

An Integrated Assessment System as a Medium for Teacher Change and the
Organizational Factors that Mediate Science Teachers' Professional Development

Lily Roberts and Mark Wilson
University of California, Berkeley
April 1998

ORIGINAL VERSION

Running Head: Integrated Assessment System as a Medium for Teacher Change

Key descriptors: embedded assessment; middle school science

Integrated Assessment System as a Medium for Teacher Change

Abstract

Local assessment moderation, a scorer-calibration process for alternative assessments, is the central component of an integrated assessment system, which serves as a medium for science teacher change. This paper describes the assessment system and provides a summary of the literature about organizational factors that mediate teacher professional development. The qualitative methods and analysis of features used to measure differences across Assessment Development Centers is presented. The findings suggest that leadership, institutional support, and teacher proximity and collaboration were the significant features that contributed to between-Center differences on teacher change in their level of success with implementing the integrated assessment system.

Integrated Assessment System as a Medium for Teacher Change

Introduction

There are numerous models for teacher professional development in science and each has its merits as well as its drawbacks (Loucks-Horsley, Hewson, Love, & Stiles, 1998). In the design of an integrated, classroom-based assessment system for a middle school science curriculum, we focused on assessment development. We developed a model that combines the benefits of consensus moderation and of effective professional development principles, which we call local assessment moderation (Roberts, Sloane, & Wilson, 1996).

Local assessment moderation is an example of a community of judgment (Wilson, 1994) that functions as a motivation for teacher change as well as a catalyst for changing the assessment culture of the school or district. Below we describe the integrated assessment system and define both community of judgment and local assessment moderation in more detail.

In our evaluation of the implementation of local assessment moderation, we paid particular attention to the organizational context factors that mediate teacher change. The qualitative evidence reported here is part of a larger evaluation of the effects of this integrated assessment system on teacher professional development (Roberts, 1996) and student achievement (Wilson & Draney, 1997). The organizational context factors that appear to affect successful implementation of local assessment moderation include: quality and strength of leadership; level of institutional support for teacher professional development; and teacher proximity and collaboration.

The SEPUP Assessment System

The Science Education for Public Understanding Program received funds from the National Science Foundation (NSF) to develop and field-test a year-long, middle school science course with an integrated assessment system, entitled *Issues, Evidence and You (IEY)*. The SEPUP Assessment Project was designed to apply new theories and methodologies in the field of assessment to the practice of teacher-managed, classroom-based assessment of student performance (Wilson & Adams, 1996).

Integrated Assessment System as a Medium for Teacher Change

In congruence with the National Science Education Standards (National Research Council, 1996), the SEPUP *IEY* curriculum is designed to engage students in an “issues-oriented, hands-on” approach to thinking about scientific issues that are relevant to their daily lives (e.g., water, waste, energy, and environment). Further, student understanding is assessed in an ongoing manner using embedded tasks. The students’ role is changed as they conduct labs and other activities that are designed to help them understand that science is really a way of asking and answering questions and not just a collection of established facts that they are asked to memorize. The teachers’ role is that of facilitator in the students’ development. The aim is that assessment information becomes a scaffolding mechanism for instructional change that further facilitates student learning. Tobin (1995) indicates that any “critical approach” to teacher development must provide teachers with “autonomy to identify problems in their classrooms and seek solutions that make sense to them” (p. 147). With SEPUP assessment, the long-term goal is for teachers to become autonomous assessors of their students’ understanding of science, which will be evidenced by their own professional development in terms of changing instructional and assessment practices.

This classroom-based assessment system provides a set of tools for teachers to use to: (a) assess student performance on core concepts and skills in the curriculum; (b) set standards of student performance; (c) track student progress over the year on the central concepts; and (d) provide feedback (to teachers, students, administrators, parents, or other audiences) on student progress as well as on the effectiveness of the curriculum materials and classroom instruction. Initially, managing a new classroom-based system of embedded assessment demands much of the teacher. To encourage and support teacher professionalization in the field of assessment, we incorporated local assessment moderation as a central component of this system. The eight components of the SEPUP Assessment System are presented in Figure 1.

Insert Figure 1 about here

Integrated Assessment System as a Medium for Teacher Change

An Integrated Assessment System as a Medium for Teacher Change:

Creating a Community of Judgment

The strategy of employing an integrated assessment system that encompasses many of the characteristics of high quality professional development evolved from the concept that teachers need to form a community of judgment (Wilson, 1994), which can serve as a network much like the “professional-area movement organizations” described by Pennell and Firestone (1996). Wilson (1994) proposed the community of judgment as one approach to changing the “assessment culture” from one that relies primarily on standardized tests to a teacher-centered approach to educational accountability. Efforts to change the assessment culture are direct responses to calls for better assessments given the limitations of standardized tests (Bullough, 1988; Linn, 1987; Office of Technology Assessment, 1992; Shepard, 1989; Spring, 1988; Wiggins, 1989). To change the assessment culture, it seems clear that one must begin with the teacher who works directly with the students being measured. Shepard (1989) suggests that teacher enhancement is needed and that teachers should be directly involved in developing and scoring assessments. However, a comprehensive embedded assessment system goes far beyond the scoring task.

Community of Judgment

Wiggins (1989) suggested that any system of truly authentic assessment should meet several general criteria; the system should be: (a) criterion-referenced; (b) formative; (c) moderated; and (d) clear in the progression of educational development (e.g., score levels on a rubric would reflect the criteria). Wilson’s (1994) community of judgment for educational accountability satisfies all of these criteria, while adding a fifth essential component -- a substantive framework that describes the achievement variables that are valued and thus worth assessing. Broad frameworks already exist in various subject areas, such as the Benchmarks for Science Literacy (AAAS, 1993) or the Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989). These broad frameworks serve as guidelines for local development and implementation of curriculum and assessment.

Integrated Assessment System as a Medium for Teacher Change

The substantive framework for assessment and accountability suggested by Wilson (1994) requires specific variables within a particular curriculum to be defined and measured, for example, throughout a school year to chart students' progress. Wilson's assessment framework specifies the constructs embedded in a course, such as the Designing and Conducting Investigations variable in Issues, Evidence and You -- thus assessment and instruction are conjoint rather than disparate elements of teaching.

In addition to a substantive framework, the community of judgment (Wilson, 1994) model also contains (a) a set of assessment modes for each variable and scoring guidelines for rating responses; (b) moderation of the teachers' judgments that allows for rater improvement (i.e., teacher enhancement in assessment) and rater adjustment (i.e., consensus building to set standards for student performance); and (c) methods of quality control that ensure technical measurement standards (e.g., reliability and validity). Assessment modes are a variety of indicators of student achievement, including tasks such as lab reports, performances (e.g., a town meeting) or projects. These modes are designed to promote higher order thinking, to engage students in real-world, meaningful activities, and to be as closely aligned as possible to the criteria of interest. Moderation is a scorer-calibration process that can also serve as a mechanism for teacher professional development (Ingvarson, 1990; Linn, 1994; Roberts, Sloane, & Wilson, 1996). However, moderation is being used on a very limited basis in this country (Linn & Baker, 1996). Quality control is needed to ensure comparability of teacher judgments across different settings and to authenticate student work (i.e., prevent cheating).

For teachers to assume a central role in this community of judgment for educational accountability, professional development is a must (Jett & Schafer, 1992; Linn, 1994; Newell, 1992). Particular to improving science education, good teaching materials, "vigorous and ongoing" professional development, and new "assessment tools" are basic components of successful reform (Lopez & Tuomi, 1995, p. 78). Richardson (1990) notes that "teachers themselves must be involved in making judgments about what change is worthwhile and significant" and "that practices and ways of thinking outside an individual teacher's own

Integrated Assessment System as a Medium for Teacher Change

experiences should be introduced into the dialogue” (p. 14). The community of judgment model can be used to engage teachers in making such judgments while benefiting from the interaction with colleagues. Interaction with colleagues is a key element for many forms of science teacher professional development (Loucks-Horsley, Hewson, Love, & Stiles, 1998).

Alternative science assessments will be implemented “only if teachers understand their use and the depth of the content they demand, are empowered to make instructional decisions, and are supported by school districts which encourage teacher change” (Harmon, 1995, p. 46). If teachers are engaged in a community of judgment for educational accountability, then conditions will hold and hence it is far more likely that alternative assessments will be implemented and that they will be meaningful and useful in terms of informing teachers’ instructional practices. Moderation provides teachers a forum for discussing science education and assessment in a meaningful way.

Furthermore, moderation affects the norms of the professional culture within in a school or district, which is essential for “changing norms of practice and pedagogy” (Loucks-Horsley, Hewson, Love, & Stiles, 1998, p. 195).

To meet the challenges of science education reform, the professionalization of teaching demands that teachers change not only their instructional practices, but also their assessment practices. Further, this change in teachers cannot and, in fact, will not occur in a vacuum; therefore providing for a professional community of judgment that serves as a medium for teacher change is essential. Local assessment moderation serves this function in the SEPUP Assessment System.

Local Assessment Moderation

SEPUP teachers engaged in local assessment moderation function as a community of judgment, and this task engagement serves as a medium for high quality professional development. Principles of high quality professional development, such as ongoing support rather than one-shot experiences (Fullan, 1991) and “support for informed dissent” (Little, 1993, p. 138), delineated by these and other researchers (e.g., Sparks & Loucks-Horsley, 1990) were adhered to in the design of local assessment moderation for SEPUP. Local assessment moderation is a critical element of

the SEPUP Assessment System as it supports teacher professional growth in the politically charged arena of assessment reform.

Moderation in SEPUP consisted of monthly meetings of small groups of teachers (4 to 6 per Assessment Development Center (ADC)). The teachers selected three to five samples of student work from the same activity and brought these to the moderation meeting. The moderation participants then scored all the samples of student work, shared their scores with the group, and then discussed variations in scoring to reach consensus. Teachers also used this to discuss instructional implications as well as reflect on their own practice.

An ongoing evaluation of the implementation of the consensus or assessment moderation process in the state of Victoria in Australia was conducted from 1981 to 1984, with a follow-up study in 1989. Based on the four year evaluation in Victoria, Ingvarson (1990, p. 9) noted “the importance of regarding moderation as a complex innovation requiring a considerable period of time for [teachers’] learning and unlearning during its implementation.” The Victorian study indicates that the consensus moderation process had “impressive side effects on the professional development and accountability of teachers” (p. 2). Many of the comments from the SEPUP teachers are consistent with the findings of the Victorian evaluation (Ingvarson, 1990). In the Victorian evaluation, it was found that involvement in the consensus moderation process:

- (a) added significantly to teachers’ skills for assessing student learning;
- (b) enhanced teachers’ ability to evaluate and improve their teaching;
- (c) significantly increased teachers’ access to useful ideas for teaching;
- (d) enhanced the quality of learning of students;
- (e) affected positively participants’ teaching in non-project classes; and
- (f) supported, rather than intimidated, beginning teachers.

Ingvarson (1990) also reported that the positive responses increased as teachers had more experience with moderation, which again reflects the need for teachers to have time to become knowledgeable and skilled in using this process. This experience factor was also evident during the evaluation of SEPUP (Roberts, 1996).

Organizational Context Factors: The Importance of Situation

In this section, we frame the organizational context for teaching from the inside out, that is from the classroom to the national level. We then present a table that summarizes the key elements of the organizational factors that mediate teacher professional development by level of organization.

First, teachers operate autonomously in their own classrooms with a group of students. This group may change each period as in secondary schools or be fairly constant throughout the day. The characteristics of the students influence teachers' instructional practices as well as personal reactions. Teachers' beliefs and values do affect their classroom practice.

Second, teachers work within a school and their relationships with the principal and other teachers influence their teaching in various ways. The norms, values and expectations at the school level can function to support teacher change or to facilitate teacher isolation (McLaughlin, 1991). The climate of the school is based on both the community in which it exists and the community that exists within. In some unique cases, the community within can overcome the deficits of the external community, such is the case of Central Park East in East Harlem (Meier, 1995).

Third, teachers work within a district that sets local curriculum, grading and other policies while controlling the purse for staff development, school improvement and other activities. The personnel office is located at the school district; salaries and other incentives are discharged by the district; and the district administers the mandated tests. Besides funding and coordinating the staff development function at the local level, the district sets the tone for the schools within its purview. As McLaughlin (1991) notes, the district sets the norms, values and expectations for teacher change. School boards and teachers' unions are two other spheres of influence that vary by locale, and in some cases, wield a great deal of power and influence.

Fourth, teachers work within a state that has an education code that directs the course of curriculum, testing, special programs for students, and so forth. Some states have exit exams for high school while others have curriculum frameworks to guide instruction from grades K-12. State education agencies also offer teacher professional development programs that serve various

Integrated Assessment System as a Medium for Teacher Change

needs -- from smoothing the teacher induction process to helping elementary teachers be better prepared to teach mathematics to disadvantaged students. At the state level, there are also professional organizations or teacher networks, such as the various state Science Teachers' Associations or other professional organizations like the California Subject Matter Projects (in mathematics, science and so on), that offer annual conferences and other professional development opportunities.

Fifth, teachers work within a national education system that serves a very heterogeneous student population. The future work force is expected to be highly mobile because jobs and careers are expected to change rapidly. Teachers are expected to teach an ever changing student population with limited resources for themselves and the students. However, national goals for educational attainment do not directly support change at the classroom level. Teachers go to work, facing the tremendous challenge of educating youth for a dynamic future, and the year 2000 is just around the corner.

Given past research, how will these different levels of organization in which teachers are nested promote teacher change and professional learning? Guskey's (1986) model of the teacher change process based on staff development activities indicated that teachers did not begin to change their beliefs or attitudes until they experienced change in their students' learning outcomes presumably based on some change in the teacher's classroom practice. Newell (1992) also found this to be a critical factor in middle school science teachers' adoption of alternative assessment. Guskey writes "teachers' knowledge of teaching is validated very pragmatically, and that without verification from the classroom, attitude change among teachers with regard to any new program or innovation is very unlikely" (p. 7).

McLaughlin (1991) enumerates several factors that are required to enable teachers to change; these are: ongoing assistance, opportunities for collegial activities, concrete training and follow through, and principal's support and encouragement. Figure 2 provides a summary of the literature related to the organizational contexts that can mediate teacher change. The rows represent the levels described earlier from the classroom to the national level. The four columns represent

the primary areas reported by various researchers. The cells identify specific issues in effect at a particular level in each of the areas for which differential responses at the organizational level may affect teacher change. For example at the classroom level, student characteristics may affect teacher change.

Insert Figure 2 about here

The factors presented in Figure 2 are drawn from the following literature:

1. The role of the principal and the organizational context in teacher change has been widely studied. One major example is the Rand Change Agent Study (McLaughlin & Marsh, 1978), which was updated to indicate factors that influence high quality professional development (McLaughlin, 1991). Many others have also addressed the importance of the principal's leadership to the success of teacher change (Fullan, 1991; Hawley & Rosenholtz, 1984; Murnane & Raizen, 1988; Rosenholtz, Bassler, & Dempsey, 1986; Warren, 1975; Weiss, 1993).
2. Beginning with Little's (1982) study of norms of collegiality and experimentation, others have continued to focus on the "norms of the workplace" (Richardson, 1994) or school climate, especially in terms of collegiality and the influence of collaboration on teacher change (Hollon, Roth, & Andersen, 1991; Ross & Regan, 1993; Tobin & Espinet, 1989).
3. As noted earlier, student characteristics and the classroom context play an extremely important role in teacher change (Hawley & Rosenholtz, 1984; Lederman, 1992; Lee, Dedrick, & Smith, 1991; Lee & Porter, 1993; McLaughlin, 1993; Warren, 1975).
4. The school climate and the sociocultural context (i.e., parents and the community) in which the school exists influences teachers in many ways (Richardson, 1990; Warren, 1975).
5. Administrative policies affect the daily lives of teachers (Lederman, 1992; McLaughlin, 1991) including local and state-mandated tests, grading and reporting requirements as well as the norms and expectations set at different organizational levels for teacher performance. Sparks and Loucks-Horsley (1990) recommend that there be a "common, coherent set of goals" for teacher professional development at both the school and district levels (p. 245).

Integrated Assessment System as a Medium for Teacher Change

The overwhelming evidence from these various studies confirms that teacher professional development alone, regardless of quality, does not necessarily result in significant differences in classroom practices and student experiences. St. John (1991) notes, “As with the history of curriculum development, the lessons learned about professional development indicate that it also must address the powerful broader contextual factors that implicitly but strongly shape classroom practice” (p. 13).

Methods

This section presents the qualitative elements of data collection, sources and techniques, and strategies for data retrieval and display. The qualitative data analytic procedures are also described. The case-ordered matrix approach was central to characterizing the features of SEPUP Assessment Development Centers (ADCs) that mediate teacher change.

Subjects

Three treatment groups were identified for the overall study: SEPUP Assessment Development Centers (ADCs); SEPUP Professional Development Centers (PDCs); and a non-SEPUP comparison group. Teachers in all three groups taught science in grades 7, 8 or 9. All SEPUP teachers used the Issues, Evidence and You curriculum and the non-SEPUP teachers used their traditional science curriculum. The SEPUP ADC teachers received the most comprehensive treatment, both the curriculum and the assessment system components, which included regular participation in local assessment moderation. The SEPUP PDC teachers received the curriculum containing the embedded assessments and scoring guides, but no support for using the assessment system.

This purposive sample consists of all teachers involved in the 1994-95 field test of SEPUP's IEY and the embedded assessment system. There were 26 SEPUP teachers and seven comparison teachers in six ADCs. In the six PDCs, there were 25 SEPUP teachers and five comparison teachers. Some Centers were located in a single, large school district and a district representative functioned as the ADC director. Other Centers were organized less centrally, for

Integrated Assessment System as a Medium for Teacher Change

example, one ADC had five SEPUP teachers from five different districts, and one of these teachers also served as the ADC director. In some cases, the Center director was not from the local school or district, but rather a university person involved in science education.

Data Sources and Techniques

The qualitative study focused on the teachers within in Assessment Development Centers. The major sources of qualitative information about the impact of the program on ADC teachers were self-reports and field notes.

Self-report techniques.

The two self-report techniques used to collect data from the ADCs were personal interviews and focus groups. During pre and post site visits, personal interviews were conducted with most of the SEPUP teachers, some principals, and the ADC Directors (post-only). All interviews were conducted using a structured interview protocol containing both closed- and open-ended questions. The goal was to obtain evidence on common factors (e.g., external testing mandates faced by teachers) from multiple sources of informants, thus enhancing validity as well as reliability. We are using reliability here as an aspect of construct validity as proposed by Moss (1994, p. 7), where she notes reliability in this context as “consonance among multiple lines of evidence supporting the intended interpretation over alternative interpretations.”

The focus groups were conducted using a structured, open-ended guide (i.e., protocol). The face validity of focus groups is considered to be quite high (Krueger, 1994). The focus group was deemed an appropriate method to use with the ADC groups, because they had spent at least a year working together in moderation meetings. The focus group was an opportunity to obtain information on the participants’ feelings about the moderation process, such as perceptions of its usefulness, advantages and disadvantages, and feasibility for future use. The focus group was conducted primarily for the purpose of collecting corroborating evidence. This technique was useful for triangulation with the teacher interviews as well as a source from which to check the representativeness of the cases (e.g., look at the outliers). In qualitative research, it is important to be aware of potential bias stemming from a researcher effect (Miles & Huberman, 1994, p. 265-

266). One way to avoid such bias is through multiple data collection methods, so the focus group served this purpose as well.

Field note techniques.

The second method for data collection was field notes. Techniques used to collect field notes included unobtrusive classroom observations, observation of moderation sessions, and informal and formal discussions with students and other teachers. During and after site visits (both pre and post site visits), summary reports were produced based on field notes. The synthesis of the field notes into site visit summaries served as a data reduction mechanism, weeding out rival explanations and identifying plausible interpretations. The post site visit summaries were shared with the ADC Directors for the purpose of double checking researcher interpretations for accuracy, and as suggested by Miles and Huberman (1994, p. 275-277), gaining corroboration from key informants.

Data Retrieval and Display

The interviews were tape recorded to facilitate information collection and reduce potential researcher bias. Comments and responses were also recorded by hand on the interview protocol forms. Transcription was not used. Since one of the primary purposes of the qualitative data was to answer the question -- How do ADCs differ? -- it was not necessary to use transcribed dialogues to answer this question.

A matrix approach was used for data display purposes. The case-ordered descriptive matrix was useful in identifying patterns and “seeing plausibility” (Miles & Huberman, 1994, p. 246-248) in the data as well as making comparisons across ADCs. For the matrix cell entries, researcher explanations were used. These were corroborated in two ways. First, the qualitative data syntheses were reviewed by several members of the SEPUP Assessment Team and one other outside reviewer. And second, the ADC Directors were given the opportunity to review the site visit report summaries that were the basis for the case-ordered matrix. This latter form of corroboration is also referred to as “member checks” (Rudestam & Newton, 1992, p. 76), where

participants are given the opportunity to reconsider the interpretations of the data made by the researcher.

Qualitative Data Analysis

The condensed field notes and self-reports were used to build a case-ordered matrix in which the Centers were ordered empirically from least to most successful with the local assessment moderation process. Using three features that were drawn from the literature cited earlier, the qualitative data was analyzed systematically using the matrix approach to identify the features that contributed to the between-Center variation in teacher change that was identified in the quantitative analysis of measures of teacher change reported elsewhere (Roberts, 1996; 1997).

Results

Teacher change can be described as a *situated process* (Roberts, 1996). Situated in two senses; first, as a process that occurs as a teacher engages in actively integrating theory and practice, and second, situated in terms of the organizational context. Local assessment moderation provided a window through which teacher change could be studied systematically in terms of both individual growth as well as organizational impact on teacher development.

Contextual Factors: Qualitative Evidence

The level of success with implementation of the SEPUP Assessment System was determined to a large extent by the organizational context; in other words, the Center mattered. Features of the Assessment Development Centers (ADCs) that mattered include: quality and strength of leadership; institutional support for teacher professional development; and teacher proximity and collaboration. Figure 3 summarizes the qualitative evidence for the four ADCs, which are ordered from most to least successful with local assessment moderation. Success was measured in terms of: (1) completion of field test requirements by teachers within each ADC, such as number of activities moderated and selection of student exemplar papers; (2) teachers' reported levels of growth based on personal interviews and interviews with their principals; and (3) observations of the moderation process corroborated by member checks with the ADC Directors.

INSERT FIGURE 3 ABOUT HERE

Leadership for Change.

The ADC teachers in general came from different schools, so their experiences with “norms of collegiality” (Little, 1982) varied. Some teachers felt isolated in their schools while others participated in science department or grade level teams. The key to bringing the group together cohesively was the leadership provided by the ADC Director or by strong teachers within the group who had prior SEPUP experience. The importance of leadership runs the gamut of ensuring that the group develops a rapport, to preparing teachers to use local assessment moderation, to facilitating moderation sessions.

Who the leader is may not be all that important, but according to the ADC teachers, there needs to be someone in a leadership position to move the group toward consensus, to intervene in personal conflicts, to diffuse philosophical differences that digress from the work at hand, and overall to keep the group on task. Time, as always, is a factor for teachers, so maintaining task orientation is important.

Institutional support.

As Little (1993) and others have suggested, a quality professional development program needs to balance support for institutional initiatives with support for those initiated by teachers. For some of the ADCs, the purpose of changing teachers’ assessment practices was consistent with state initiatives or local school or district-level staff development goals. These goals were balanced with the needs of the participating teachers, who for the most part were very successful science teachers, but were interested in learning about new assessment strategies. In terms of school-based administration, teachers need support from their principals, and in the case of SEPUP this was generally not a problem. Many of the SEPUP teachers have a long history of involvement in innovations and have been supported by their principals.

Strong leadership was also important in securing district-level support. The most critical support factor related to the district was gaining access to the ADC teachers for whole-day or half-

Integrated Assessment System as a Medium for Teacher Change

day meetings rather than after-school sessions, including release time and substitutes. Having sufficient time as a group was important for three fundamental reasons; time is needed: (1) to build a group rapport or a SEPUP norm of collegiality; (2) to build teacher understanding of the SEPUP assessment system; and (3) for teachers to function as reflective practitioners. Having district support has broader implications as well, but with time being an oft-noted issue for teachers, it is critical to not lose sight of the opportunity cost of teachers' time. In two of the ADCs, the ADC director was a representative from the district.

Teacher Proximity and Collaboration.

Having teachers from more than one district can work. In ADC 2 (refer to Figure 3), two teachers from a smaller district, an hour's drive north of the larger participating district, communicated with each other between moderation meetings. Teachers in the larger district tended to call the ADC Director (who worked at the District office) or one of the experienced SEPUP teachers in their own district. Given the long commute to attend meetings for three of the six teachers, this ADC met for whole days. These two districts had an outstanding history of collaboration. However as noted above, too many districts can be detrimental and minimize the leadership of the ADC director.

Teacher collaboration is related to proximity in that teachers who are closer tend to be able to spend more time communicating with one another. One of the teachers in ADC 2 noted that she would have liked to confer with one of the teachers from the other district, but that it was a long distance call, so she tended not to contact her between meetings. Teachers most often called one another to ask a question about an assessment task or to clarify something on a scoring guide. On occasion, teachers would actually meet or pursue a mutual project together.

Based on site visit observations, the quality of moderation varied, but so did the amount of experience with using local assessment moderation and the amount of time devoted to preparing teachers to engage in moderation. As noted above, the full days that some ADCs used were critical not only to forming a cohesive group, but to learning about the moderation process and learning from each other how the assessments were working in their classrooms. In this way, the

Integrated Assessment System as a Medium for Teacher Change

moderation meetings became the ongoing support that teachers needed as they learned about and implemented the SEPUP course and assessment system.

Discussion

Providing a forum for teacher discourse about assessment that enhances professional growth is the principle purpose for incorporating local assessment moderation in the SEPUP Assessment System. Without building a community of judgment, we believe efforts to change the assessment culture within schools or districts are likely to fail. Further, without ongoing support teachers are not likely to make progress in changing their assessment practices.

Contribution of Local Assessment Moderation to Teacher Change

The situated nature of teacher change requires that any evaluation of growth take into account the context, including organizational barriers and supports as well as the teachers' own proximity to colleagues as well as predisposition to collaborate. As noted by Loucks-Horsley and her colleagues (1998), professional communities thrive on collaboration, experimentation and challenging discourse" (p. 197). However, they admonish that collaboration alone is an insufficient condition to promote teacher change. The community of judgment created by teachers engaging in local assessment moderation can be a thriving community if the situation is amenable. Yet their caveat is duly acknowledged. Simply coming together to moderate in a rote fashion, as was the case for the least successful ADC in Figure 3, was an example of how collaboration can be a meaningless exercise. To borrow from Shakespeare, this ADC's moderation meetings were "full of sound and fury signifying nothing." There were also success stories -- those ADCs where teachers came together to moderate student papers, engage in rich dialogue about assessment and instruction, provide constructive criticism in a safe environment, and trade strategies for improving classroom practice. These teachers communicated about what students know about science and how they could help them to progress in their understandings or to undo misconceptions.

For ADC teachers to manage the SEPUP Assessment System, ongoing support and professional development was essential for successful implementation. SEPUP field test teachers

Integrated Assessment System as a Medium for Teacher Change

told us that “teachers need to internalize the [SEPUP] variables.” In other words, teachers need to develop an understanding of what it means, for example, to have their students “using evidence to make tradeoffs.” Simply giving teachers scoring guides (as was the case with the teachers in the Professional Development Centers) did not sufficiently promote an understanding of the SEPUP variables, such as Evidence and Tradeoffs (Roberts, 1996). Teachers learned about variables by engaging in local assessment moderation in which they were able to set common standards for judging student performance.

The consensus of the ADC participants was that ongoing support through local assessment moderation was critical to teachers’ use of the SEPUP curriculum as well as the assessment system. As teachers used assessment tasks and the scoring guides on their own, they encountered questions or issues of implementation that they shared during moderation meetings with their colleagues. These meetings provided the ADC teachers the ongoing support necessary to change their instructional and assessment practices. Through moderation, teachers reported that they were able to: (a) build an understanding of embedded assessment and use it effectively; (b) gain insights about student learning through discussion with colleagues; and (c) reflect on their teaching practice.

Mediating Factors for Assessment Reform

The organizational context factors identified in this study that affected teachers’ success with use of the SEPUP Assessment System were: leadership; institutional support; and teacher proximity and collaboration. The Assessment Development Centers with strong local leadership, institutional support from the district and/or school level, and teachers who collaborated regularly were the most successful in implementing the SEPUP Assessment System.

These three factors are consistent with the literature about variables that mediate teacher professional development (see Figure 2). Leadership for change is consistent with studies related to school climate and professional support, which suggest that a supportive environment can be conducive to teacher change. Institutional support is consistent with the literature about ongoing assistance and administrative policies that facilitate teacher development. Teacher collaboration is

Integrated Assessment System as a Medium for Teacher Change

congruent with the research that reports norms, values and expectations for change that “celebrate the professionalism” of teachers (McLaughlin, 1991, p. 77).

We conclude that district level support was the most conducive to success in the Assessment Development Centers in this study. Districts can support teacher professional development in various ways, such as financial resources or release time. Districts are more likely to implement reform activities than an individual school (Levine, 1995) and assessment reform once implemented is more likely to be sustained with district-level commitment (Chrispeel, 1997).

Integrated Assessment System as a Medium for Teacher Change

Reference List

- American Association for the Advancement of Science (AAAS), Project 2061. (1993). Benchmarks for science literacy. New York: Oxford University Press.
- Bullough, Jr., R.V. (1988). Excellence and testing. Ch. 4 in The Forgotten Dream of American Public Education. Ames: Iowa State University Press.
- Chrispeel, J.H. (1997). Educational policy implementation in a shifting political climate: The California experience. American Educational Research Journal, 34 (3), pp. 453-481.
- Fullan, M. (1991). The new meaning of educational change (2nd edition). New York: Teachers College Press.
- Harmon, M. (1995). The changing role of assessment in evaluating science education reform. In R.G. O'Sullivan (Ed.) Emerging roles of evaluation in science education reform (pp. 31-52). New Directions for Program Evaluation, Number 65. San Francisco, CA: Jossey-Bass Publishers.
- Hawley, W., & Rosenholtz, S.J. (1984). School leadership and student learning. Peabody Journal of Education, 61 (4), pp. 53-83.
- Hollon, R.E., Roth, K.J., & Andersen, C.W. (1991). Science teachers' conceptions of teaching and learning. In Brophy (ed.), Advances in Research on Teaching (Volume 2), pp. 145-185. Greenwich, CN: JAI Press Inc.
- Ingvarson, L. (1990). Enhancing professional skill and accountability in the assessment of student learning. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. 327 558)
- Jett, D.L., & Schafer, W.D. (1992). Classroom teachers move to center stage in the assessment area--ready or not! Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. 346 144)
- Krueger, R.A. (1994). Focus groups: A practical guide for applied research (2nd ed.). Thousand Oaks, CA: Sage.

Integrated Assessment System as a Medium for Teacher Change

- Lederman, N.G. (1992). Students' and teachers' conceptions of the nature of science: A review of the literature. Journal of Research in Science Teaching, 29 (4), pp. 331- 359.
- Lee, V.E., Dedrick, R.F., & Smith, J.B. (1991, July). The effect of the social organization of schools on teachers' efficacy and satisfaction. Sociology of Education, 64, pp. 190-208.
- Lee, O., & Porter, A.C. (1993). A teachers' bounded rationality in middle school science. Teaching and Teacher Education, 9 (4), pp. 397-409.
- Levine, H. (1995). The district as the fundamental unit of change. In St. John, M., Levine, H., Håkansson, S., Lopez-Freeman, M., Roberts, L., Shattuck, J., & Von Blum, R., A Study of the California Dwight D. Eisenhower Mathematics and Science Education State Grant Program: Lessons Learned about Designing Professional Development Projects, (Essay III). Inverness, CA Inverness Research Associates.
- Linn, R.L. (1987). Accountability: The comparison of educational systems and the quality of test results. Educational Policy, 1 (2), pp. 181-198.
- Linn, R.L. (1994). Performance assessment: Policy promises and technical measurement standards. Educational Researcher, 23 (9), 4-14.
- Little, J.W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. American Educational Research Journal, 19 (3), pp. 325-340.
- Little, J.W. (1993). Teacher professional development in a climate of educational reform. Educational Evaluation and Policy Analyses, 15 (2), pp. 129-151.
- Lopez, R.E., & Tuomi, J. (1995). Student-centered inquiry. Educational Leadership, 52 (8), pp. 78-79.
- Loucks-Horsley, S., Hewson, P.W., Love, N., & Stiles, K.E. (1998). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin Press.
- McLaughlin, M.W. (1991). Enabling professional development: What have we learned? In A. Lieberman & L. Miller (Eds.), Staff development for education in the '90s: New demands, new realities, new perspectives (2nd ed.; pp. 61-82). New York: Teachers College Press.

Integrated Assessment System as a Medium for Teacher Change

- McLaughlin, M.W., & Marsh, D.D. (1978). Staff development and school change. Teachers College Record, 80 (1), pp. 69-94.
- Meier, D. (1995). The power of their ideas: Lessons for America from a small school in Harlem. Boston: Beacon Press.
- Moss, P.A. (1994). Can there be validity without reliability? Educational Researcher, 23 (2), pp. 5-12.
- Murnane, R.J., & Raizen, S.A. (1988). Indicators of Teaching Quality (Ch. 6, pp. 90-118). In Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12. Washington, D.C.: National Academy Press.
- National Council of Teachers of Mathematics (NCTM). (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM.
- National Research Council. (1996). National science education standards. Washington, D.C.: National Academy Press.
- Newell, S.T. (1992). Science teachers' perspectives on alternate assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1992.
- Office of Technology Assessment. (1992). Testing in American Schools: Asking the Right Questions. Washington, D.C.: Congress of the United States.
- Pennell, J.R., & Firestone, W.A. (1996). Changing classroom practices through teacher networks: Matching program features with teacher characteristics. Paper presented at the annual meeting of the American Educational Research Association, New York, NY, April.
- Richardson, V. (1990). Significant and worthwhile change in teaching practice. Educational Researcher, 19 (7), pp. 10-18.
- Richardson, V. (1994). Conducting research on practice. Educational Researcher, 23 (5), pp. 5-10.

Integrated Assessment System as a Medium for Teacher Change

- Roberts, L. (1997). Evaluating Teacher Professional Development: Local Assessment Moderation and the Challenge of Multisite Evaluation. Paper presented at the annual meeting of the National Evaluation Institute. Indianapolis, July 9, 1997.
- Roberts, L. (1995). Assessment as a vehicle for teacher professional development: Moving teachers further along the continuum. In St. John, M., Levine, H., Håkansson, S., Lopez-Freeman, M., Roberts, L., Shattuck, J., & Von Blum, R., A Study of the California Dwight D. Eisenhower Mathematics and Science Education State Grant Program: Lessons Learned about Designing Professional Development Projects, (Essay V). Inverness, CA Inverness Research Associates.
- Roberts, L.L.C. (1996). Methods of evaluation for a complex treatment and its effects on teacher professional development: A case study of the Science Education for Public Understanding Program. Unpublished dissertation, University of California at Berkeley.
- Roberts, L., Sloane, K., & Wilson, M. (1996). Local assessment moderation in SEPUP. Paper presented at the Annual Meeting of the American Educational Research Association. New York, April.
- Rosenholtz, S.J., Bassler, O., & Hoover-Dempsey, K. (1986). Organizational conditions of teacher learning. Teaching and Teacher Education, 2 (2), pp. 91-104.
- Ross, J.A., & Regan, E.M. (1993). Sharing professional experience: Its impact on professional development. Teaching and Teacher Education, 9 (1), pp. 91-106.
- Rudestam, K.E., & Newton, R.R. (1992). Surviving your dissertation: A comprehensive guide to content and process. Newbury Park, CA: Sage Publications.
- Shepard, L.A. (1989). Why we need better assessments. Educational Leadership, 46 (7), pp. 4-9.
- Sparks, D., & Loucks-Horsley, S. (1990). Models of staff development. In Houston, W.R. (ed.) Handbook of Research on Teacher Education. New York: Macmillan Publishing Company.
- Spring, J. (1988). Conflict of interests: The politics of American education. New York: Longman.

Integrated Assessment System as a Medium for Teacher Change

- Tobin, K. (1995). Teacher change and the assessment of teacher performance. In B.J. Fraser & H.J. Walberg (Eds.), Improving Science Education. Chicago, IL: University of Chicago Press.
- Tobin, K., & Espinet, M. (1989). Impediments to change: Applications of coaching in high school science teaching. Journal of Research in Science Teaching, 26 (2), pp. 105-120.
- Warren, R.L. (1975). Context and isolation: The teaching experience in an elementary school. Human Organization, 34 (2), pp. 139-148.
- Weiss, C.H. (1993). Principals' roles in shared decision making. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 16, 1993.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, pp. 703-713.
- Wilson, M. (1994). Community of judgment: A teacher-centered approach to educational accountability. In Office of Technology Assessment (Ed.), Issues in Educational Accountability. Washington, D.C.: Office of Technology Assessment, United States Congress.
- Wilson, M., & Adams, R.J. (1996). Evaluating progress with alternative assessments: A model for Chapter 1. In M.B. Kane (Ed.), Implementing performance assessment: Promise, Problems and Challenges. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Wilson, M., & Draney, K. (1997). Developing maps for student progress in the SEPUP assessment system. Paper presented at the Middle School meeting of American Association for the Advancement of Science (AAAS). Seattle, WA, February.

Figure Captions

Figure 1. SEPUP assessment components at a glance

Figure 2. Factors that mediate teacher professional development by level of organization

Figure 3. Level of success with local assessment moderation by features of the Assessment Development Centers

- 1. SEPUP Variables** - Five variables which represent student learning in terms of the core concepts of IEY.
- 2. Assessment Tasks** - Activities that are an integral part of regular instruction.
- 3. Scoring Guides** - Rubrics that establish baseline criteria for assessing levels of student performance.
- 4. Assessment Blueprints** - Sequential list of IEY activities and opportune points in instruction for assessing student learning.
- 5. Exemplars** - Examples of student work that have been scored and moderated for each variable and score level.
- 6. Assessment Moderation** - A process by which teachers discuss and reach consensus on local standards for scoring student work.
- 7. Performance Maps** - Graphical representations of student development on the SEPUP Variables.
- 8. Link Tests** - Additional assessment activities for teachers' use at major course transitions that are also based on the SEPUP Variables.

Level	School Climate	Professional Support/ Ongoing Assistance	Administrative Policies	Norms, Values and Expectations for Change
Classroom	<ul style="list-style-type: none"> - Student characteristics - Student behavior - Parental involvement 	<ul style="list-style-type: none"> -Substitutes -Sufficient preparation time 	<ul style="list-style-type: none"> - Testing - Grading/report cards 	<ul style="list-style-type: none"> - Meet needs of all students - Norm of collegiality rather than privacy
School	<ul style="list-style-type: none"> - Peers/Colleagues - Recruitment and socialization of new teachers - Principals promote a norm of collegiality 	<ul style="list-style-type: none"> - Instructional coordination - Common prep time for teachers (by subject and/or grade) - Collaborative style of leadership and interaction 	<ul style="list-style-type: none"> - Common, coherent set of goals for teacher professional development - Shared decision making 	<ul style="list-style-type: none"> - Principal sets the tone for teacher change - Norm of collegiality and acceptance of innovation
District	<ul style="list-style-type: none"> - Accountability - Testing/evaluation 	<ul style="list-style-type: none"> - District consultants - Resources - Concrete/intensive training rather than one-shot workshops 	<ul style="list-style-type: none"> - Set testing/grading policy - Common, coherent set of goals for teacher professional development 	<ul style="list-style-type: none"> - iCelebrate the professionalism of teachers (McLaughlin, 1991, p. 77)
State (government and professional organizations)	<ul style="list-style-type: none"> - Accountability - Testing/evaluation - Teacher certification - Teacher networks 	<ul style="list-style-type: none"> - Special projects funds - Higher education and continuing education - Inservice provided by professional organizations 	<ul style="list-style-type: none"> - State-mandated tests - Curriculum Frameworks or Standards - Graduation requirements 	<ul style="list-style-type: none"> - Set standards, such as education for all children - Provide inservice on the frameworks or standards to clarify expectations
National (government and professional organizations)	<ul style="list-style-type: none"> - Testing - Teacher networks or iProfessional-area movement organizationi (Pennell & Firestone, 1996) 	<ul style="list-style-type: none"> - Teacher enhancement funds - Inservice by professional organizations - Annual conferences 	<ul style="list-style-type: none"> - National tests (e.g., NAEP) - Educational policy 	<ul style="list-style-type: none"> - Set standards (e.g., Bench- marks for Science Literacy (AAAS, 1993); Professional Standards for Teaching Mathematics (NCTM, 1991))

Integrated Assessment System as a Medium for Teacher Change

Level of Success with Implementation of Local Assessment Moderation	Leadership	Institutional Support	Teacher Proximity and Collaboration
Very High (ADC 4)	Strong leadership of ADC director. Former SEPUP teacher functioned as moderation facilitator.	Single district with district administrator as ADC director. Principals supported teacher release time. Efforts parallel State mandates for change.	All four teachers in the same district, but at different schools. Teachers work on other projects together. Teachers communicate via electronic mail or telephone.
High (ADC 3)	Teacher leader in place of ADC director who took on other duties.	Two schools in a single district. Principals supportive of staff development efforts.	Teachers paired at the two schools. Beginning teachers were supported by experienced teachers.
Moderate (ADC 2)	Two districts involved. Two teachers co-facilitated moderation meetings. The ADC director missed a few meetings due to other district obligations.	Districts emphasizing assessment as staff development focus. Have several federally-funded science education reform projects in the State.	Teachers who co-facilitated were accepted in this shared role by the other teachers. Teachers called others in the same district between meetings (otherwise long distance).
Limited (ADC 1)	Five teachers from five different districts. Limited by the fact that the ADC director was also a classroom teacher.	Varying district mandates diluted level of success. Principals' support varied, and there was no apparent district support. State drafting standards for science (K-12).	Limited somewhat by physical distance. Limited collaboration; called someone if a problem arose. Voice mail established by ADC director to provide weekly updates, but was not used by all teachers.