portfolios they constructed. A content-specific assessment program can strengthen teachers' content and pedagogical content knowledge by engaging teachers with close interpretation of student work, instructional follow-up, and feedback to students. What are the appropriate purposes and contexts for a generic vs. a content-focused classroom assessment program, and how could they be productively coordinated? Continued research on these issues will provide educators more complete understandings of the contexts in which teachers learn best about assessment and ways to use assessment to support student learning.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement and Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Aschbacher, P. (1999). Helping educators to develop and use alternative assessments: Barriers and facilitators. *Educational Policy, 8,* 202–223.

Aschbacher, P., & Alonzo, A. (2006). Examining the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment, 11*, 279–203.

Atkin, J. M., Coffey, J. E., Moorthy, S., Sato, M., & Thibeault, M. (2005). *Designing everyday assessment in the science classroom.* New York: Teachers College Press.

Ball, D. L., Hill, H. C., & Bass, H. (2005, Fall). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator,* 14–23.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Birman, B. F., Desimone, L., Garet, M. S., Porter, A. C., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38,* 915–945

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7–74.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice.* Buckingham, England: Open University Press.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13,* 259–278.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*, 5–12.

Diranna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., et al. (2007). *Assessment-centered teaching.* Thousand Oaks, CA: Corwin Press.

Falk, B., & Ort, S. (1998, September). Sitting down to score: Teacher learning through assessment. *Phi Delta Kappan, 80*(1)*,* 59–64.

Garet, M., Porter, A. C., Desimone, C., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38,* 915–945.

Gearhart, M., & Saxe, G. B. (2004). When teachers know what students know: Integrating assessment in elementary mathematics. *Theory Into Practice, 43,* 304–313.

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S., Schwab, C., Vendlinski, T., et al. (2006). Developing expertise with classroom assessment in K–12 science: Learning to interpret student work. *Educational Assess*ment*, 11,* 237–263.

Goldberg, G. L, & Roswell, B. S. (1999–2000). From perception to practice: The impact of teachers' scoring experience on performance-based instruction and classroom assessment. *Educational Assessment, 6*, 257–290.

Guskey, T. R., (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASP Bulletin, 87*(637), 28–54.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook for policy and practice* (pp. 127–150). San Francisco: Jossey-Bass.

Herman, J. L. (2005, September). *Using assessment to improve school and classroom learning: Critical ingredients.* Presentation at the Annual Conference of the Center for Research on Evaluation, Standards, and Student Testing. University of California, Los Angeles.

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2005, April). *The nature and impact of teachers' formative assessment practices*. Paper presented as part of the symposium, Building Science Assessment Systems that Serve Accountability and Student Learning: The CAESL Model, at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada.

Hill, H. C. & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's Mathematics Professional Development Institutes. *Journal for Research in Mathematics Education, 35,* 330–351.

Kennedy, C. A., Brown, N. J. S., Draney, K., & Wilson, M. (2005, April). *Using progress variables and embedded assessment to improve teaching and learning.* Paper presented as part of the symposium Building Science Assessment Systems that Serve Accountability and Student Learning: The CAESL Model, at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada.

Laguarda, K. G., & Anderson, L. M. (1998). *Partnerships for standards-based professional development: Final report of the evaluation.* Washington, DC: Policy Studies Associates, Inc.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics.* Thousand Oaks, CA: Sage Publications.

Mansvelder-Longayroux, D., Beijard, D., Verloop, N., & Vermunt, J. D. (2007). Functions of the learning portfolio in student teachers' learning process. *Teachers College Record, 109,* 126–159.

National Research Council, National Committee on Science Education Standards. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. Committee on Classroom Assessment and the National Science Education Standards. Center for Education, Division of Behavior and Social Sciences and Education. (2001a). In J. M. Atkin, P. Black, & J. E. Coffey (Eds.). *Classroom assessment and the National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council. Committee on the Foundations of Assessment. Board on Testing and Assessment, Center for Education, Division of Behavior and Social Sciences and Education. *Knowing what students know: The science and design of educational assessment.* (2001b). In J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.). Washington, DC: National Academy Press.

Popham, W. J. (2004.) *Classroom assessment: What teachers need to know* (3rd ed.). Boston: Allyn & Bacon.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner.* San Francisco: Jossey-Bass.

Shavelson, R., Stanford Educational Assessment Laboratory (SEAL) and Curriculum Research & Development Group (CRDG). (2005). *Embedding assessments in the FAST curriculum: The romance between curriculum and assessment.* (Final Report). Palo Alto, CA: Stanford University

Sheingold, K., Heller, J. I., & Paulukonis, S. T. (1995). *Actively seeking evidence: Teacher change through assessment development* (Rep. No. MS #94-04). Princeton, NJ: Educational Testing Service, Center for Performance Assessment.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *The handbook of research on teaching,* (4th ed., pp. 1066–1101). Washington, DC: American Educational Research Association.

Stiggins, R. J. (2005). *Student-involved assessment FOR learning* (4th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Taylor, C. S. (1997) Using portfolios to teach teachers about assessment: How to survive. *Educational Assessment, 4,* 123–147.

Taylor, C. S. & Nolen, S. B. (1996a). A contextualized approach to teaching teachers about classroom-based assessment. *Educational Psychologist, 31,* 77–88.

Taylor, C. S., & Nolen, S. B. (1996b). What does the psychometrician's classroom look like? Reframing assessment concepts in the context of learning. *Educational Policy Analysis Archives, 4*(17).

Taylor, C. S. & Nolen, S. B. (2004). *Classroom assessment.* Saddleback, NJ: Prentice Hall.

Watson, A. (2000). Mathematics teachers acting as informal assessors: Practices, problems and recommendations. *Educational Studies in Mathematics, 41,* 69–91.

Weiss, I. R., & Miller, B. (2006, October). *Deepening teacher content knowledge for teaching: A review of the evidence.* Paper prepared for the Second Mathematics Science Partnership (MSP) Evaluation Summit, Minneapolis, MI.

Wiggins, G., & McTighe, J. (2005). *Understanding by design.* Baltimore: Association for Supervision and Curriculum Development.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*, 49–65.

Wilson, S. M. & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education, 24,* 173–209

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13,* 181–208.

Wilson, S. M. (2004). Student assessment as an opportunity to learn in and from one's teaching practice. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education, Part 2* (pp. 264–271). Chicago: University of Chicago Press.

Wolf, K., & Dietz, M. (1998). Teaching portfolios: Purposes and possibilities. *Teacher Education Quarterly, 25,* 9–22.

Zeichner, K. & Wray S. (2001). The teaching portfolio in U.S. teacher education programs: What we know and what we need to know. *Teaching and Teacher Education 17,* 613–621.