DOCUMENT RESUME

ED 405 346 TM 026 120

AUTHOR	Tittle, Carol Kehr; Pape, Stephen
TITLE	A Framework and Classification of Procedures for Use in Evaluation of Mathematics and Science Teaching.
PUB DATE	18 Sep 96
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Classification; *Educational Change; Educational Practices; Elementary Secondary Education; *Evaluation Methods; Knowledge Base for Teaching; *Mathematics Instruction; Models; *Science Instruction; Teacher Attitudes; Teaching Methods; Test Construction; Test Use
IDENTIFIERS	Subject Content Knowledge

ABSTRACT

Alternative perspectives and practices for describing and documenting teaching practice and student learning relevant to classroom reform are identified, and a framework is provided for characterizing these process-focused instruments and indicators. Diverse descriptors were used to locate documents describing instruments and procedures used to evaluate mathematics and science activities, teaching methods, and related aspects of instruction. From nine selected instruments or sets of procedures a detailed framework was developed with the following attributes: (1) author's stated purpose; (2) subject matter; (3) classroom interaction; (4) types of student knowledge/expected learning outcome/cognitive processes; (5) teacher knowledge and beliefs related to teaching and learning science and mathematics; (6) methods and procedures; and (7) demographics collected. The framework provides a stimulus to examine the role of theory and reform visions in evaluation procedures. The instruments and procedures that were identified point out the need to re-examine the role of procedures used in evaluations, as well as their specific characteristics, for support of reforms in classroom teaching and learning. An appendix presents the classification framework. (Contains 1 figure and 36 references.)(SLD)

m1aere1r 9/18/96

A framework and classification of procedures for use in evaluation of mathematics and science teaching

Carol Kehr Tittle
Stephen Pape
Graduate School and University Center
City University of New York

Introduction

The role of theory in evaluation has been discussed in several contexts. One context is theories about evaluation and another context is theories about the "object" of evaluation, e.g., educational reform. It is in this latter context that Cook and Campbell (1979) and Lipsey (1994) consider the role of theory in evaluation, in strengthening causal interpretation in nonexperimental applied research. Cook and Campbell separate out and explicitly identify the issues of the construct validity of causes (treatments) as well as effects. The idea of construct validity of the treatment, the "black box" or treatment theory characterized by Lipsey, is that the causal analysis is "...strengthened by an explicit theory about the nature and details of the change mechanism through which the cause of interest is expected to produce the effect(s) of interest" (Lipsey, 1994, p.6).

Current educational reform efforts derive from changing perspectives on teaching and learning. And, in many educational research and evaluation efforts, the black box is the classroom. In science and mathematics reforms, perspectives on classroom processes and outcomes are stated in the standards' documents of the National Council of Teachers of Mathematics (1989, 1991, 1995) and of the National Research Council (1996). These reform documents draw on theories of knowledge construction and instruction that can be broadly characterized as developmental and apprenticeship in their orientation (Farnham-Diggory, 1994) or constructivist, emergent and sociocultural (Cobb and Yackel, in press). As a result, the documents propose different roles for teachers and students, changes in classroom interactions, different emphases in student understandings in problem solving and inquiry processes, as well as changes in the focus of subject matter.

As with theories about the object of evaluation (the reform efforts), the role of theories or models of evaluation are of concern. For example, evaluators are examining the "Emerging roles of evaluation in science education reform" (O'Sullivan, 1995), considering strategies for non-traditional program evaluation (Frechtling, 1995), and archiving case studies of mathematics and science teacher

preparation (reform) projects (Stake, et al, 1993; Trumbull, 1993a, 1993b). Examples of other uses of evaluation models in the science context are provided by Altschuld and Kumar (1995). In mathematics, the extensive documentation and assessment for the QUASAR project (e.g., Silver & Cai, 1993; Stein & Lane, in press) provides an example of a reform-based project with several suggestions for models of evaluation (discussed further below).

Evaluators and researchers are also involved in evaluations of major statewide and urban systemic reform efforts that draw on these changed perspectives on teaching and learning mathematics and science. In this instance, work on opportunity to learn (OTL) indicators and school delivery standards play a role in studies evaluating school and system-wide change. Porter (1991) described a model of school process indicators and their importance in monitoring and understanding the relationship between student performance and schools. He has also examined empirically the relationship of classroom process variables to changes in student opportunity to learn and achievement (Porter, 1993, 1995).

Evaluators of reform in mathematics and science are thus concerned with both theories of teaching and learning and theories or models for evaluation. In this paper we focus on the 'black box' of the classroom, in particular the classroom interactions of teachers and learners. First, we identify alternative perspectives and methods for describing and documenting teaching practice and student learning relevant to classroom reforms. Second, we provide a framework for characterizing these process-focused instruments and indicators. Third, we use examples of observation instruments and procedures to illustrate the framework. Finally, we consider the diverse examples for their implications for evaluation in support of reform.

<u>Procedures and framework</u>

A diverse set of descriptors were used to locate documents describing instruments and procedures used to describe and evaluate mathematics and science activities, programs, instruction, teacher education and evaluation, teacher and student discourse and interaction analysis, protocol materials and urban programs. Documents were selected if they emphasized: 1. a view of the student as an active participant in the learning process--problem solving and conducting scientific inquiries; 2. the teacher as facilitator of student development in a particular subject matter; and 3) teachers themselves actively developing and reflecting on classroom practice. Key ideas also used to identify documents were: mathematical problem solving in groups and hands-on science inquiry; communities of learners, shared agreements between teachers and students

about the nature of discourse on mathematical and scientific problems, about evidence, explanation, and justification; and an emphasis on communication about mathematical and scientific ideas.

We selected a group of nine instruments and/or set of procedures that represented different purposes for collecting information on teacher-student interactions and classroom processes. (The term "instrument" is used here as a generic category for what was, frequently, a set of procedures that included classroom observations or indicators of classroom processes.) The instrument purposes ranged from use in large-scale indicator studies, research, evaluation, and teacher professional development.

A detailed framework was developed drawing on the work of Porter and his colleagues for the Reform-up-Close Study (Smithson and Porter, 1993), the National Council of Mathematics standards documents (1989; 1991) and the National Science Education Standards (1996). The framework has the following major attributes:

I. Author's stated purpose
II. Subject matter
III. Classroom interactions
IV. Types of student knowledge/expected learning outcome/cognitive processes
V. Teacher knowledge and beliefs related to teaching and learning science/mathematics
VI. Methods and procedures
VII. Demographics collected

The framework is presented in Figure 1, Instruments and procedures for observing and evaluating mathematics and science reform classrooms: A framework for classification of observations and indicators. The framework with the detailed list of attributes is given in the Appendix. (See also note 1.)

The classifications in Figure 1 are based on the information provided in documents which varied in the level of information they provided. The major sources for each of the nine instruments is identified here by date and given in full in the references:

QUASAR (QUASAR, 1992; personal communication on coding and pattern analysis, Mary Kay Stein, March 15, 1996);

ESTEEM (Expert Science Teaching Educational Evaluation Model, Burry-Stock, 1995);

Young (Classroom Observation Protocol, in Young, Brett, Squires, & Lemire, 1995);

Authentic Pedagogy (School Restructuring Study, in Newmann, Marks, & Gamoran, 1995);

A & A-T Artzt & Armour-Thomas (Phase-Dimension Framework for Assessment of Mathematics Teaching, and Teacher Cognitions Framework, in Artzt & Armour-Thomas, 1995);

Forman (Forman, Stein, Brown, & Larreamendy-Joerns, 1995);

Porter (Reform Up Close study, in Porter, 1993; Smithson & Porter, 1993);

NAEP (National Assessment of Educational Progress, 1990, 1992);

CLAS (California Learning Assessment System, in Wiley & Yoon, 1995).

Classification description

The nine instruments are identified as observations or indicators. In general, observations means that there is an independent observer(s) of classroom processes. Indicators means that information is based on teacher self-reports of classroom processes, subject matter, materials, goals and objectives, and so on. Each of the attribute classifications for the nine instruments are briefly summarized below:

I. Author's stated purpose.

Five instruments are identified as developed in the context of research studies (QUASAR, Authentic Pedagogy, Artzt & Armour-Thomas, Forman and Porter). Two instruments are developed specifically for use in observations as part of teacher professional development (ESTEEM and A & A-T) and two others are potentially useful for teacher professional development (QUASAR and Forman). The three indicator/opportunity to learn (OTL) instruments are Porter, NAEP and CLAS. Two, CLAS and NAEP, have been used state-wide or at state and national levels of data collection. Young's Classroom Observation Protocol for inquiry-based science teaching and learning is the only one specifically identified for local evaluations.

Categories II.-V. attempt to encompass the specific mathematics and science reform "visions" that may be the focus of an instrument, that is, science as inquiry or mathematics as problem solving. Pedagogical emphasis, classroom interactions, and student expected learning outcomes are attributes that these instruments, grounded to varying degrees in theory, research, professional standards and practice, are in the process of trying to define and capture. We judged the extent to which these different

attributes are identified and included in particular instruments.

II. Subject matter.

Two observation instruments are specifically related to science (ESTEEM and Young), and four encompass mathematics (although Forman's discourse analysis can be done in science classrooms as well). Two of the indicator instruments (Porter and NAEP) are for both math and science; CLAS is for mathematics.

All provided for observations or indicators of activities and tasks, typically with some criteria. The most detailed criteria were available for describing tasks on the QUASAR project in terms of their mathematical cognitive demands of students and then student engagement at those levels. On the indicator instruments, teachers were asked to report on freqency of types of activities or tasks. A classification of pedagogical meaningfulness was entered for most of the observation instruments.

III. Classroom interactions.

Several instruments focused on (A) classroom presentations-representations (of concepts) that teachers used. For Porter the type of representation (graphical, concrete, etc.) was identified; for QUASAR the quality of the representation was also identified, and observers described the advantages and disadvantages of the representation. All instruments included (B) instructional practice descriptions, at variable levels of detail. A range of teacher centered and student centered instructional practices were listed for the indicator instruments (Porter, NAEP and CLAS). With three exceptions, the indicators and observation instruments provide for descriptions of student activities (C).

All of the observation instruments focus on (D), the interpersonal level of analysis, emphasizing a student centered-teacher facilitor pattern of interaction. Ratings on this quality of interaction are not evident in the indicator instruments. Similarly, another important characteristic of reform classrooms is the quality of (E) the discourse level of analysis (e.g., Cobb, Wood, & Yackel, 1993; Cobb & Yackel, in press; Cobb & Bauersfeld, 1995). Both QUASAR and Forman (and Forman, 1996) provide examples of discourse level analysis, with Forman using detailed codings of discourse.

Several of the instruments also include observations of (F), assessments (QUASAR, ESTEEM, Young), collect examples of student performance on assessments (Authentic Pedagogy), or ask teachers about assessment procedures (NAEP Mathematics). The instruments were not consistent in the extent to which

there were provisions for categories (G) management and administrative, instruction-related and (H) non-instruction, administrative, off-task.

IV. <u>Types of student knowledge/expected learning outcomes.</u>

The list of expected student outcomes varies from (A) facts, to (G) build and revise theory, develop proofs, build arguments, explanations, pose questions, hypotheses. All of the instruments included conceptual understanding; some instruments (Young, and the three indicators) included facts. The (A) facts and (C) basic procedures categories provide opportunities to contrast with the higher levels of cognitive outcomes. For some instruments (e.g., ESTEEM, Young, A & A-T) it was not possible to be sure whether the higher categories, (F) and (G), were included. There is a lack of common language across the instruments, and also a lack of examples to anchor many of the ratings on the observations as well as the indicators. Further, it is not clear whether teachers' interpretations of indicators would be what the instrument developers intended.

Several instruments provide for focused observations of students and evaluation of the quality of student performance (see also VI., methods and procedures). In particular, Authentic Pedagogy and QUASAR both included evaluation of samples of student work on performance assessment tasks independently of the observation process.

V. <u>Teacher knowledge and beliefs.</u>

In the first category (A), teacher knowledge of content and pedagogical knowledge are specifically described or rated by observers in QUASAR, ESTEEM, and Artzt and Armour-Thomas. NAEP includes teacher self-ratings on extent of content knowledge. The projects that examine teacher knowledge of national and local reform documents/curricula are Porter (in questions that ask about a range of influences on teaching) and CLAS. In the second category (B), teacher beliefs about reform and about teaching and learning are elicited in those instruments which include teacher interviews about their goals and aims for teaching and learning. These include QUASAR, Artzt and Armour-Thomas, and Forman (using QUASAR interviews).

VI. <u>Methods and procedures.</u>

This section examines the extent to which the instrument incorporates (A) Sources of classroom data, (B) Scoring/evaluation, and (C) Preparation/reporting. There is a wide range of sources of classroom data. QUASAR represents the most extensive documentation, with pre/post observation teacher interviews, three videotaped observations per teacher, classroom materials, observation

of a target student, and student group interviews in connection with the observed lesson. ESTEEM has a set of instruments in connection with this professional development program: pre- post-teacher observation interviews; a series of observations over time (videotaped where possible); teacher self-report questionnaires on classroom practices and assessment practices; student outcome assessment rubric; and student concept mapping rubric.

Young and Authentic Pedagogy used observers as raters, with no videotaping. Both NAEP and CLAS used paper and pencil teacher questionnaires. Porter studied several ways to collect data, including teacher logs which were compared with classroom observers' reports.

All nine instruments collect information on teachers, and several included separate student information (QUASAR, ESTEEM, Authentic Pedagogy, CLAS). NAEP also collects student performance information (not examined here).

In (B) instruments were examined for scoring procedures and evaluation. Procedures compared instruments as to whether they used (1) detailed coding categories; (2) defined rating scales (i.e., a brief description anchored several points on the rating scale); or (3) holistic ratings (i.e., 1-5 rating scale with a general standard or description). All of the indicator instruments had detailed coding schemes, as did QUASAR.

The evaluation category examined whether ratings/codes of individuals were compared to one another (i.e., among teachers, as with a distribution of 'scores'), were compared against a standard, and/or whether patterns of ratings or codings were examined. All of the observation instruments held the ratings against standards rather than comparative descriptions.

C. <u>Preparation/reporting.</u>

Evaluation instruments and procedures may "feedback" into reform efforts by communicating goals and values. This category examined, for example, whether teachers were prepared for observations and knew about the purposes of the observations. This was difficult to determine for many of the instruments, with the exception of those intended for teacher professional development. It appeared that any instrument which required on-site observations or extensive teacher participation (e.g., Porter) required detailed directions and contact with teachers. Similarly, feedback can occur through reporting to teachers or follow-up of observations, etc. The only extensive reporting is done in the context of instruments for teacher professional development (Artzt & Armour-Thomas; ESTEEM). QUASAR has

also reported some observation information to teachers (Stein & Smith, in press)

VII. Demographics collected.

Detailed information on schools, class characteristics, and teacher background was described primarily for the indicator studies of NAEP and Porter, as well as Authentic Pedagogy (School Restructuring Study). Grade levels of each of the studies/instruments are indicated also, with the majority of instruments used at middle and high school levels.

The summaries above do not do justice to the efforts involved in using several of the instruments in order to describe the quality of the academic experience of students. The two instruments for which this is key are QUASAR and Authentic Pedagogy. The value of the QUASAR theory-guided extensive documentation is that it permits re-entry to the data base for researchers interested in other levels of analysis (e.g., Forman et al., 1995; 1996). While it is not practical except on large-scale, well-funded reform efforts, it offers a model for case work even in local, small-scale evaluations. Authentic Pedagogy focuses specifically on evaluating the quality of classroom activities and student performance assessments (Newmann, Secada & Wehlage, 1995).

Implications for evaluation to support reform

In many evaluations of reforms there is a need to provide observations or indicators that can support the intended direction of change or "vision" of teaching and learning described in the mathematics and science standards. Thus, a first goal was to identify examples of instruments that are emerging to meet the challenge of describing and documenting teaching practice and student learning that is related to the reform visions for classrooms. A second goal was to provide a framework that would assist in describing the characteristics of such instruments. Based on the framework that evolved, we have identified two broad implications or concerns for evaluation to support reform.

The first concern is to specifically examine evaluation procedures for the feedback or communication provided to teachers about reform efforts. The second concern is with the potential effectiveness of evaluation to support reform. Both of these concerns arise from the usefulness of the formative role of evaluation to support reform and the importance of using evaluation procedures that have meaning to teacher participants as well as other stakeholders.

Evaluation procedures and feedback to teachers. With respect to the first concern, the design of evaluation

procedures to include reporting or feedback of observations to support reform requires fidelity between the vision of reform and the inferences and interpretations that are drawn by participants. This is particularly important for teachers and others involved in teacher professional development. Do these interpretations and any subsequent uses support the desired classroom practices? or, are there unintended (perhaps negative) consequences?

The instruments examined in Figure 1 illustrate the present diversity in approaches intended to describe processes and content in mathematics or science classrooms. The desire for fidelity between reform vision, instrumentation, and teacher interpretations suggests the importance of both consistency and interpretability in level of description for observations and indicators of "visions."

The observation instruments range from thorough documentation, highly focused on tasks and processes as part of a "vision," to instruments that provide a few ratings to define a classroom, to the indicators that rely on teacher interpretations of indicator statements. There is an inconsistent use of language to define observations and indicators over the set reviewed. As a result, evaluators need to compare prospective instruments in detail and consider their strengths and weaknesses for feedback and formative reporting to teachers, as well as for other intended uses.

The range in level of description suggests a need for evaluators to conduct research on the interpretations and uses teachers (and others) make based on evaluation data. Some instruments may provide more direct or transparent meanings and suggestions to teachers that support their efforts to change practice (e.g., see Stein & Smith, in press; Forman, 1995). Some instruments, particularly the indicators, may need to be piloted with teachers identified as expert and not-expert in the reform vision of classroom teaching to see if distinguishing patterns of responses can be identified. How do teachers interpret these indicator statements in think-aloud protocols? How close are their interpretations to classroom practice? In sum, to what extent does an evaluation instrument(s) support teacher reform efforts?

<u>Potential effectiveness of evaluation to support reform.</u> There are implications for evaluation to support reform in the various instruments we examined. These implications are in the use of the professional development instruments as models and in considering the development of alternative processes and procedures for use with teacher participants.

The teacher professional development instruments examined suggest that in some evaluations it might be feasible to use

them for teachers who want to participate over time in examining their own classroom practice. For evaluation purposes, these teachers would agree to make selected observations or videotapes available for independent observers to review. Research on teachers and alternative observation practices might inform evaluation procedures. Particularly for smaller-scale evaluation projects, an important criteria for evaluation procedures can be the educational function (and thus reform-supporting function) of the processes or procedures for participating teachers.

The context of the evaluation will have implications for what "instruments" evaluators can use. Extensive documentation on any substantial scale is costly, yet there are resulting benefits in using the data (videotapes and extensive observer notes) in multiple studies. For most evaluations, extensive documentation is not possible, and evaluators may combine several procedures with "light sampling" in each (see, for example, Huetinck, Munshin, & Murray-Ward, 1995). Research studies are needed to compare evaluation methods. As one example, Porter (1993) compared teacher logs and observer's records, finding substantial agreements. Studies to examine the agreement between observers and teacher logs focusing on specific processes that teachers are changing (or content being learned) would yield important information for evaluators. Building on such research and existing procedures might indicate that examples of extensive documentation for a few classrooms, and use of logs, indicator questionnaires, and student performance may be feasible alternatives to extensive documentation on a wider sample.

In summary, the framework we used to examine this set of instruments provides a stimulus to examine the role of theory and reform visions in evaluation procedures. The instruments and procedures we found suggest the need to re-examine the role of procedures used in evaluations, as well as their specific characteristics, for support of reforms in classroom teaching and learning.

Notes

Paper presented as part of a symposium, "Evaluating mathematics and science reform in school classrooms: The role of theories in frameworks for evaluation" (Carol Kehr Tittle, Chair) at the annual meeting of the American Educational Research Association, New York, NY, April 1996.

1. Figure 1 does not include the detailed lists of categories for several of the attributes. For example, the subject matter lists of topics are not provided here since the majority of instruments did not include specific topics.

Specific topics are typically recorded in observations, but specific lists of topics were provided only on the indicators instruments. Of these three, the most detailed was the work of Porter and his colleagues, since the coding was intended to include mathematics and science in grades 9-12.

References

Altschuld, J.W., & Kumar, D. (1995). Program evaluation in science education: The model perspective. New Directions for Program Evaluation, 65(Spring), 5-17.

Artzt, A., & Armour-Thomas, E.(1995). A model for studying the relationship between instructional practice and teacher cognition. Manuscript submitted for publication. Flushing, NY: Queens College of the City University of New York.

Burry-Stock, J.A. (1995). Expert science teaching educational evaluation model (ESTEEM), (1st Edition). Tuscaloosa, AL: Box 870231, Educational Research, University of Alabama.

Cobb, P., & Bauersfeld, H.(Eds.)(1995). The emergence of mathematical meaning: Interaction in classroom cultures. Hillsdale, NJ: Erlbaum.

Cobb, P., & Yackel, E.(in press). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. Educational Psychologist.

Cobb, P., Wood, T., & Yackel, E. (1993). Discourse, mathematical thinking, and classroom practice. In E.A.Forman, N.Minick, & C.A. Stone (Eds.), Contexts for learning: Sociocultural dynamics in children's development (pp. 91-119). New York: Oxford University Press.

Cook, T. & Campbell, D.T.(1979). Quasi-experimentation: Design & analysis issues for field settings. Chicago, IL: Rand McNally.

Farnham-Diggory, S. (1994). Paradigms of knowledge and instruction. Review of Educational Research, 64, 463-477.

Forman, E.A. (1996, April). Mathematics and talk: Evaluating discourse in reform classrooms. Paper presented as part of a symposium, "Evaluating mathematics and science reform in school classrooms: The role of theories in frameworks for evaluation" (Carol Kehr Tittle, Chair) at the Annual meeting of the American Educational Research Association, New York, NY.

Forman, E., Stein, M.K.,Brown, C., & Larreamendy-Joerns,J. (1995, April). The socialization of mathematical thinking. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Frechtling, J.A. (Ed.). (1995, January). Footprints: Strategies for non-traditional program evaluation. Rockville, MD: Westat.

Huetinck, L., Munshin, S., & Murray-Ward, M.(1995). Eight methods to evaluate and support reform in the secondary-level mathematics classroom. Evaluation Review, 19(6), 646-662.

Lipsey, M.W.(1994). Theory as method: Small theories of treatments. In L.B.Sechrest & A.G.Scott (Eds.), Understanding causes and generalizing about them. New Directions for Program Evaluation, 63 (Fall), 5-25.

National Assessment of Educational Progress (1990). Science teacher questionnaire, Grade 8. Washington, DC.: USOE.

National Assessment of Educational Progress (1992). Math teacher questionnaire, Grade 4. Washington, DC: USOE.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1991). Professional standards for teaching mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1995). Assessment standards for school mathematics. Reston, VA: Author.

National Research Council (1994, November). National science education standards. Washington, DC: National Academy Press.

National Research Council (1996). National science education standards. Washington, DC: National Academy Press.

Newmann, F.M., Marks, H.M., & Gamoran, A. (1995, April). Authentic pedagogy and student performance. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Newmann, F.M., Secada, W.G., & Wehlage, G.G.(1995). A guide to authentic instruction and assessment: Vision, standards and scoring. Madison, WI: Wisconsin Center for Education Research.

O'Sullivan, R.G. (Ed.). (1995). Emerging roles of evaluation in science education reform [Special issue]. New Directions for Program Evaluation, 65 (Spring).

Porter, A.C.(1991). Creating a system of school process indicators. Educational Evaluation and Policy Analysis, 13(1), 13-29.

Porter, A.C.(1993). Defining and measuring opportunity to learn. Paper prepared for the National Governors' Association. Madison, WI: Center for Policy Research in Education.

Porter, A.C.(1995, April). Standard setting and the reform of high school mathematics and science. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

QUASAR (1992, September). QUASAR Documentation: Classroom Observation Instrument. Pittsburgh, PA: QUASAR Project, Edward A.

Silver, Director, Learning Research and Development Center, University of Pittsburgh.

Silver, E.A., & Cai, J. (1993, June). Schemes for analyzing student responses to QUASAR's performance assessments: Blending cognitive and psychometric considerations. Paper presented at the meeting of the American Educational Research Association, Atlanta, GA., April 13, 1993.

Smithson, J.L., & Porter, A.C. (1993, April). Measuring classroom practice. Paper presented at the meeting of the American Educational Research Association, Atlanta, GA.

Stake, R., Raths, J., St.John, M., Trumbull, D., Jenness, D., Foster, M., Denny, T., & Easley, J. (1993). Teacher preparation archives: Case studies of NSF-funded middle school science and mathematics teacher preparation projects. Urbana, IL: University of Illinois, College of Education.

Stein, M.K., & Lane, S. (in press). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. Educational Research and Evaluation.

Stein, M.K., & Smith, M.S. (in press). Mathematical tasks as a framework for reflection. Mathematics Teaching in the Middle School.

Trumbull, D. (1993a). Boston University: Project PROMISE: Program for Middle School Science Education. In R. Stake et al., Teacher preparation archives (pp. 119-135). Urbana, IL: University of Illinois, College of Education.

Trumbull, D. (1993b). Conceptualizations of science and mathematics education in and for the middle schools. In R. Stake, et al., Teacher preparation archives (pp. 245-256). Urbana, IL: University of Illinois, College of Education.

Wiley, D.E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). Educational Evaluation and Policy Analysis, 17(3), 355-370.

Young, M.J., Brett, B., Squires, S., & Lemire, N. (1995, April). Symposium: A new observation tool for looking at inquir-based teaching and learning. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

(mlaerref 9/19/96)

Figure 1

Instruments and procedures for observing and evaluating mathematics and science reform classrooms: A framework for classification of observations and indicators.

	Observational Tools						OTL Indicators		
	QUASAR	ESTEEM	Young	Auth Ped	A & A-T	Forman	Porter	NAEP	CLAS
I. Author's Stated Purpose									
A. Research	X			X	X	X	X		
B. Program Evaluation (Local, distr, state, national)	ANY LEVEL		LOCAL	ANY LEVEL			STATE & LOCAL	NATIONAL	STATE
C. Teacher/Professional Dev't	X	X			X	X (?)			
II. Subject Matter									
A. Science as inquiry		X	X				X*	X* (1990)	
B. Math as problem solving	X			X	X	X	X*	X* (1992)	X (1993)
C. Activity/task/criteria	X	X	X	X	X	X	X	X	X
D. Pedagogical meaningfulness	X	X	?	X	X	X			
III. Classroom Interactions									
A. Classroom Presentation - Representation	X	X			X	X	X		
B. Instructional Practice/Desc.	X	X	X	X	X	X	X	X	
C. Student Activities (work sheets, presentations) groups, pairs	X		X			X	X	X	X
D. Interpersonal Level analysis - student centered-teacher facilitator vs teacher centered	X	X	X	X	X	X			
E. Discourse Level Analysis - S-S and T-S	X					X			
F. Assessment	X	X	X	X				X (MATH)	
G. Management/administrative-instruction-related	X				X	X	X	X	
H. non-instr/admin/off-task	X		X				X		

\TITTLE\AERA\CLASTAB2.DOC\7-Sep-96
Classification table for Observation Tools

	Observational Tools						OTL Indicators		
	QUASAR	ESTEEM	Young	Auth Ped	A & A-T	Forman	Porter	NAEP	CLAS
IV. Types of student knowledge/expected learning outcome/Cognitive processes									
A. Facts (memorization, definitions, equations)			X				X	X	
B. Conceptual understanding	X	X	X	X	X	X	X	X	
C. Procedures									
1. Collect data (e.g. observe, measure)			X				X	X	
2. Order, compare, estimate, approximate				X			X		X
3. Perform procedures: execute algorithms/routine procedures, classify	X		X		X		X	X	X
4. Communicate Understanding in different forms	X	X	X	X	X	X		X	X
D. Solve routine problems, replicate exper/proofs							X	X	
E. Interpret data, recognize patterns and relationships	X	X		X	X		X		X
F. Recognize, formulate, and solve novel problems/design experiments	X	X		X			X		X
G. Build and revise theory, develop proofs, build arguments, explanations, pose questions, hypotheses	X			X		X	X		X
V. Teacher Knowledge and Beliefs related to teaching and learning science/math									
A. Teacher Knowledge									
1. Content - science and math	X	X			X			X	
2. Content - pedagogical knowledge	X	X			X			X	
3. Reform documents							X		X
B. Teacher beliefs: reform, teaching and learning epistemology	X				X	X			

\TITTLE\AERA\CLASTAB2.DOC\7-Sep-96
Classification table for Observation Tools

	Observational Tools						OTL Indicators		
	QUASAR	ESTEEM	Young	Auth Ped	A & A-T	Forman	Porter	NAEP	CLAS
VI. Methods and Procedures									
A. Source of classroom data - print/materials, audio, video, observ, intervw (pre/post), logs, portfolios, questionnaires, student products	materials VT obs (3) pre/post interview	materials pre-intvw VT obs self-rprt	materials obs pre/post interview	quest. Assess tasks obs (4)	pre/post interview VT obs	materials VT & notes of obs (3) pre/post interview	materials T. Logs obs T. quest.	quest. self-rprt demogr	T. self-rprt & std. assess survey
1. Teacher	X	X	X	X	X	X	X	X	X
2. Student	X	X	?	X					X
B. Scoring/Evaluation									
1. Procedures - (1) defined coding scheme/categories; (2) defined rating scale; (3) holistic rating	1	2	3	2	2	1	1	1	1
2. Evaluation - comparison vs others, standard, or pattern analysis	standard & pattern	standard	standard	standard	standard	pattern	comp. vs others & pattern	comp. vs others	comp. vs others
C. Preparation/Reporting Procedures/Description									
1. Preparation prior to observation - none vs some	X	X	X	?	X	X	X		?
2. Report to teachers/follow-up - none or elaborate	X	X	?	?	X	?	?	?	?
VII. Demographics Collected									
A. School/class/teacher	X			X		X	X	X	
B. Grade Level	6-8	7-12 (?)	K-12	4-5, 7-8, 9-10	9-12	6-8	9-12	8	4, 8, 10

\TITTLE\AERA\CLASTAB2.DOC\7-Sep-96
Classification table for Observation Tools

APPENDIX

Classification framework: Instruments and procedures for observing and evaluating mathematics and science classroom interaction processes

I. Author's stated purposes for instrument/procedure
   A. research
   B. evaluation/indicators/OTL (local, state, national)
   C. teacher professional development

II. Subject matter
   A. science
      1. science as inquiry
      2. topics and domains
         physical science, including chemistry and physics; life science, earth and space science
         general topics: science and technology, history and nature of science, unifying concepts & processes

   B. mathematics
      1. mathematics as problem solving and understanding
      2. topics and domains
         estimation, number sense/theory/mathematical structure & numeration, arithmetic operations, geometry, measurement, statistics & probability, fractions & decimals, algebra, functions, trigonometry, discrete mathematics, calculus

   C. activity/task criteria
      Science:
      central event/phenomenon in the natural world
      central scientific idea and organizing principle (explanatory power, fruitful, investigation,applies to common everyday experiences, links to meaningful learning experiences, developmentally appropriate for diverse students--prior experiences, etc.)
      Mathematics:
      significant mathematics
      developmentally appropriate(experience, interest, diverse students, difficulty level, sequencing, and motivational strategies

   D. pedagogical meaningfulness
      1. fosters mathematical/scientific understanding, and communication
      2. develops beliefs about mathematics/science as an ongoing human activity
      3. builds connections, interest, student curiosity, and speculation
      4. requires problem (question) formulation, problem solving/gathering evidence, mathematical reasoning/ proposing scientific explanations, extended problem exploration/scientific investigation

III. Classroom interactions
A. general content presentation/ representations
1. exposition-verbal & written
2. pictoral models
3. concrete models (e.g.,manipulatives)
4. equations/formulas (e.g., symbolic)
5. graphical
6. laboratory work
7. field work
B. instructional practice (description)
C. student activities (work sheets, presentations) groups, pairs

D. interpersonal level analysis
1. student centered
student centered-teacher facilitator
2. teacher centered
setting up task and conditions
surveying answers
asking questions
summarizing

E. discourse level analysis
1. initiations -requests for answers; requests for explanations
2. responses --state answer; state explanation
3. reconceptualization --restatement; expansion; rephrasing; evaluation

F. assessment
G. management/administrative instruction-related
H. non-instructional/administrative/off-task category

IV. Type of student knowledge/ expected learning outcome/ cognitive processes
A. facts (memorizing facts,definitions,equations)
B. conceptual understanding
C. procedures
1. collect data (e.g., observe, measure)
2. order, compare, estimate, approximate
3. perform procedures: execute algorithms/routine procedures (including factoring), classsify
4. communicate understanding, use different representations--symbolic, written, oral
D. solve routine problems (including word problems), replicate experiments, replicate proofs
E. interpret data, recognize patterns and relationships
F. recognize, formulate, and solve novel problems/ design experiments
G. build and revise theory, develop proofs, build arguments, explanations, pose questions, conjecture, hypotheses

V. Teacher knowledge and beliefs related to teaching and learning science/ mathematics
   A. teacher knowledge
      1. content --science or mathematics
      2. content --pedagogical knowledge, e.g., represention s (science and mathematical concepts)
      3. reform documents
   B. teacher beliefs/cognitions, reform-related

VI. Methods and procedures
   A. source of classroom data: print/materials, audio, video, observation, pre/post interviews, logs, portfolios, questionnaires, student products
      1. teacher
      2. student interview/questionnaire/tasks
   B. scoring/evaluation
      1. scoring procedures
         1 defined coding scheme/ categories
         2 defined ratings: brief description/rating scale anchored and points "defined"
         3 holistic, e.g., 1-5 scale, ends may be anchored
      2. evaluation
         compared with others vs. defined standard, and/or pattern analysis
   C. preparation/reporting:procedures/description
      1. preparation: none-> at least some
      2. report to teacher: none->elaborate (interpretive reports, videotape of exemplars, etc.)

VII. Demographics collected
   A. school, class, teacher
   B. grade level

N.B.Use of any framework/procedures requires an understanding of the context within which teaching and learning occur; might be provided by a description of the context of the observation/evaluation including:
demographics: type of school; grade level; N students;
activity/task descriptions (running log)
(task, content, what students doing, what teacher is doing, materials used, assessments,
context: teacher goals (teaching & learning)

References: This framework draws on the National Science Education Standards (1994), NCTM standards 1989, 1991, and Forman et al (1995); content categories from NSES (1994) & NCTM (1989) & Smithson & Porter (1993)

sciev1r 9/18/96

AERA April 8-12, 1996

TM026120

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A framework and classification of procedures for use in evaluation of mathematics and science teaching	
Author(s): Carol Kehr Tittle and Stephen J. Pape	
Corporate Source: Graduate School and University Center City University of New York	Publication Date: September 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[XX] Sample sticker to be affixed to document

Check here

Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here

Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Carol Kehr Tittle	Position: Professor
Printed Name: Carol Kehr Tittle	Organization: Graduate School and University Center - City University of New York
Address: PhD Program in Educational Psychology CUNY - Graduate School 33 West 42 Street New York, NY 10036	Telephone Number: (212) 642-2254
	Date: October 9, 1996