MH
M.A. HENRY
CONSULTING, LLC

Meeting the Challenge of STEM Classroom Observation in Evaluating Teacher Development Projects

A Comparison of Two Widely Used Instruments

Martha A. Henry
mahenry@mahenryconsulting.com

Keith S. Murray
keithsmurray@mahenryconsulting.com

Katherine A. Phillips
kathyphillips@mahenryconsulting.com

November 30, 2007

Introduction

Federally funded education projects committed to the enhancement of teacher knowledge and skills and student achievement in science, technology, engineering and mathematics (STEM) benefit from a look inside the classroom (Knapp, 2003; Medina, et al., 2000). Documentation of classroom implementation is necessary to assess performance, discern changes in practice, and analyze the effects of a project's activities. Without observation, the causal or correlative relationships between a project's treatment efforts and subsequent teacher performance are more difficult to determine (Shavelson & Towne, 2003).

However, the process of observing classrooms is costly in terms of personnel and time. The stakes involved in ensuring comparability and relevance of observations across settings and observers requires that evaluators and others must select observers and instruments wisely.

Instrument selection therefore is a challenge evaluators must confront with each project that requires documentation of classroom implementation. The need exists for an instrument that is aligned to the philosophical underpinning of the project, reflects reform teaching, requires a minimum of training, is available without cost, provides complete and valid data, and is not cumbersome to use. In addition, the instrument must reflect the specific objectives of the project if the indicators are to provide data that will accurately reflect project outcomes.

This is a tall order for any one instrument. The added need for an instrument that reflects a particular content, such as mathematics or science, or a particular focus, such as the implementation of technology, further narrows viable options for evaluators.

The option of creating a customized observation instrument for a particular project may seem attractive. However, few projects are funded with adequate resources to support such instrumentation requirements as repeated validity and reliability testing. The use of ad hoc, untested instruments may be common but is unacceptable. Grafting project-specific items onto existing protocols produces the same instrumentation issues. Such approaches produce results whose validity and reliability must be questioned.

Given these factors, plus the cost of training and supporting observers, many evaluators have eliminated classroom observations altogether as a means of confirming implementation of projects. In such cases, evaluators have defaulted to teacher self-reports, with their inherent limitations (Baron, 2006). Such a choice is more indicative of an inability to surmount the challenges of observation than confidence in the efficacy of self-reporting.

What instruments, then, are available for use? In the process of investigating observation instruments for four Mathematics and Science Partnerships (MSPs), the authors saw two tools emerge as the most frequently recommended and used: the *Inside the Classroom Observation and Analytic Protocol* (ITC COP) (Horizon Research, Inc., 2003) and the *Reform Teaching Observation Protocol* (RTOP) (Arizona Collaborative for Excellence in the Preparation of Teachers [ACEPT]) (Sawada, et al., 2000). With varying levels of satisfaction, the authors have used both instruments in evaluating MSPs. As the leading instruments available, these two instruments are the focus of the comparison presented in this paper. The instruments can be

downloaded at http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/index.htm for the RTOP and http://www.horizon-research.com/insidetheclassroom/reports/looking/appendixa.pdf for the ITC COP.

Because of the importance and challenge of classroom observation, closer examination of the characteristics of these instruments and their usefulness to evaluators as they enter and observe science and mathematics classrooms is needed. This paper examines the RTOP and ITC COP instruments based on the author's experiences and analysis and addresses these questions:

1. What are the underlying assumptions about reform teaching in these documents?
2. By whom should each instrument be used?
3. Which instrument provides ease of use to the rater while meeting evaluation needs?
4. Which instrument provides data that represents actual classroom implementation?

Methods

The present paper represents the first phase of a two-phase analysis conducted by the authors. This first phase represents an analysis of instruments, condensing to the two instruments under consideration. Their philosophical underpinnings as claimed (when available) or inferred (when not available), the testing data available, and the research supporting the instruments were reviewed. Indicators from each instrument were examined for comparability and for weighting in scoring rubrics.

The authors researched available documentation supporting the RTOP and ITC COP. Seminal documents were obtained from the developers at the Arizona State University and Horizon Research, Inc. These documents were examined for evidence of the bases on which the instruments were developed, the developers' definition of reform teaching, the purpose for development, and any subsequent testing for validity and reliability.

Following research into the background of the instruments, analysis focused on the indicators in the rating sections of the instruments. Categories of each instrument were examined for congruence. Given some degree of congruence among these categories found between instruments, specific items within the categories were aligned.

Issues deriving from categorical misalignment arose because some items did not appear in similar categories. A closer examination of items across categories was undertaken, with matches resulting from three approaches: (1) items judged to be identical or to indicate the same behavior, (2) overarching items in one instrument (RTOP) judged to subsume items in the other instrument (ITC COP) and (3) overarching items in the second instrument (ITC COP) judged to subsume items in the first one (RTOP).

Once these alignments were selected, the resulting information was sent to two evaluators experienced with both instruments for their review and confirmation of the alignment. After several adjustments, the final set was reviewed by the three evaluators for agreement and the analysis was undertaken.

The final comparison of the two instruments was then sent to five experts familiar with the two instruments and who had used them in the field. These expert reviewers included an MSP evaluator, an evaluator of state MSP projects, a project investigator on an MSP project, a university professor holding appointments in education and a STEM content area, and a field staff person who works with an evaluation firm.

The second phase of the analysis will involve a more systematic look at the indicators. Both instruments are currently being used by field staff as they observe approximately 20 teachers for an ongoing MSP evaluation. Each teacher will be rated by the same rater using the two instruments. Statistical analysis of indicators and domains across instruments will be conducted for later publication.

Results

The Instruments

The *Inside the Classroom Observation Protocol* (ITC COP) and the *Reform Teaching Observation Protocol* (RTOP) are two classroom observation instruments in widespread use among evaluators of mathematics and science professional development projects. They also are used in intra-project assessments and as tools for teacher coaching. Both instruments provide indicators within domains for observers to consider and rate as they observe classrooms. They contain supplemental pages for recording observation notes, demographic data, and classroom environment information. Both recommend a formal course of training, with interrater reliability testing prior to use.
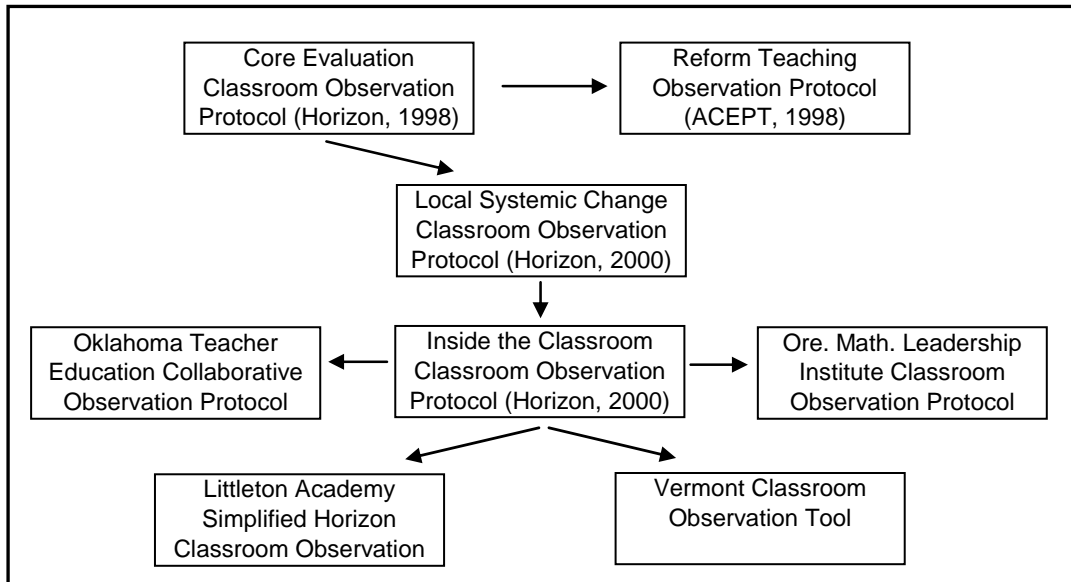
Each instrument has strengths and weaknesses for STEM evaluative purposes, including the extent to which metrics are developed and available, options for recording all relevant classroom conditions, the type of data produced, and ease of use. These features will be discussed further as each instrument is examined.

The RTOP was developed in 1998 and is the older of the two instruments, although it is partially based on antecedents of the ITC COP. Widely used, the RTOP does not appear to have generated other widely used instruments similar either in content or organization.

In contrast, several instruments have been derived from the ITC COP. These derivations exhibit several similarities and differences. Those instruments directly developed from the ITC COP, as indicated either through descriptions of the instruments by their developers or by comparison of indicators and format, include the *Core Evaluation COP* for the Centers for Enhancement of Teacher Preparation (CECOP), the *Vermont Classroom Observation Tool* (VCOT), the *Littleton Academy Simplified Horizon Classroom Observation Protocol*, the *InSITE Fellows Teaching Observation Protocol*, and the *Oregon Mathematics Leadership Institute Classroom Observation Protocol* (OMLI COP).

Figure 1 shows the authors' understanding of the development course and relationships among selected observational instruments.

Figure 1. Derivations of classroom observation instruments

```
┌─────────────────────────────────────────────────────────────────────────┐
│   ┌──────────────────────┐                    ┌──────────────────────┐    │
│   │   Core Evaluation    │                    │   Reform Teaching    │    │
│   │ Classroom Observation│ ─────────────────▶ │  Observation Protocol│    │
│   │ Protocol (Horizon,   │                    │     (ACEPT, 1998)    │    │
│   │        1998)         │                    │                      │    │
│   └──────────────────────┘                    └──────────────────────┘    │
│              │                                                             │
│              ▼                                                             │
│        ┌──────────────────────┐                                           │
│        │ Local Systemic Change│                                           │
│        │ Classroom Observation│                                           │
│        │ Protocol (Horizon,   │                                           │
│        │        2000)         │                                           │
│        └──────────────────────┘                                           │
│                   │                                                        │
│                   ▼                                                        │
│ ┌────────────────┐  ┌──────────────────────┐  ┌──────────────────────┐   │
│ │Oklahoma Teacher│◀─│  Inside the Classroom│─▶│ Ore. Math. Leadership│   │
│ │Education Collab│  │ Classroom Observation│  │  Institute Classroom │   │
│ │orative         │  │ Protocol (Horizon,   │  │ Observation Protocol │   │
│ │Observation     │  │        2000)         │  │                      │   │
│ │Protocol        │  └──────────────────────┘  └──────────────────────┘   │
│ └────────────────┘      ╱          ╲                                      │
│            ┌──────────────────┐  ┌──────────────────────┐                 │
│            │Littleton Academy │  │  Vermont Classroom   │                 │
│            │Simplified Horizon│  │   Observation Tool   │                 │
│            │Classroom         │  │                      │                 │
│            │Observation       │  │                      │                 │
│            └──────────────────┘  └──────────────────────┘                 │
└─────────────────────────────────────────────────────────────────────────┘
```

Recent instruments represent various adaptations and changes to the original ITC COP. Additions to or deletions of the pages supplemental to the rating scales, changes to the cover page, and changes to the number and wording of rating indicators within the domains are among the refinements that appear.

Since the purpose of this study is the comparison of the two most popular instruments used within the MSPs, a detailed analysis of each of these adaptations will not be conducted here. Instead, the RTOP will be the primary instrument for comparison because of its use by MSP projects that have not developed unique instruments or that do not use the ITC COP.

Inside the Classroom Observation and Analytic Protocol (ITC COP) (2000)

ITC COP Organization

The ITC COP, along with the Local Systemic Change through Teacher Enhancement Observation Protocol (LSC COP) (2000), seems to be among the early instruments of many available to observers of teachers of mathematics and science. The instrument is designed to assess the quality of the design and implementation of mathematics and science lessons (Weiss, et al., 2003). An examination of 12 observation instruments (see Appendix) shows that of the eight directly applicable for observations of mathematics and science reform teaching, six were directly derived from or show characteristics in indicator language or rating options from the ITC COP. The *Local Systemic Change Classroom Observation Protocol* (2000) and the ITC COP share almost duplicate item language and show major changes only in the supporting pages, which are directed toward the LSC goals for the LSC COP.

A three-day, for-fee, intensive training is available for the ITC COP from Horizon Research. The training involves the analysis of several classroom videos with interrater reliability determined among the group. The training manual is provided at that training.

The full ITC COP is a lengthy document of 22 pages, designed with three parts. Part One includes six sections:

Section A, Basic Descriptive Information, describes the teacher, the subject and the class being observed.

Section B, Purpose of the Lesson, is a brief teacher-derived report of the purpose of the lesson and a descriptive check list of the focus of the lesson.

Section C, Lesson Ratings, comprises indicators used for actual rating of the lesson implementation. Each of the indicators in the four domains are rated on a scale of 1 = "Not at all" to 5 = "To a great extent." In addition, there is an option to rate "Don't know" or "N/A." The domains rated are Design, Implementation, Mathematics/Science Content and Classroom Culture. The ratings section (Section C) will be the focus of examination for this study.

Section D, Lesson Arrangements and Activities, charts the number of minutes spent on various parts of the lesson and the percent of instructional time allotted.

Section E, Overall Ratings of the Lesson, provides six indicators for which the lesson may have had a negative, mixed/neutral, or positive effect. The second part of that section is the Capsule Rating of the Quality of the Lesson consisting of five levels with descriptions that characterize a lesson. The rater most closely matches the lesson observed to these descriptions.

Section F, Descriptive Rationale, consists of two pages of narrative including a general narrative of the Lesson and Lesson Features.

Part Two, Influences on the Selection of Topics/Instructional Materials/Pedagogy Used in Planning This Lesson, examines several areas that can influence a lesson including Policy and Support Infrastructure, the Physical Environment, Instructional Materials, Student Characteristics, The Teacher and the interaction of all of these in Section B, Why This Lesson.

Part Three, Putting It All Together, provides an opportunity to summarize the lesson in a brief story, a tag line, and overall assessment of quality, and to provide any additional information that would inform the evaluation. There is an option to recommend this lesson as a vignette study for Horizon Research in their study of classrooms.

The protocol is a comprehensive document, promoting the examination of in-class implementation, planning, influences on curriculum and teaching, and the supporting environment, including materials support. When focusing on each of these aspects, the time required to complete the document may be warranted. However, its length may be one reason other instruments have been developed with similar characteristics that omit some of the detail and documentation of outside influences.

ITC COP - Definition of Reform Teaching

The set of ITC COP instruments from Horizon was developed based on national science standards from the *National Science Education Standards* (National Research Council, 1996) and the national mathematics standards from the National Council of Teachers of Mathematics documents *Curriculum and Evaluation Standards for School Mathematics* (1989), the *Professional Teaching Standards for School Mathematics* (1991), and the *Assessment Standards for School Mathematics* (1995). The separation of the indicators into the four domains of Design, Implementation, Mathematics/Science Content, and Classroom Culture supports the belief that reform teaching occurs in or influences each of these domains.

The Design section reflects lessons that incorporate the inclusion of activities, tasks, roles and interactions leading to a collaborative approach to learning. Included also are indicators of careful planning and time for concluding the lesson, including an overt plan for sense-making.

Implementation indicators reflect the importance of a teacher's ability to carry out the design confidently, incorporating a pace that adapts to students' levels of understanding, and questioning applied to develop conceptual understanding.

The Mathematics/Science Content section places importance on significant, accurate content appropriate to the development of the student, including appropriate levels of abstraction, presented contextually, and promoting conceptual development and sense making. Many of these indicators could be categorized as pedagogical content knowledge indicators.

Classroom Culture indicators describe a classroom where there is respect for students and teachers, ideas and rigor are valued, and collaborative work is the norm. The inferred definition of reform teaching puts the student in the center of the planning and implementation. It is an active learning process where rigorous content is presented in a way that challenges students to work together to make sense of the content, building their conceptual understanding through interactions with the content and other learners, including teachers and other students.

ITC COP - Data Collection

If the complete instrument is used, information collected by observers includes demographic data on the teacher and brief description of the course, including numbers of males and females (Section A). No race or ethnicity information is requested. Preliminary information about the lesson, obtained from the Horizon Teacher Interview Protocol (Horizon Research, Inc., 2000), is to be used to assist the observer in completing the next section (Section B). It includes the purpose and focus of the lesson as described by the teacher. The third section (Section C), as noted, includes the ratings of the four domains with an associated synthesis rating and an explanation, the overall ratings and the capsule rating. This third section will be the focus of the comparison with the RTOP. The next section (Section D) provides for a narrative description of the lesson with prompts. Lesson features provide another form of analysis of what the lesson contained in terms of instructions pedagogy and student activities.

Part Two of the instrument focuses on influences on the selection of topics/instructional materials/pedagogy used in planning the lesson. This information is obtained from the teacher in

a post-lesson discussion. This section is unique to the ITC COP and does not appear in any of the adaptations. Areas examined that may influence teacher decision-making about instruction include policy and support infrastructure, the physical environment, instructional materials, student characteristics (intellectual level, special needs, ESL, cognitive abilities), and specific teacher characteristics such as tenure in teaching, professional training, and teacher attitudes and beliefs. Part Three synthesizes why the lesson was chosen, designed and taught by the teacher, as determined by the rater and based on all of the data collected. The last section is for Horizon's use in determining the usefulness of this lesson as a vignette for their research purposes.

Section C, comprising the individual ratings by domains, is the source of most of the quantitative data from this instrument. The 30 indicators are unequally divided among the domains. There are nine indicators in the Design section, seven in the Implementation section, nine for assessing Mathematics/Science Content, and five for Classroom Culture. These data are often analyzed in various ways depending upon the intent of the evaluation.

One overall score can be obtained for each teacher, sub-groups of teachers, or the total number of teachers being evaluated. Pre/post scores with interventions may be analyzed for mean gain or normalized mean gain for each of the previously mentioned groups. In addition, domains can be analyzed for these groups separate from and compared to the total score. In this instrument, a synthesis rating for each domain is provided. These synthesis ratings on a 5-point Likert scale have end points labeled according to the domain. The 1 rating indicates that the lesson is "…not at all reflective of best practice…" in that domain, and the 5 rating indicates that the lesson is "…extremely reflective of best practice in …" the domain. These ratings may be compared from pre- to post administration of the instrument to examine teacher change.

Because these are Likert ratings, caution should be taken during the analysis. Likert data are ordinal data so only nonparametric statistical analysis is appropriately administered unless one can be assured that the intervals between numbers are equal.

This instrument, though comprehensive, requires a considerable amount of time to fully complete, including pre/post in-depth teacher interviews and a classroom observation in addition to the time required to synthesize the information and make sense of it as a unique lesson. Though an admirable goal, there is seldom enough time to utilize all of the tools within this instrument. That could be the reason there are so many variations of the instrument in use today.

ITC COP – Research

One reference to research on the validity and reliability of the Inside the Classroom Observation Protocol is available from Horizon Research, Inc. (Weiss, et al., 2003)

> "For *Inside the Classroom*, the study coordinators adapted the classroom observation instrument originally developed by HRI as part of the core evaluation of National Science Foundation's Local Systemic Change initiative (p. ix)…As part of the core evaluation of NSF's Local Systemic Change Initiative, Horizon Research, Inc. (HRI) field-tested, revised, and demonstrated the reliability of a classroom observation instrument for assessing the quality of the design and implementation of mathematics and science lessons" (p. 2).

However, no metrics were provided in this article nor have they been provided when requested from Horizon. Two studies have been conducted on the Local Systemic Change Observation Protocol, which, though not identical in two domains (Design and Implementation), has identical items in the other two (Mathematics/Science Content and Classroom Culture). For these last two domains, items and individual wording within the items are identical. Horizon (2000) reports that the internal consistency for item sets within these two domains is .93 (Chronbach's α) for both Content and Culture, both with and without the synthesis rating. Although the other two domains show equally high ratings, the addition of two additional items in each category prevents any inferences about the ITC COP categories.

One study released by Horizon Research on the predictive validity of the LSC COP (Banilower, 2005) is inconclusive about the ability of mean observation scores on three observations of 16 teachers to predict student achievement scores on mid-year mathematics assessment. Despite similarities across instruments, no inferences can be made about the predictive validity of the ITC COP from this study.

These two studies are cited because of an apparent lack of research attention to the ITC COP, even though it is used widely in Math/Science Partnership projects. Because of the research available on the RTOP, comparison of this instrument and the RTOP seems useful.

<u>Reform Teaching Observation Protocol (RTOP)</u>

RTOP Organization
The RTOP was created as an instrument to use when observing science classrooms. Its contents were published as a technical report NO. IN00-1 by the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) in 1998. It was soon revised to incorporate mathematics.

Following several tests for validity and reliability and following the use of the instrument by hundreds of teachers, it was again revised in 2000 (Pilburn & Sawada, no date). The RTOP has an associated training manual for observers (Sawada, et al., 2000) with a supporting web site, (http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/using_RTOP_1.html) containing video clips for the training process. Though helpful for training, these videos show university-level teaching and have limited application for K-12 observers. Studies on validity and reliability are available in the technical report NO. IN00-1 (Sawada, 2000).

"[RTOP] is an observational instrument designed to measure 'reformed' teaching'" (Pilburn & Sawada, no date). The analysis of reformed teaching takes place through three sections in the RTOP instrument. Section One, Background Information, focuses specifically on the school, teacher, and observation time and date. There is no student information requested. The second section, Contextual Background and Activities, provides a space for a description of the lesson, the classroom setting, and other details about the students and teacher.

The rating section consists of indicators for various parts of a lesson including Lesson Design and Implementation, Content (Propositional Knowledge and Procedural Knowledge), and Classroom Culture (Communicative Interactions and Student/Teacher Relationships). Each of these domains and sub-domains is rated using five indicators. Ratings range from 0 = "Never

Occurred" to 4 = "Very Descriptive" for a possible total of 100 on the instrument. There is no opportunity to rate "Not Observed" or "N/A," as in the ITC COP. Two pages are included for taking notes and for additional comments. This five-page document may be attractive to those projects and evaluators who wish to examine these domains in a comprehensive but concise manner.

RTOP - Definition of Inquiry

The RTOP is a reform-based instrument built on recommendations from *Project 2061* (AAAS 1989), the *National Science Education Standards* (National Academy of Science, 1996) and mathematics reform documents from the National Council of Teachers of Mathematics including the *Curriculum and Evaluation Standards* (1989), the *Professional Standards* (1991) and the *Assessment Standards* (1995) incorporating the principles of equity, curriculum, teaching, learning, assessment, technology from *Standards 2000* (NCTM, 2000). It shares these foundational documents with the ITC COP.

Reform, as illustrated in this instrument, focuses on lessons where students participate in inquiry and have early and continuous engagement in collection and use of evidence. The lessons build on student preconceptions, promote discourse among students about scientific and mathematical ideas, and challenge students to take responsibility for their own learning. It is based in the philosophy that learning progresses from concrete to abstract, thus lessons should begin with "active manipulation of physical experimentation, or…involve the use of existing evidence of data sets" (Pilburn & Sawada, [no date], p. 4). Learning does not occur in isolation and should occur as students work in groups. The teacher's role is as a support for the student as they are participating in inquiry and discourse, to encourage all students to accept responsibility for their own learning, and to be a model of scientific processes and states of mind. (Pilburn & Sawada, [no date], p. 5).

RTOP - Data Collection

Background information includes the number of years of teaching and the teaching certification held by the teacher. The starting time and end time can indicate the length of the lesson. All of these can be covariates in any statistical analysis of the observation ratings.

Qualitative data are derived from the initial contextual description of the classroom setting, students, and description of the lesson in addition to the observer notes that assist in documenting the ratings. These pages prompt for "salient notes" (Sawada, Piburn, Turley, Falconer, Benford, R., Bloom, I., & Judson, E. (2000 Revision), p. 4) with no further description.

As previously described, each of the three domains (totaling 5 subdomains) is defined by five individual indicators rated on a scale from 0 (never occurred) to 4 (very descriptive) with a total of 100 possible. An average can be calculated for each subdomain for each teacher and followed over time if the teacher is part of a group with a long-term intervention. In addition, mean ratings for teachers who belong to a specific subset of teachers within a project (all math teachers within the project or all teachers in a particular building) may be calculated. Statistical tests of mean differences can be conducted to document changes in implementation of a project over time. A total score for each teacher or mean score for the group may also be used to indicate the level of

implementation for that group. Because these ratings based on Likert ordinal scales, caution regarding the depth of statistical analysis should be taken.

RTOP – Research

All research reported in this section was drawn from the ACEPT Technical Report IN00-3 (Pilburn & Sawada, [no date]). The instrument was originally tested for reliability and validity. Results are reported within this document and are summarized here.

Reliability. Reliability focused on the total instrument and the five subscales. Observations were conducted by two observers three times for each of 20 teachers. Sixteen pairs of observations were used in the final analysis. Reliability scores can be found in Table 1. With the exception of Subscale 2, similar reliabilities were obtained between the subscales (0.872 to 0.946) and the whole test (0.954).

Table 1. Reliability results for subscales and RTOP instrument

Subscale	R-squared
1. Lesson Design and Implementation	0.915
2. Content: Propositional Knowledge	0.670
3. Content: Procedural Knowledge	0.946
4. Classroom Culture: Communicative Interactions	0.907
5. Classroom Culture: Student/Teacher Relationships	0.872
RTOP Total Instrument	0.954

ACEPT Technical Report IN00-3 (Pilburn & Sawada)

Additional reliability studies using fewer observations were conducted with similar results.

Validity. Face validity is based on the three national mathematics and science standards documents that served as the source for the development of the items. Construct validity is based in the ability of the document to explore and document the principles of "inquiry-based" and "standards-based." A correlational analysis on the five subscales on their predictive ability for the total scores was applied. Table 2 shows that subscale three was the strongest predictor, with all but subscale 2 being strong predictors of the total instrument.

Table 2. Construct validity for RTOP subscales

Subscale	R-squared
1. Lesson Design and Implementation	0.956
2. Content: Propositional Knowledge	0.769
3. Content: Procedural Knowledge	0.971
4. Classroom Culture: Communicative Interactions	0.967
5. Classroom Culture: Student/Teacher Relationships	0.941

ACEPT Technical Report IN00-3 (Pilburn & Sawada)

Predictive validity was conducted in four settings in sixteen classrooms in community colleges and universities. Comparisons of teachers' normalized gain scores on the RTOP to their students' achievement in physical science, physics and mathematics were conducted. Correlations are

shown in Table 3. These correlations are significant at the p<0.01 level. Low Ns caution limited interpretation on predictive ability despite the strong correlation.

Table 3. Correlation of physical science, physics and mathematics students to teacher's RTOP normalized gain score

Content Area	Correlation: RTOP and Normalized Gains
Mathematics (n = 6)	
Conceptual Understanding	0.94
Number Sense	0.92
Physical Science (n = 6)	0.88
Physics (n = 4)	0.97

ACEPT Technical Report IN00-3 (Pilburn & Sawada)

Factor Analysis. A comprehensive factor analysis revealed three factors represented by the items. Factor 1, Inquiry, includes items 3, 4, 11, 12, 13, 14, 16. Factor 2 aligns with subscale 2, Content Prepositional Knowledge and incorporates items 6, 7, and 10. Factor 3 was an unanticipated factor when compared to the initial purposes for the instrument. This factor is described as reflecting the qualities of fairness, justice or equity and focuses on the student as leading critical components of his or her own learning. This factor was labeled Collaboration.

A refined factor analysis yielded three additional factors sharing loadings across factors: content pedagogical knowledge, community of learners, and reform teaching. More information can be found in the ACEPT Technical Report IN00-3 (Pilburn & Sawada, [no date]).

Norms. Norms were derived from a sample of 153 classes that incorporated 38 mathematics classes, 51 science classes and 12 methods courses. Various educational levels were represented: 62 university courses, 26 community college courses, 37 high school classes and 28 middle school classes. Norms for assistance in interpreting results can be found in Table 4.

Table 4. RTOP norms for mathematics and science classes in various educational levels

Level	Mathematics			Science			Total		
	n	mean	s.d.	N	mean	s.d.	N	mean	s.d.
University	10	63.9	22.0	40	58.3	21.3	50	59.4	21.3
CC	3	48.0	11.8	23	50.1	21.6	26	49.9	20.6
HS	12	48.8	10.8	25	41.8	20.2	37	44.1	17.8
MS	13	46.8	19.0	15	50.0	14.1	28	48.5	16.3
RTOP	38	52.0		103	51.0	20.9	141	51.3	50.1

ACEPT Technical Report IN00-3 (Pilburn & Sawada)

For teachers in middle school science, a score of 50 or above indicates the presence of some level of reform teaching. Likewise, a score of 46.8 for middle school mathematics teachers provides the same indicator.

College content faculty who were trained in the ACEPT program were compared to those faculty teaching content courses who were not trained. The comparison of mean observation scores of

both sets of teachers shows a difference of 24 points (reform classes mean = 61.7; non-reform mean = 37.6). RTOP is able to distinguish reform teaching as defined by this instrument in teachers at all levels of instruction.

Findings

Comparison of ITC COP and RTOP

The two instruments share obvious similarities. Their philosophical basis, indicators rated on a Likert scale, and prompts to record environmental and explanatory data to support and explain the ratings provide evaluative information on project implementation the extent to which practice is based on reform teaching. Both instruments yield sufficient quantitative data for limited statistical analysis. Qualitative data are compiled that can support and inform the results of such analysis.

Comparisons of the two instruments' indicators show both areas of commonality and difference. Both instruments contain rating categories for lesson design, implementation, content and classroom culture. The most obvious difference is the number of indicators attributed to each category. The RTOP allows only 5 indicators for each category or sub-category, whereas the ITC COP is designed with a varying number of indicators per category. This may have the affect of differential weighting of categories in the ITC COP. Likewise, with the RTOP limit of five indicators per subcategory, critical elements may be omitted or subsumed in an indicator, limiting the ability to differentiate to a useful extent.

The descriptors on the RTOP for the extreme ratings, "Never Occurred" for 1 and "Very Descriptive" for 4, may cause problems. These two descriptors are not indicators of opposite situations and therefore are imbalanced.

Another issue arises when an indicator may not be appropriate for the reform lesson being observed. A rater using the RTOP has no option to score an indicator that would not appropriately occur in the phase of an inquiry lesson being observed. This situation appears to be a result of there being no provision in the RTOP for a rating of "Not Applicable" or "Didn't Occur" outside of the rating of the indicator itself, as appears in the ITC COP.

Item Matches

A comparative analysis of the items on the RTOP and ITC COP was conducted. Items were examined in three ways. First, they were analyzed to see if they matched closely enough to be considered as identical indicators (See Table 5). If not, they were examined to see if they were included as part of an indicator on the other instrument (e.g. ITC COP indicators within an RTOP indicator). Last, they were analyzed in the reverse to see if the ITC COP indicators were written broadly enough to include RTOP indicators.

Table 5. Match of ITC COP items to RTOP items

RTOP Category	Design and Implementation			Content: Propositional Knowledge				Content: Procedural Knowledge		Classroom Culture: Communicative Interactions	Classroom Culture: Student/Teacher Relationships	
RTOP Items	1	2	4	7	8	9	10	11	14	20	21	22
ITC COP Items	I.3	I.6	I.1 II.1	I.7	III.3 III.5	III.7	III.8	III.7	III.9	IV.2	IV.1	IV.5

Twelve of the 25 RTOP items (48%) had matches with 14 of the 30 ITC COP items (43%). Fifty-two percent of the RTOP items and 57% of the ITC COP items did not have similar indicators in the other instrument. Items are described below. Parenthetical/explanatory language in the indicators has been omitted.

Design and Implementation. The RTOP category, Design and Implementation, matches indicators from ITC COP category I, Design, and II, Implementation.

RTOP 1 The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.
    ITC COP I.3 The instructions strategies and activities used in this lesson reflected attention to students' experiences, preparedness, prior knowledge, and/or learning styles.
RTOP 2 The lesson was designed to engage students as members of a learning community
    ITC COP I.6 The design of the lesson encouraged a collaborative approach to learning among the students.
RTOP 4 The lesson encouraged students to seek and value alternative modes of investigation or of problem solving.
    ITC COP I.1 The design of the lesson incorporated tasks, roles, and interactions consistent with investigative mathematics/science.
    ITC COP II.1 The instructional strategies were consistent with investigative mathematics/science.

Content: Propositional Knowledge. Of the five indicators in RTOP Content: Propositional Knowledge, four had matches in the ITC COP instrument. All but one of these four matches occurred in ITC COP category III, Mathematics and Science Content.

RTOP 7. The lesson promoted strongly coherent conceptual understanding.
    ITC COP I.7 Adequate time and structure were provided for "sense-making."
RTOP 8. The teachers had a solid grasp of the subject matter content inherent in the lesson.
    ITC COP III.3 Teacher-provided content information was accurate.
    ITC COP III.5 The teacher displayed an understanding of mathematics/science concepts.
RTOP 9. Elements of abstraction were encouraged when it was important to do so.
    ITC COP III.7 Elements of mathematical/science abstraction were included when it was important to do so.
RTOP 10. Connections with other content disciplines and/or real world phenomena were explored and valued.
    ITC COP III.8 Appropriate connections were made to other areas of mathematics/science, to other disciplines, and/or to real-world contexts.

Content: Procedural Knowledge. There were two ITC COP matches for two RTOP indicators in Content: Procedural Knowledge. Again, these ITC COP matches were in category III, Mathematics and Science Content, indicating some cross-over in the ITC COP from Mathematics/Science Content with the RTOP's Content: Procedural Knowledge category.

RTOP 11. Students used a variety of means to represent phenomena.
       ITC COP III.7 Elements of mathematical/science abstractions were included when it was important to do so.
RTOP 14. Students were reflective about their learning.
       ITC COP III.9 The degree of "sense-making: of mathematics/science content within this lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.

Classroom Culture: Communicative Interactions. Only one indicator, the RTOP Classroom Culture: Communicative Interactions category, matched one indicator in ITC COP's Classroom Culture category.

RTOP 20.There was a climate of respect for what others had to say.
       ITC COP IV.2 There was a climate of respect for students' ideas, questions and contributions.

There were two matches in the RTOP's Classroom Culture: Student/Teacher Relationships each matching one ITC indicator in the Classroom Culture category.

RTOP 21. Active participation of students was encouraged and valued.
       ITC COP I.5 The instructional strategies and activities reflected attention to issues of access, equity, and diversity for students.
RTOP 22. Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.
       ITC COP IV.5 The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/or propositions.

As can be seen, these instruments are designed to look for similar indicators of reform teaching. However, about half (13 of 25) of the RTOP indicators do not have matching or comparable indicators on the ITC COP. Likewise, 16 of 30 indicators on the ITC COP show no matches or similar indicators on the RTOP. Many of these items occur in one of the next two groupings.

ITC COP indicators that are included within the RTOP indicators

An examination of categories in each document stated broadly enough to be inclusive of indicators of the other instrument was undertaken. These matches are not as obvious, and require interpretation into the test developers' intentions. The categories within which they fall provide certain clues, but when examining the indicators outside of the category, other characteristics become evident. Indicators with multiple descriptors measure more than one attribute, so will appear in more than one category.

Looking first at the RTOP general indicators, only four RTOP indicators are stated broadly enough to incorporate individual indicators from the ITC COP. They fall into the four categories shown in table 6.

Table 6. Indicators from the ITC COP falling within RTOP indicators

RTOP Category	Content: Propositional Knowledge	Content: Procedural Knowledge	Culture: Communicative Interactions	Culture: Student/Teacher Relationship
RTOP Items	7	15	18	21
ITC COP Items within RTOP Items	III.4	III.4 IV.5	I.6	I.5

As can be seen, none of the items except for RTOP 7 (Content: Propositional Knowledge) and ITC COP III.4 (Mathematics/Science Content) align in categories as would be expected.

RTOP 7. The lesson promoted strongly coherent conceptual understanding.
       ITC COP III.4 Students were intellectually engaged with important ideas relevant to the focus of the lesson.

The other matches in RTOP Procedural Knowledge are from the ITC COP categories of Mathematics/Science Content and Classroom Culture.

RTOP 15. Intellectual rigor, constructive criticism, and the challenging of ideas were valued.
       ITC COP III.4 Students were intellectually engaged with important ideas relevant to the focus of the lesson.
       ITC COP IV.5 The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/or propositions.

Both RTOP categories of Culture: Communicative Interactions and Culture: Student/Teacher Relationships matched ITC COP indicators in the Design category (I).

RTOP 18. There was a high proportion of student talk and a significant amount of it occurred between and among students.
       ITC COP I.6 The design of the lesson encouraged a collaborative approach to learning among the students.
RTOP 21. Active participation of students was encouraged and valued.
       ITC COP 1.5 The instructional strategies and activities reflected attention to issues of access, equity, and diversity for students.

RTOP Indicators that are included in ITC COP Indicators

A reverse process to that described above was undertaken. An examination of ITC COP indicators written broadly enough to include indicators from the RTOP begins to illustrate the differences between the two instruments. About 43% of the ITC COP indicators are written in a manner in which the specific indicators in the RTOP could be incorporated within them. This is in contrast to only four of the 25 RTOP indicators (16%) written in such a broad way. The specific RTOP items that fall within ITC indicators are shown in Table 7. Alignment between the two instruments is presented using ITC COP conceptual organizers.

Table 7. Indicators from the RTOP that fall into the ITC COP indicators

ITC COP Indicators	Design					Implementation		Mathematics/Science Content			Classroom Culture		
ITC COP Items	I.1	I.2	I.3	I.7	I.9	II.6	II.7	III.4	III.6	III.9	IV.1	IV.2	IV.4
RTOP Items within ITC COP Items	2	1	1	9	20	17	4	13	4	9	1	5	5
	3	2	3	10	21		12		13	11	2	18	18
	5	3		13			22		15	16	5	19	19
	12			14			24		22	22	13	21	20
	13			22							16	24	21
	22										19		24
											22		

Design

ITC COP I.1. The design of the lesson incorporated tasks, roles, and interactions consistent with investigative mathematics/science.
    RTOP 2 The lesson was designed to engage students as members of a learning community.
    RTOP 3 In this lesson, student exploration preceded formal presentation.
    RTOP 5 The focus and direction of the lesson was often determined by ideas originating with students.
    RTOP 12 Students made predictions, estimations, and/or hypotheses and devised means for testing them.
    RTOP 13 Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
    RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.
ITC COP I.2. The design of the lesson reflected careful planning and organization
    RTOP 1 The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.
    RTOP 2 The lesson was designed to engage students as members of a learning community.
    RTOP 3 In this lesson, student exploration preceded formal presentation.
ITC COP I.3. The instructional strategies and activities used in this lesson reflected attention to students' experience, preparedness, prior knowledge, and/or learning styles.
    RTOP 1 The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.
    RTOP 3 In this lesson, student exploration preceded formal presentation.
ITC COP I.7. Adequate time and structure were provided for "sense-making."
    RTOP 9 Elements of abstraction were encouraged when it was important to do so.
    RTOP 10 Connections with other content disciplines and/or real world phenomena were explored and valued
    RTOP 13 Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
    RTOP 14 Students were reflective about their learning.
    RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.
ITC COP I.9. Adequate time allotted for students to react to other students' questions or comments.
    RTOP 20 There was a climate of respect for what others had to say.
    RTOP 21 Active participation of students was encouraged and valued.

Implementation

ITC COP II.6. The teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem solving.
        RTOP 17 The teacher's questions triggered divergent modes of thinking.
ITC COP II.7. Encouraged independent problem-solving.
        RTOP 4 This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.
        RTOP 12 Students made predictions, estimations, and/or hypotheses and devised means for testing them.
        RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.
        RTOP 24 The teacher acted as a resource person, working to support and enhance student investigation.

Mathematics and Science Content

ITC COP III.4. Students were intellectually engaged with important ideas relevant to the focus of the lesson.
        RTOP 13 Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
ITC COP III.6. Mathematics/science was portrayed as a dynamic body of knowledge continually enriched by conjecture, investigation analysis, and/or proof/justification.
        RTOP 4 This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.
        RTOP 13 Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
        RTOP 15 Intellectual rigor, constructive criticism and the challenging of ideas were valued.
        RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.
ITC COP III.9. The degree of "sense-making" of mathematics/science content within this lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.
        RTOP 9 Elements of abstraction were encouraged when it was important to do so.
        RTOP 11 Students used a variety of means to represent phenomena.
        RTOP 16 Students were involved in the communication of their ideas to others using a variety of means and media.
        RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.

Classroom Culture

ITC COP IV.1. Active participation of all was encouraged and valued.
        RTOP 1 The instructional strategies and activities respected students' prior knowledge and preconceptions inherent therein.
        RTOP 2 The lesson was designed to engage students as members of a learning community.
        RTOP 5 The focus and direction of the lesson was often determined by ideas originating with students.
        RTOP 13 Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
        RTOP 16 Students were involved in the communication of their ideas to others using a variety of means and media.
        RTOP 19 Student questions and comments often determined the focus and direction of classroom discourse.
        RTOP 22 Students were encouraged to generate conjectures, alternative solutions, strategies, and ways of interpreting evidence.

ITC COP IV.2. There was a climate of respect for students' ideas, questions and contributions.
RTOP 5 The focus and direction of the lesson was often determined by ideas originating with students.
RTOP 18 There was a high proportion of student talk and a significant amount of it occurred between and among students.
RTOP 19 Student questions and comments often determined the focus and direction of classroom discourse.
RTOP 21 Active participation of students was encouraged and valued.
RTOP 24 The teacher acted as a resource person, working to support and enhance student investigation.
ITC COP IV.4. Students were intellectually engaged with important ideas relevant to the focus of the lesson.
RTOP 5 The focus and direction of the lesson was often determined by ideas originating with students.
RTOP 18 There was a high proportion of student talk and a significant amount of it occurred between and among students.
RTOP 19 Student questions and comments often determined the focus and direction of classroom discourse.
RTOP 20 There was a climate of respect for what others had to say.
RTOP 21 Active participation of students was encouraged and valued.
RTOP 24 The teacher acted as a resource person, working to support and enhance student investigation.

From the last two comparisons, it becomes clear that the RTOP indicators are more precisely focused. For the most part, they make up a checklist for the observer. Did the students use a variety of means to represent phenomena (#11)? Did students make predictions, estimations, and/or hypotheses and devise means for testing them (#12)? Were students involved in the communication of their ideas to others using a variety of means and media (#16)? These indicators raise the question of the appropriateness of a rating scale in collecting data for these indicators on the RTOP when a simple "yes" or "no" would most appropriately address these indicators.

The ITC COP indicators require a deeper understanding of what reform teaching is and asks for a judgment about the appropriateness of a specified strategy to the particular lesson being observed. There is more of a focus on pedagogical content knowledge and its appropriate application to the particular lesson in the classroom with the students involved. Observers could not rate these indicators without a deep understanding of content, pedagogy, knowledge of students and classroom experience. The rating for these items requires more than an observation of the classroom.

For example, consider II.4: The pace of the lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson, or II.6: The teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem solving… . These two examples show that a deep understanding of the students, how students at that age learn, lesson design, content, and proposed lesson outcomes are required to accurately rate these indicators. These are not indicators that can be observed as the class is taking place, but one that must be analyzed in the context of prior knowledge and experience of the rater.

Summary

Based on these findings, the authors offer the following answers to questions posed in the introduction of this paper.

1. What are the underlying assumptions about reform teaching in these documents?

Both the RTOP and the ITC COP are instruments that are in wide use among evaluators and project staff seeking to determine if reform teaching as a result of a project intervention is being implemented. These instruments focus primarily on classroom implementation and provide opportunities for documenting environmental, demographic, and teacher perceptions associated with the lesson. Both are based on the same reform documents in mathematics and science and assumptions about the nature of reform teaching as interactive, collaborative, and student directed are evident within both.

2. By whom should each instrument be used?

The RTOP is an instrument that can be used by most educators given training on the instrument and agreement on what reform teaching looks like. Indicators are stated precisely, seldom requiring deep interpretation or extensive background knowledge or classroom extensive experience. The instrument could be used by graduate students lacking classroom experience with appropriate interrater reliability training.

On the other hand, the ITC COP requires a deeper understanding of the classroom, the content and the student. It assumes a nuanced understanding of reform teaching that may not be present among all potential observers. In addition to interrater reliability training and retraining on a regular basis, this instrument is best used by experienced educators and evaluators with grounding in the project being evaluated, along with the project's definition of reform teaching.

3. Which instrument provides ease of use to the rater while meeting evaluation needs?

Using only the rating sections of both instruments provides comparable ease of use. However, the addition of the "N/A" and "Don't Know" ratings on the ITC COP provide options that will provide a more realistic picture of the specific lesson being observed. The forced rating on some RTOP indicators may produce frustration on the part of the rater in forcing a choice when one was not appropriate or not observed in this specific lesson. The extended additional documentation on the ITC COP is useful when qualitative analysis of the teaching process is required. This documentation requires additional time, including pre and post interviews with the teacher, and may not be practical for all projects.

4. Which instrument provides data that represent actual classroom practice?

The ability to rate an indicator as not appropriate (N/A) for a lesson provides a more accurate representation about what should be and is occurring in a lesson. For example, exploration may not be observed because it was completed several lessons ago and students are now developing a process for testing a hypothesis. Observers are inclined to give that rating a "0" if it is not there,

whether it should have been or not. To be rated low on such a rating is not representative of what is or what should be occurring in that lesson on that day. This is a major drawback to the RTOP, which forces a rating on all indicators. Despite repeated interrater reliability training, the authors have found that some raters are reluctant to rate a teacher low on such indicators because they realize that some indicators are not appropriate for that lesson and raters do not want it to negatively affect the total score for the teacher. A tendency to infer capacity or a likelihood that the teacher would positively demonstrate the indicator in another situation has been observed.

Both the ITC COP and the RTOP are useful instruments in documenting classroom implementation of reform teaching if they are aligned with the philosophy of the project and are used by raters with the knowledge and experience required to accurately interpret an indicator. This is more problematic for the ITC COP than the RTOP. In determining which instrument to use in evaluation, discussions with the project leadership regarding the characteristics, strengths, and weaknesses of each instrument, along with project alignment will provide the opportunity for collecting the data that best represents the project's effects.

In summary, the two instruments and, by extension, the various derivations and site-specific modifications of the ITC COP are potentially useful instruments for purposes of rating classroom implementation of mathematics and science curriculum across domains of interest from a reform teaching perspective. While the RTOP offers perhaps an easier instrument for novice professionals to use, the ITC COP's inclusion of "N/A" and "Not Observed" ratings makes for greater sensitivity in assessing classroom performance. The role of proper training and oversight of inexperienced classroom observers cannot be ignored in either case, as the inherent complexity of both instruments requires care and knowledge.

The authors recognize the limitations of this type of analysis. It, of necessity, is based on the analysts' and reviewers' experiential knowledge and understanding of the standards documents' definitions of reform teaching, informed by the documentation supporting these instruments. The lack of metrics for the ITC COP is a limitation of the analysis, as inference from another document similar to the ITC COP is weak support for its reliability and validity, though the ITC COP continues to be used widely in MSP evaluations.

The authors urge ongoing research into validity and reliability of the instruments that may inform some of the deficiencies of these instruments. If these data are available, release of the information can assist the field in understanding what is happening in the classroom.

Phase two of the analysis of the RTOP and ITC COP, which will compare ratings between the instruments for observations of the same lessons, may assist in deepening evaluators' understanding of the appropriate use of these instruments, as well as areas ripe for refinement. The authors welcome feedback and discussion of these two instruments.

References

American Association for the Advancement of Science (AAAS). (1989). *Project 2061: Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology.* Washington, DC: Author.

Banilower, E.F. (2005). *A study of the predictive validity of the LSC Classroom Observation Protocol.* Retrieved September 30, 2006 from http://www.horizon-research.com/reports/2005/COP_validity.phprl.

Baron, J. (2006). <DED MSP evaluation meeting presentation>

Horizon Research, Inc. (2002). *Inside the classroom interview protocol.* Retrieved September 30, 2006 from http://www.horizon-research.com/instruments/clas/interview/pdf.

Horizon Research, Inc. (March 2000). *Validity and reliability information for the LSC Classroom Observation Protocol.* Retrieved April 15, 2006 from http://www.horizon-research.com/LSC/news/cop_validity_2000.pdf.

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

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

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

National Council of Teachers of Mathematics (NCTM). (2000). *Principals and standards for school mathematics.* Reston, VA: Author.

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

Pilburn, M. & Sawada, D. (No date.). *Reformed Teaching Observation Protocol (RTOP): Reference Manual*. (ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona State University.

Sawada, D., Pilburn, M., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed Teaching Observation Protocol (RTOP) (ACEPT Technical Report NO. IN00-1).* Tempe, AZ: Arizona State University.

Sawada, D., Piburn, M., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000 Revision). *Reformed teaching observation protocol (RTOP).* Tempe, AZ: Arizona State University.

Sawada, D., Piburn, M., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). *Reformed teaching observation protocol (RTOP) Training Guide (ACEPT Technical Report No. IN00-2)* Tempe, AZ: Arizona State University.

Weiss, I.R., Pasley, J.D., Smith, P.S., Banilor, E.R., and Heck, D.J. (2003). *Looking inside the classroom: A study of K-12 science and mathematics in the united states.* Horizon Research, Inc.

Appendix

Observation Instruments Derived from or Informing
Inside the Classroom Observation and Analytic Protocol

Core Evaluation Classroom Observation Protocol. Center for Excellence in Teacher Preparation (CETP). University of Minnesota. Available from http://cehd.umn.edu/CAREI/CETP/PDF/COP2002.pdf

InSITE Classroom Observation Protocol. Available from John Lannin, University of Missouri, Columbia, Missouri. Lanninj@missouri.edu.

Knapp, M.S. (2003). Professional development as a policy pathway. In R. E. Floden (Ed). *Review of research in education 27* (pp. 116, 120). Washington DC: American Educational Research Association.

Littleton Academy Simplified Horizon COP. Available from http://www.cde.state.co.us/cdechart/guidebook/adm/pdf/LAClassroomObservationTeacher.pdf

2005-06 Local Systemic Change Classroom Observation Protocol (LSC COP). Horizon Research, Inc. www.horizon-research.com/instruments

Medina, K., Pollard, J., Schneider, D., & Leonhardt, C. (2000). *How do students understand the discipline of history as an outcome of teachers' professional development?"* Oakland, CA: Regents of the University of California.

Oregon Mathematics Leadership Institute Classroom observation Protocol (OMLI). RMC Research Corporation, Portland, OR.

Oklahoma Teacher Education Collaborative (O-TEC). Core Evaluation Classroom Observation Protocol (CECOP). Available from Mary T. Stewart mary-stewart@tulsa.edu; Joyce C. Townsend joyce-townsend@utusa.edu, or Curtis E. Miller curtis-miller@utulsa.edu.

Shavelson, R. J., & Towne, L. (Eds). (2003). *Scientific research in education* (pp. 105-108). Washington, DC: National Academy Press.

Vermont Classroom Observation Tool (VCOT). The Vermont Institutes. Available from www.vermontinstitutes.org/leadership.html.