Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol

Daiyo Sawada
University of Alberta

Michael D. Piburn and Eugene Judson
Arizona State University

Jeff Turley, Kathleen Falconer, Russell Benford and Irene Bloom
Arizona Collaborative for Excellence in the Preparation of Teachers

The National Science Foundation has funded 22 Collaboratives for Excellence in Teacher Preparation. Despite the remarkable allocation of resources to this effort, it has proven exceptionally difficult to demonstrate the effectiveness of collaborative reform. In large part, this has resulted because of the difficulty of defining and measuring reform. The Reformed Teaching Observation Protocol (RTOP) was designed by the Evaluation Facilitation Group of the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT). It is a 25-item classroom observation protocol that is (a) standards based, (b) inquiry oriented, and (c) student centered. This instrument has provided the definition for reform and the basis for evaluation of the ACEPT collaborative. The data upon which this report is based were collected over a period of more than 2 years from 153 public school, college, and university mathematics and science classrooms. A trained team of observers consisting of two faculty members and seven graduate students was able to achieve exceptionally high levels of interrater reliability. Internal consistency, as estimated by Cronbach's alpha, was also remarkably high. Correlation coefficients ranging from 0.88 to 0.97 between RTOP scores for classrooms, and mean normalized gain scores for students in those classrooms on achievement measures demonstrate that reform, as defined by ACEPT and measured by the RTOP, has been effective.

The reform of science and mathematics education over the past decade has been spurred by high-profile policy documents produced by major science and mathematics bodies (American Association for the Advancement of Science, 1989, 1993; National Council Teachers of Mathematics, 1989, 1991, 1995, 2000; National Research Council, 1996, 2000). Despite the plethora of literature advocating a shift in teaching and learning of science and mathematics toward student-centeredness, the development of sensitive evaluation frameworks and data collection instruments appropriately aligned to these efforts has been both difficult (National Institute for Science Education, 1999) and controversial (Linn, 2000). In particular, constructing classroom observation instruments to measure the degree to which classrooms have become aligned with reform principles has only just begun (Tittle & Pape, 1996).

In the absence of strong quantitative evidence, debate over the impact of reform rages unabated (Calhoun, Bohlin, Bohlin, & Tracz, 1997; Mayer, 1999). The evidential basis supporting reform remains soft, despite research summaries to the contrary (Hiebert, 1999; Westat*McKenzie Consortium, 1998). For example, a 5-year, $4.6 million evaluation of the National Science Foundation's program of Statewide Systemic Initiatives concluded that the impact of this $300 million program "has been extremely hard to measure" and that "evidence of improved test scores as a direct result of the reforms is even more tenuous" (Mervis, 1998, p. 1800).

Statement of the Problem

Overcoming these evidential difficulties was the focus of the Evaluation Facilitation Group (EFG) of the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT), one of 22 Collaboratives for Excellence in Teacher Preparation (CETP) funded by NSF. The EFG consisted of an external and an internal evaluator and seven graduate assistants. The charge to this group was to examine the reform efforts of ACEPT and to report on their efficacy. To this end, the greatest challenge confronting EFG was to define and measure reform in a manner that was intellectually compelling and empirically convincing to a diverse group of science and mathematics educators. This task required 2 years for completion and resulted in the development of the Reformed Teaching Observation Protocol (RTOP), an instrument that eventually anchored and validated the ACEPT evaluation (Sawada & Piburn, 2000).

Defining Reform

Students of science and mathematics education of the past century would be wholly justified in claiming that a stable constant in the field is that of reformation. Reform is a term often used in the annals, and it is a term connoting various definitions. It is fitting, then, to render precisely what the authors mean when using the term reform.

Ideally, reform reaches beyond classroom practice and embodies student and teacher empowerment, development of critical thinking, and a culture that supports change (Anderson et al., 1994; Shymansky & Kyle, 1990). The reforming of instruction and learning is a movement away from the traditional didactic practice toward constructivism (Anderson et al. 1994).

The framework of reform within which the ACEPT effort was developed has been well defined by the National Science Foundation's Collaboratives for Excellence in Teacher Preparation. Conditional to this definition, reform predicates students using data to justify opinions, experiencing ambiguity as a result of learning, and learning from one another. Additionally, reform presupposes that teachers do not emphasize lecture, but rather stress a problem-solving approach and foster active learning (Frantz, Lawrenz, Kushner, & Millar, 1998).

Development of the Instrument

The question that focused the development of the RTOP was this: "How would you know if a mathematics or science classroom was reformed?" The EFG began by conducting a review of existing observation instruments and of the literature articulating the science and mathematics reforms. The literature reviews enabled us to articulate a basic set of criteria for the instrument. The instrument would have to be focused on both science and mathematics, standards based, focused exclusively on reform rather than the generic characteristics of good teaching, easy to administer, appropriate for classrooms K-20, valid, and reliable.

Development began with a pool of existing items drawn from four major sources: (a) research studies grounded in the 1989-2000 science and mathematics standards; (b) existing instruments, especially the Horizon Research protocol (Horizon Research, 1998); (c) the work of ACEPT staff, especially that of Baker and Piburn (1997), Carlson, Bukirk, and Halloun (1999) and Lawson (1995); and (d) the ACEPT Evaluation Facilitation Group. Items that were neither standards based nor inquiry oriented or had a more generic focus were discarded, leaving a pool of about 60. From this group, 25 items were assigned to a system of categories consistent with the reform literature.

The final version of the RTOP (Appendix A) contains three scales, the first consisting of five items and the second and third of 10 items each. The second and third scales are further subdivided into two subscales of five. The three scales are as follows:

Lesson Design and Implementation: These five items in the first scale of the RTOP reflect the ACEPT definition of the structure and nature of a reformed lesson. Reformed teaching is enhanced by the design and sequence, as well as the pedagogical setting within which it takes place. For example, a reformed lesson begins by acknowledging and respecting ideas that students bring to the classroom. Students are envisioned as a community of inquirers and, as such, engage in exploration before attempting explication or definition.

Content: While appropriate content is an important element of any teaching, a definition of reformed teaching must acknowledge the distinction between content knowledge and pedagogical content knowledge that was first proposed by Shulman (1986). We take this distinction one step further in the second scale of the RTOP by subdividing pedagogical content knowledge into two kinds: (a) Propositional Knowledge (e.g., "the lesson promoted strongly coherent conceptual understanding" or "connections with other content disciplines and/or real world phenomena were explored and valued) and (b) Procedural Knowledge (e.g., "students made predictions, estimations and/or hypotheses and devised means for testing them" or "students were reflective about their learning").

Classroom Culture: In a reformed classroom, (a) communicative interactions are very diverse and decentralized. As a result, (b) student/teacher relationships are more egalitarian with teachers supporting initiatives coming from students.

Using the 25-item instrument, all members of the EFG observed and scored videotaped science and mathematics lessons taught by student teachers. Using these data, interrater reliabilities were computed, and the judgments of the reviewers were discussed. This process was continued over three semesters using new videotaped lessons. In this way, the RTOP items were continuously revised. As this was happening, a training manual that could be used to convey the developing interpretive consensus underlying the increasing reliability estimates was written (Sawada et al., 2000).

With the training manual as a guide, members of the EFG began visiting university and community college classrooms during the spring of 1999 to make live observations designed to further refine and improve the RTOP. Teams of at least two, and often many more, completed RTOP observations of the same class and met immediately afterwards to discuss and critique. This process continued through the summer. The July 1999 version of the RTOP (Sawada & Piburn, 2000) marked the end of the developmental process. When the final 25 items were compared with the initial 25, only one remained in its original form. All others had undergone substantial revision or had been replaced with new items.

As ACEPT approached its final year, the EFG designed a set of more formal studies that would contribute to the final evaluation report. These included a new set of quasi-experimental comparisons of traditional vs. reformed teaching. The total sample consisted of 287 RTOP forms collected from over 141 mathematics and science classrooms in middle schools, high schools, community colleges, and universities. These were gathered by nine trained observers and constitute the data set upon which this report is based.

Reliability

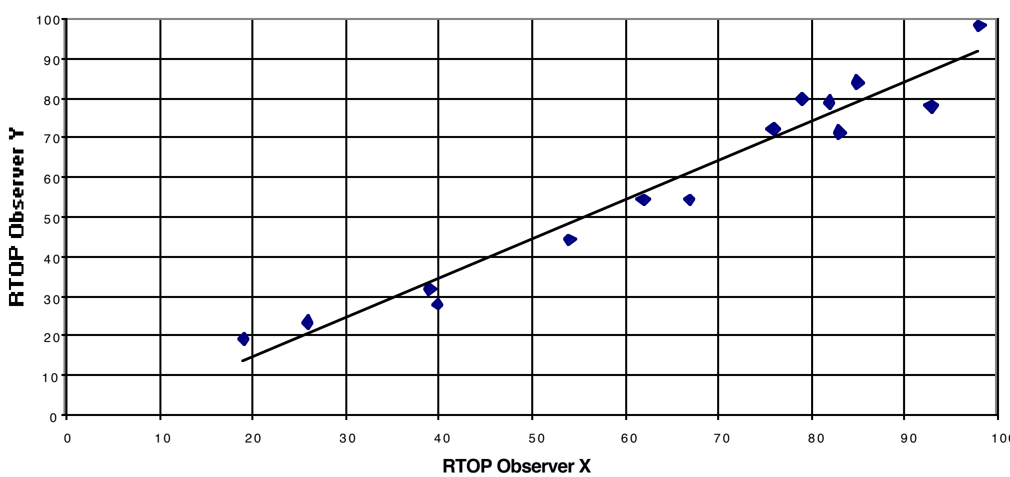
The RTOP was used on all courses included in the fall 1999 evaluation of ACEPT. Each of the courses was to be observed three times, once toward the beginning of the course, again during the middle, and a third time toward the close of the course. In order to get an early reading of interrater reliability, observers agreed to work in pairs for some of the initial observations.

As part of this plan, two members of the EFG paired up to do a subset of observations on the same classes. The first 16 such pairs (a total of 32 independent observations) were used to calculate estimates of interrater reliability. Two items of technical significance should be noted. First, 17 pairs were available for analysis, but one of the lessons was so strikingly unique that it prompted discussion between the two observers. The ratings could no longer be considered independent, and the observations were excluded from the analysis. Second, for three of the paired observations, the instructor was the same but the paired observation was of a lesson taught on a different day though with the same class. These three non-paired data points were still included in the analysis but variability introduced by this circumstance may produce an underestimate of reliability.

Estimates of interrater reliability were obtained by computing a best-fit linear regression of the observations of one observer on those of the other. Figure 1 shows a scatter plot of the 32 data points (some fall on each other). The equation for the best fitting line and the proportion of variance accounted for by that line are shown. The correlation coefficient is 0.98, and the variance shared between observers is 95%.

Further data suitable for estimating interrater reliability became available in the fall of 1999 when, as part of the ACEPT evaluation, two different members of EFG gathered RTOP observations on eight college and

Figure 1. *Estimate of the interrater reliability of the RTOP from observations in mathematics and physics classes. (*r*-squared = .954)*



university biology instructors. While the number of paired observations is not high, the correlation coefficient was 0.90. Figure 2 shows the scatterplot of the observations and the best-fitting line. This second data set confirms the high reliabilities that paired observers who have received training are able to obtain with the RTOP.

A more classical psychometric assessment of the reliability of the RTOP was conducted by computing Cronbach's alpha for the instrument as a whole, as well as for each subscale (Table 1). The standardized item alpha for the entire instrument was a remarkable 0.97, which suggests a high degree of consistency across items. Subscale alphas were also high, despite the fact that each consisted of only five items, ranging from 0.80 to 0.93

Validity

The face validity of RTOP draws on three major sources: (a) *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), *Professional Standards for Teaching Mathematics* (NCTM, 1991), *Assessment Standards for School Mathematics* (NCTM, 1995), and *Principles and Standards for School Mathematics* (NCTM, 2000); (b) *National Science Education Standards* (NRC, 1996) and *Inquiry and the National Science Education Standards* (NRC, 2000); and (c) *Project 2061: Science for All Americans* (AAAS, 1989) and *Benchmarks for Scientific Literacy* (AAAS, 1993). A detailed discussion of these documents and their relationship to the RTOP can be found in the Reformed Teaching Observation Protocol reference manual (Piburn et al., 2000).

Table 1
Values of Cronbach's Alpha for Individual RTOP Scales and Subscales.

Scale	alpha
Lesson design and implementation	0.91
Content	
1. Propositional knowledge	0.80
2. Procedural knowledge	0.93
Classroom culture	
1. Communicative interactions	0.91
2. Student/teacher relationships	0.91

To test the hypothesis that reform is a powerful integrating force in the structure of RTOP, a factor analysis was performed on the instrument. The 25-item RTOP protocol was analyzed using a database containing observations from 141 public school, college, and university classrooms. The principal components extraction method and the principal axes extraction method were both performed, resulting in similar analyses (to be expected given the high reliability estimates). Because the sample size was adequate, the principal components analysis followed by a Varimax rotation is reported here.

Solutions asking for two, three, and four principle components to be extracted resulted in two strong factors and a weaker third factor as shown in Table 2. To confirm whether the third factor with eigenvalue 1.18 was a "legitimate" component, a Scree test was also performed. It showed that the third component is definitely located in the curvilinear region, thus justifying

Figure 2. *Estimate of the interrater reliability of the RTOP from observations in biology classes.* (r-*squared = 0.803)*

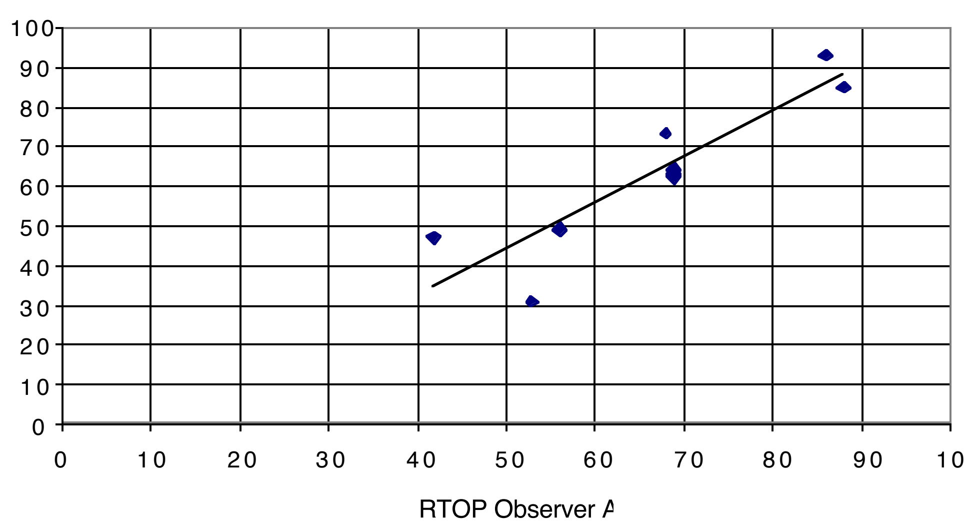


Table 2
Principal Components: Variance Distribution for Unrotated and Rotated Solutions.

		Unrotated Solution		Varimax Rotation	
Component	Eigenvalue	% of Variance Accounted For	Cumulative %	% of Variance Accounted for	Cumulative %
1	14.72	58.89	58.89	42.39	42.39
2	2.08	8.31	67.70	15.38	57.76
3	1.18	4.72	71.92	14.16	71.92

it as a legitimate component. Three factors were, therefore, retained and interpreted.

To visually and numerically simplify the factor pattern, a simple iconic coding was imposed on the coefficients in the factor pattern (see Piburn et al., 2000, for the actual values of the factor loadings). Using strings of asterisks to signify the magnitude of a coefficient, a visually more parsimonious pattern is revealed in Appendix A. The coding scheme, which only included coefficients equal to or greater than 0.50, is indicated at the bottom of the table.

The first factor draws heavily on all subscales except subscale 2. As mentioned in the construct validity section of this article, this general factor represents the overall thrust of the instrument. As such, the most appropriate name for this factor seems to be inquiry orientation.

The second factor, on the other hand, draws exclusively on subscale 2, a subscale that is characterized as representing propositional pedagogical knowledge. Because all five items of the subscale load on this factor, the same label seems appropriate for this factor.

The first two factors were expected in that they reflect the face validity of the items. The third factor was not anticipated. While accounting for less than 5% of the total variance in the instrument, it met both the eigenvalue and Scree criterion for inclusion. However, its emergence forced a closer look at the instrument.

The three items loading most heavily on Factor 3 come from the last section of the Classroom Culture portion of the instrument. That section was labeled, Student/Teacher Relationships. However, not all of the items in that section loaded on the third factor.

Achievement and the RTOP

A great deal of evidence has been collected confirming the predictive validity of the RTOP in different instructional settings on community college and university campuses. In the evaluation of introductory mathematics, physical science, and physics courses the RTOP was administered to instructors who had attended ACEPT workshops (experimental instructors) and to instructors who had not (control instructors). Content assessments were given as pre- and posttests in all classes.

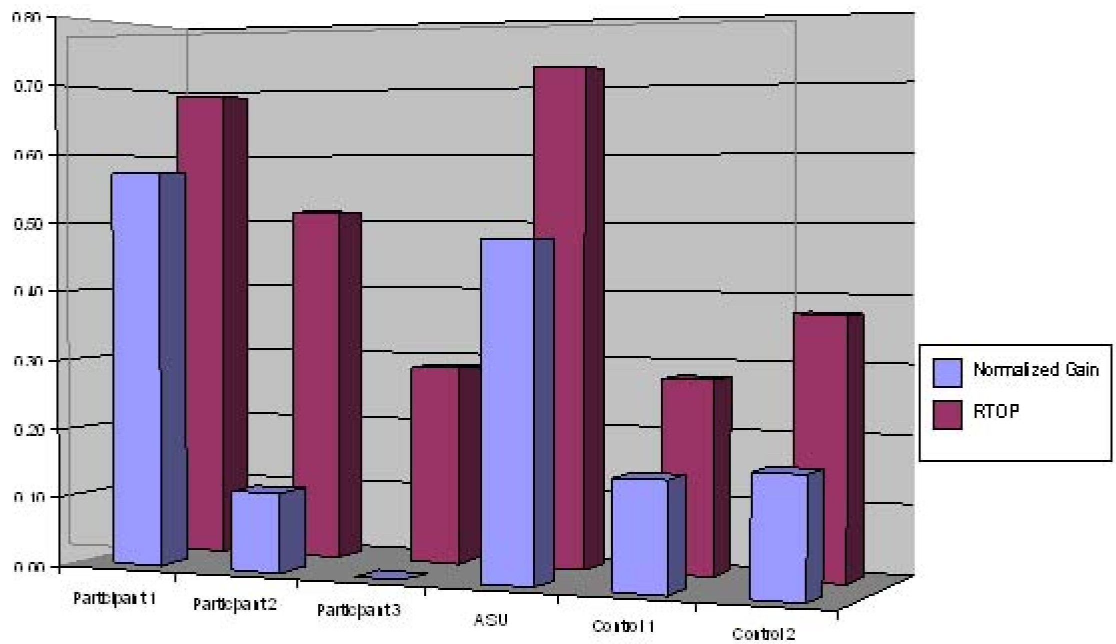
In these studies, multiple instructors were involved. There were 6 instructors in mathematics, 6 in physical science, and 4 in physics. Each instructor was observed a minimum of two times during the fall semester of 1999. The mean RTOP for each instructor was used as the RTOP score for that class. Normalized gain scores (often called the "Hake Factor" after physicist Richard Hake) were also calculated for each class. This score is used in preference to simple gain scores (post minus pre) because it takes into account initial differences on the pretest. Formulaically, Normalized Gain = (Post - Pre)/(Total - Pre). Conceptually, the normalized gain is the proportion of the possible increase in achievement that is accomplished by an individual or a group. It is a unitless value.

As an example, the RTOP and normalized gain scores for Physical Sciences 110 are presented in Figure 3. It can be seen that the normalized gain rises or falls much in the same manner as the RTOP score of the instructor of the class. The coefficient of correlation between RTOP scores and normalized gain scores for these six classrooms was 0.88. Despite the small sample size, a correlation of this magnitude is significant at the 0.01 probability level. Similar correlations were obtained in mathematics for conceptual understanding ($r = 0.94$), number sense ($r = 0.92$), and physics ($r = 0.97$).

Norms

It is important for users of an instrument like the RTOP to have some standards of performance against which to assess the scores achieved by individuals or samples in their own data sets. For those purposes,

Figure 3. *Covariation of mean RTOP scores with mean normalized gain scores in Physical Sciences 110 classes. (Normalized Gain vs. Avg RTOP on PCS PHS 110 Fall 1999)*



The correlation coefficient between Normalized Gain and RTOP is 0.88.

Table 3
Norms for RTOP Scores in Mathematics and Science Classrooms by Subject and Educational Level

	Mathematics			Science			Total		
	n	mean	*SD*	*n*	mean	*SD*	*n*	mean	*SD*
University	10	63.9	22.0	40	58.25	21.3	50	59.4	21.3
Community College	3	48.0	11.8	23	50.1	21.6	26	49.9	20.6
High Sch	12	48.8	10.8	25	41.8	20.2	37	44.1	17.8
Middle Sch	13	46.8	19.0	15	50.0	14.1	28	48.5	16.3
Total	38	52.0	18.1	103	51.0	20.9	141	51.3	20.1

norms from the sample used to create the factor analysis for this report are given in Table 3.

Science and mathematics classes are presented separately in Table 3. RTOP scores for this sample ranged from a high of 98 to a low of 18. The mean for the entire sample was 51.3. The mean scores for all mathematics and all science classes are virtually identical to one another and the same as the mean for the sample. University scores tend to be somewhat higher than those for community colleges or public schools. Although no statistical comparisons were made, high school science scores seem to be the lowest among all of the comparison groups.

One possible reason for the higher scores of the community college and university samples is that they consist of a large number of faculty who were involved in the ACEPT initiative. This is more pronounced at the university level than at the college level. In order to give a more realistic estimate of a typical sample of college and university teachers, the mean scores of ACEPT and non-ACEPT faculty are given. As a further comparison, the mean score of a sample of university faculty teaching education courses for mathematics and science students is also included (Table 4).

As can be seen from this table, the lowest mean scores were those of non-ACEPT science and mathematics faculty teaching content courses. The next highest were those of ACEPT faculty teaching content courses. The highest mean RTOP scores in the entire sample were those of university faculty in the ACEPT

Table 4
A Comparison of the Mean RTOP Scores of Non-ACEPT College and University Faculty With Those of ACEPT Faculty, Including the Teachers of Methods Courses.

	N	mean	*SD*
Non-ACEPT (content courses)	16	37.6	10.8
ACEPT (content courses)	55	61.7	20.9
ACEPT (methods courses)	12	80.1	10.9

project who taught educational methods courses for prospective mathematics and science teachers.

Discussion

Almost every article written about the reform movement in mathematics and science education emphasizes the need for appropriate assessment. Typically, that assessment is characterized as a measurement of some aspect of learning on the part of students. However, that begs the question of the assessment of reformed teaching itself.

Kulm and Stuessy (1992) pointed out that "changes in curriculum goals (also) require concurrent changes in approaches used by teachers in improving learning" (p. 73). Procedures need to be developed to assess curricular innovation and teaching strategies independently of student learning. Just as goals can be set for student learning, so they can be for teaching. Teaching can then be evaluated against these goals. The RTOP provides an instrument for doing just that. By drawing heavily upon the reform literature, a model is established for reformed teaching, not only as understood within the ACEPT program, but as it would be instantiated from a consideration of national standards. The model defines reformed teaching in mathematics and science independently of student achievement, and the RTOP allows observers to arrive at a quantitative characterization of the degree to which such reform has been achieved.

Our data show that when teaching is highly reformed, student learning is significantly enhanced. This relationship between reformed teaching and student achievement holds across a variety of public school grade levels and college and university settings. It also holds across different disciplines in science and in mathematics. The fact that the RTOP is strongly predictive of how much students learn in their classrooms is an important validation of the reform recommendations that were the basis of its development.

The RTOP is a highly inductive instrument. As opposed to many other classroom observation protocols, it calls upon observers to make holistic judgments about broad categories of lesson design and classroom culture. Fortunately, it also appears to tap a kind of judgment call that is familiar to experienced educators. With training, teams of observers can reach high levels of interrater reliability.

Many questions have yet to be answered about reformed teaching in science and mathematics. Most of these require some kind of a definition and means of operationalizing that concept. The Reformed Teaching Observation Protocol offers both a specification of the concept and a means of quantifying that specification. It has potential as a research instrument that in future studies can lead to deeper understandings of the nature of reformed teaching.

References

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans.* Washington, DC: Author.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* Oxford: Oxford University Press.

Anderson, R.D., Anderson, B.L., Varank-Martin, M.A., Romagnano, L., Bielenberg, J., Flory, M., Mieras, A.B., & Whitworth, J. (1994). *Issues of curriculum reform in science, mathematics, and higher order thinking across the disciplines* (Curriculum Reform Project Series 0-16-043073-9). Washington, DC: U.S. Department of Education.

Baker, D.R., & Piburn, M.D. (1997). *Constructing science in middle and secondary school classrooms.* Needham Heights, MA: Allyn and Bacon.

Carlson, M., Buskirk, T., & Halloun, I. (1999). *Assessing college students' views about mathematics with the Views About Mathematics Survey.* Manuscript submitted for publication.

Calhoun, D.O., Bohlin, C., Bohlin, R., & Tracz, S. (March 1997). *The mathematics reform movement: Assessing the degree of reform in secondary mathematics classrooms*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Il.

Frantz, C., Lawrenz, F., Kushner, J., & Millar, S.B. (1998). *Corridors for collaboration*. Arlington, VA: National Science Foundation.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, *30*, 3–19.

Horizon Research Inc. (1998). 1997-98 Local systemic change revised classroom observation protocol. Chapel Hill, NC: Author.

Kulm, G., & Stuessy, C. (1991). Assessment in science and mathematics education reform (Chapter 5). In G. Kulm & S. Malcom (Eds.), *Science assessment in the service of reform*. Washington, DC: American Association for the Advancement of Science.

Lawson, A. E. (1995). Science teaching and the development of thinking. Belmont, CA: Wadsworth Publishing.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, *29*(2), 4-16.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, *21*(1), 29-46.

Mervis, J. (1998). Mixed grades for NSF's bold reform of statewide education. *Science*, *282*(4), 1800-1805.

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

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

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

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

National Institute for Science Education. (1999). [Online]. Available: http://www.wcer.wisc.edu/nise/

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

National Research Council. (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). *Reformed teaching observation protocol (RTOP): Reference manual* (ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers. (Eric Document Reproduction Service, ED 447 205.)

Sawada, D., & Piburn, M., (2000). *Reformed teaching observation protocol (RTOP).* (ACEPT Technical Report No. IN00-1). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

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 Collaborative for Excellence in the Preparation of Teachers.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Shymansky, J.A., & Kyle, W.C., Jr. (1990). *Establishing a research agenda: The critical issues of science curriculum reform.* Paper presented at the National Association for Research in Science Teaching, Atlanta, GA.

Tittle, C.K., & Pape, S. (1996, April). *A framework and classification of procedures for use in evaluation of mathematics and science teaching.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Wetstat*McKenzie Consortium. (1998). *The National Science Foundation's Urban Systemic Initiatives (USI) program: Models of reform of K-12 science and mathematics education.* Washington: The National Science Foundation.

Editors' Note: Contract grant sponsor: National Science Foundation; Contract grant numbers : DUE 9453610 and 0084434.

Daiyo Sawada, Elementary Education, University of Alberta; Michael D. Piburn and Eugene Judson, Science Education, Arizona State University; Jeff Turley, Kathleen Falconer, Russell Benford and Irene Bloom, The Evaluation Facilitation Group, Arizona Collaborative for Excellence in the Preparation of Teachers.

Correspondence concerning this article should be addressed to Michael D. Piburn, Science Education, P.O. Box 870911, Arizona State University, Tempe, AZ 85287-0911.

Electronic mail may be sent via Internet to mike.piburn@asu.edu

Appendix A

Factor Structure of the Reformed Teaching Observation Protocol

	Factor 1	Factor 2	Factor 3
I. Lesson Design and Implementation			
1. The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	**		
2. The lesson was designed to engage students as members of a learning community.	****		
3. In this lesson, student exploration preceded formal presentation.	****		
4. This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	****		
5. The focus and direction of the lesson was often determined by ideas originating with students.	***		
II. Content			
Propositional Knowledge			
6. The lesson involved fundamental concepts of the subject.		****	
7. The lesson promoted strongly coherent conceptual understanding.		***	
8. The teacher had a solid grasp of the subject matter content inherent in the lesson.		**	
9. Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.		*	
10. Connections with other content disciplines and/or real world phenomena were explored and valued.		**	
Procedural Knowledge			
11. Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	**		
12. Students made predictions, estimations and/or hypotheses and devised means for testing them.	****		
13. Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	***		
14. Students were reflective about their learning	***		
15. Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	***		
III. Classroom Culture			
Communicative Interactions			
16. Students were involved in the communication of their ideas to others using a variety of means and media.	***		
17. The teacher's questions triggered divergent modes of thinking.	**		
18. There was a high proportion of student talk and a significant amount of it occurred between and among students.	***		
19. Student questions and comments often determined the focus and direction of classroom discourse.	**		
20. There was a climate of respect for what others had to say.	*		**
Student/Teacher Relationships			
21. Active participation of students was encouraged and valued.	**		*
22. Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	**		
23. In general the teacher was patient with students.			***
24. The teacher acted as a resource person, working to support and enhance student investigations.	****		
25. The metaphor "teacher as listener" was very characteristic of this classroom.	***		

** (0.60 – 0.69), *** (0.70 – 0.79), **** (0.80 – 0.99)