Reformed Teaching Observation Protocol (RTOP):Reference Manual

Michael Piburn and Daiyo Sawada
ACEPT Technical Report No. IN00-3
Arizona Collaborative for Excellence in the Preparation of Teachers

Introduction

The Reformed Teaching Observation Protocol (RTOP) was created by the Evaluation Facilitation Group (EFG) of the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT). It is an observational instrument designed to measure "reformed" teaching. A complete copy of the RTOP can be found in Appendix 2.

The EFG consists of Daiyo Sawada (External Evaluator), Michael Piburn (Internal Evaluator), Bryce Bartley and Russell Benford (Biology), Apple Bloom and Matt Isom (Mathematics), Kathleen Falconer (Physics), Eugene Judson (Beginning Teacher Evaluation) and Jeff Turley (Field Experiences). The hard work and intellectual contributions of all of these people are herein acknowledged. Without their efforts, this work could not have been conducted.

The initial development of the RTOP is now complete, and the instrument is being widely circulated. Consequently, there is a need for a manual that contains the more technical information about the RTOP that might be used by scholars and researchers. This document is designed to fill that need. The theoretical constructs that guided the design of the instrument are presented here, as are reliability and validity information. In addition, the results of an exploratory factor analysis of the RTOP are presented.

The RTOP should not be used for research purposes by untrained observers. The statistical information that is contained here could not have been collected without the help of observers who spent many hours training to achieve high levels of inter-rater reliability. So that others may have similar experiences, a Training Guide (ACEPT Tech Report IN00-2) has been created to assist in the preparation of observers.

The authors welcome others who wish to use the RTOP in their own research. Inquiries and requests for additional information should be directed to mike.piburn@asu.edu or dsawada@ualberta.ca.

Background

The ACEPT Collaborative

The Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) is a program, funded by the National Science Foundation, that was designed to improve the preparation of science and mathematics teachers in elementary and secondary schools. It specifically targets pre-service teachers, but has extended its concerns to encompass the induction (1-3) years of beginning teachers.

- The primary sponsoring organization for ACEPT is Arizona State University (ASU). The Collaborative, however, encompasses a wide variety of pre-college and university educational establishments in the state. These include Northern Arizona University, the University of Arizona, all of the Community Colleges, Diné College (Navajo Community College), the Phoenix Urban Systemic Initiative, and Local Systemic Initiatives in Gilbert and Mesa.

At ASU, the collaborators come from departments in three Colleges. Biology, Chemistry, Geology, Mathematics and Physics were represented from the College of Liberal Arts and Sciences. Both Science and Mathematics education are represented in the College of Education. The College of Engineering is the third college represented in the Collaborative.

The most basic goal of ACEPT, and of all of the NSF funded Collaboratives for Excellence in Teacher Preparation (CETP's), was the *reform* of teacher education. Funding by the NSF had for many years been directed separately to academic departments and colleges of education, and the preparation of teachers had suffered from this artificial dichotomy. One of the important goals of ACEPT was to bring faculties in science and mathematics, engineering and education together in a joint effort. The desired end was that the preparation of teachers would be "seamless", eliminating the many boundaries and barriers between content and the teaching of that content.

University and community college faculty who become involved in ACEPT through collaborative curriculum development efforts and workshops develop new understandings of their role of teachers, as well as of how students learn. ACEPT prepared faculty and students also teach in a more reformed way than those who have not had the ACEPT experience.

There is a very substantial research literature about the induction of teachers into the profession, and the path that they then follow to expertise. We know that ACEPT prepared teachers are different than others who graduate from our institutions, but there is much that is not known, and the unfinished business of ACEPT includes trying to understand and improve that process.

The Reform Movement

Mathematics and science educators are engaged today in a substantial effort of reform. This is evidenced, in part, by the many recommendations being made by professional organizations for standards in mathematics and science and the teaching of those subjects. The ACEPT project is driven by this reform agenda.

There have been many reform movements in education. The most memorable one in mathematics and science education began in 1957, and continued into the 1970's. That period was characterized primarily by a concern for the structure of the disciplines and for engaging students in authentic inquiry. While those concerns remain, the new reform movement has extended its boundaries well beyond the narrower confines of the science and mathematics curriculum revision efforts of that time.

The RTOP was designed to capture the current reform movement, and especially those characteristics that define "reformed teaching." To do that, the authors of the RTOP relied heavily upon research in mathematics and science education and on the new national standards.

Constructivism

The philosophical and theoretical rationale that underlies the modern reform movement in education is called "constructivism" (von Glasersfeld, 1989). This is characterized by an assumption that "knowledge is not transmitted directly from one knower to another, but is actively built up by the learner" (Driver, et al., 1994, pg. 5).

To many educators, the benchmark work on this topic was that of Jean Piaget, who is often spoken of as the "first constructivist" (von Glasersfeld, 1989, pg. 125). In the Piagetian framework, the maturing individual moves through a series of stages in logical reasoning, from those of the youngster to those of the mature adult. This was the underlying construct for much research and curriculum development in both mathematics and science education between the 1960's and the present. Piaget referred to this focus on stages and movement (acceleration) of students through them as the "American question".

In Piagetian theory, learners could engage in new experiences in two contrasting ways. They might *assimilate* new experiences to what they already know, or they could *accommodate* their ideas to incorporate new information. Curricula constructed with these processes in mind often attempted to induce dissonance, or disequilibrium, that was designed to create conceptual conflict and then to help the student resolve that conflict. An example of such a curriculum design might be the well-known "learning cycle" (Lawson, Abraham & Renner, 1989).

Another view of constructivism has been built upon the work of L.S. Vygotsky. His idea that learning is primarily a *socio-linguistic* phenomenon has been hotly disputed among mathematics and science educators but more openly welcomed by language and reading educators. Regardless of their acceptability, his ideas provide the primary rationale for those who propose to invite and listen to new "voices" in the classroom. Vygotskians are interested in curricula that revolve around active student participation in the negotiation and resolution of meaning. Consequently, classroom discourse becomes a major focus of attention in this model.

Going beyond the socio-linguistic, Cobern (1993) argued that "constructivism is an avenue of research that departed from the neo-Piagetian mainstream 20 years ago and has continued on a distinct path of development." He would direct our attention to the role of *culture* in the learning process. Students come to our classrooms with a variety of world-views and preconceptions that they have acquired as much from socio-cultural contexts as from previous mathematics or science classes. The preferred instructional design for socio-cultural constructivists would be one that acknowledges, indeed values, a variety of alternative ideologies.

This synopsis makes one point clear: there are a wide variety of epistemological and ontological stances at play within current conceptions of constructivism. Acknowledging this variety, perhaps a beginning definition of a constructivist classroom would be one in which people are working together to learn. This has been called a "knowledge-building community" (Bereiter & Scardamalia, 1993, pg. 210-216). Such a community would be characterized by many of the elements of constructivism that have already been mentioned. It would be a place where inquiry was conducted. Discourse would be the primary mode by which participants engaged in negotiations of meaning. Cognitive, social and cultural differences among participants would be honored and alternative world-views respected. A high level of rigor, and an accompanying demand for evidence and argument, would be a hallmark of such a community. Conventions would be established for negotiating meaning but only as they facilitated the knowledge-building priorities already honored within the community.

While there is no common agreement among educators about definitions of constructivism, there is a growing unanimity regarding some of the basic elements of reformed teaching. This unanimity is well documented in the latest editions of the mathematics and science standards released by NCTM (2000) and the National Academy of Sciences (1995, 2000).

Science Education

Today's reform movement in science education can be dated approximately to the publication of Project 2061: Science for All Americans (AAAS, 1989). That document was based on recommendations of the National Council on Science and Technology Education, and was the work of Project 2061of the American Association for the Advancement of Science. This was later followed by the publication of the National Science Education Standards (NAS, 1996), prepared by the National Research Council of the National Academy of Sciences.

While many reform documents have been published, those two remain the referents to which others are compared. Across the country, state and local science education syllabi have been created to mirror these standards. More recently, as high-stakes testing has become common at the state level, the standards have increasingly become the criteria against which student performance is judged. Although it is sometimes overlooked, the standards also outlined recommendations for the teaching of science and for the preparation of science teachers.

There is an over-arching demand in the standards that "teaching should be consistent with the nature of scientific inquiry" (AAAS, 1989, p. 147). Good science teaching would (NRC, 1996, pg. 30):

- Start with questions about nature.
- Engage students actively.
- Concentrate on the collection and use of evidence.
- Not separate knowing from finding out.

A considerable body of literature indicates that it is important that science lessons take into consideration the preconceptions that students bring to the classroom. We know that "what students learn is influenced by their existing ideas" (AAAS, 1989, pg. 145). A reformed science lesson would honor students' prior knowledge, and be constructed in such a way as to challenge their ideas. The national standards require that a teacher must "select science content and adapt and design curricula to meet the interests, knowledge, understanding, abilities and experiences of students" (NRC, 1996, pg. 30).

Another principle of reform is that the "progression of learning is usually from the concrete to the abstract" (AAAS, 1989, p. 146). This suggests that a lesson should begin with the active manipulation of physical objects or data before structured abstractions are introduced. This might take the form of laboratory experimentation, or it might involve the use of existing evidence of data sets. Whichever is the case, science teaching should emphasize active student engagement, and allow generalizations to emerge from that engagement.

The authors of the standards recognized that learning does not occur in isolation. In fact, "in successful science classrooms, teachers and students collaborate in the pursuit of ideas, and students quite often initiate new activities relevant to an inquiry" (NRC, 1996, pg. 33). The notion of collaboration, not only between teacher and students, but also among students, is a crucial underpinning of reform. The national standards demand that teachers "select teaching and assessment strategies that support the development of student understanding and nurture a community of science learners" (NRC, 1996, pg. 31).

Project 2061 made it clear that "scientists and engineers work mostly in groups and less often as isolated investigators", and recommended that since learning how to work in groups is an important outcome of science education, "students should gain experience sharing responsibility for learning with each other"(AAAS, 1989, pg. 148). The national standards make essentially the same point, by insisting that "using a collaborative group structure, teachers encourage interdependency among group members, assisting students to work together in small groups…." (NRC, 1996, pg. 36).

A final imperative for reformed teaching is that students engage in activities that call for them to reflect on their own work. In reformed classrooms "students explain and justify their work to themselves and to one another" (NRC, 1995, pg. 33). They "assess the efficacy of their efforts—they evaluate the data they have collected, re-examining or collecting more if necessary, and making statements about the generalizability of their findings. They plan and make presentations to the rest of the class about their work and accept and react to the constructive criticism of others (NRC, 1996, p. 33).

It is almost impossible in a brief statement like this one to do full justice to the recommendations of Project 2061 and the national standards. However, teaching Standard B of the *National Science Education Standards* provides as good a single summary as can be found of the reform recommendations for science teaching. A teacher should (NRC, 1996, p. 32):

- Focus and support inquiries while interacting with students.
- Orchestrate discourse among students about scientific ideas.
- Challenge students to accept and share responsibility for their own learning.
- Recognize and respond to student diversity and encourage all students to participate fully in science learning.
- Encourage and model the skills of scientific inquiry as well as the curiosity, openness to new ideas and data, and skepticism that characterize science.

Mathematics Education

The New Math movement of the 1960's, despite its many significant priorities and practices, left mathematics education in a state of ambivalence. Amid the confusion, piecemeal pedagogical trajectories were established during the seventies and eighties around calls to return to the problem solving agenda first articulated by Polya in the 1940's. Indeed, the 1980 Yearbook of the NCTM was titled, "Let Problem Solving be the Focus of the Eighties". Toward the end of the eighties, the problem-solving thrust received new direction with the publication of NCTM's Curriculum and Evaluation Standards (1989). Professional Standards (1991) and Assessment Standards (1995) followed this in quick succession. Thus began the current standards-based reform movement in mathematics education. These three volumes have been synopsized and synthesized in a single volume titled Principles and Standards for School Mathematics (NCTM, 2000).

More than two years in its gestation, "Standards 2000" as it was called, received widespread input and critique from the mathematics education community the world over. Significant in acknowledging the evolving priorities fomenting in recent years, the synopsis began using the concept of "principles" as well as standards. Thus, the title itself, Principles and Standards for School Mathematics, is a definite indication of reaching beyond "standards" as a way of articulating and guiding reform.

There are six principles, five generic standards, and several specific content standards. The six principles articulate a strongly coherent picture of mathematics reform.

Principles

The Equity Principle: "Excellence in mathematics education requires equity – high expectations and strong support for all students" (p. 12). Equity acknowledges and honors the vast array of culturally, socially, ethnically, racially, and cognitively diverse experience which students necessarily bring with them wherever they go. These differences are not simply tolerated; they are a valuable resource that powers and empowers the reformed teacher and student.

The Curriculum Principle: "A curriculum is more than a collection of activities: it must be coherent, focused on important mathematics, and well articulated across the grades" (p. 14). As well, the curriculum effectively integrates fundamental mathematical concepts so that the student can build and extend ideas through establishing connections with other mathematical ideas as well as interpretations that draw upon concepts from science and other domains including the richness and nuance of everyday phenomena.

The Teaching Principle: "Effective mathematics teaching requires understanding what students know and need to learn and then challenging and supporting them to learn it well" (p. 16). This includes understanding big mathematical ideas in different representational modes, sensing when student thinking might be tapping alternate modes, and taking the risk of pursuing such possibilities as part of a teaching strategy. "The teacher is responsible for creating an intellectual environment where serious mathematical thinking is the norm" (p. 18).

The Learning Principle: "Students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge" (p.20). As confirmed by Bransford, Brown, and Cocking (1999, p. 21) all students have a "knowledge base on which to build, including ideas developed in prior school instruction and those acquired through everyday experience." Furthermore, learning with understanding can be enhanced through classroom discourse in which students propose mathematical ideas and conjectures, evaluate their own thinking as well as that of others, and revise or refine their thoughts.

The Assessment Principle: "Assessment should support the learning of important mathematics and furnish useful information to both teachers and students" (p. 22). In order to do this, assessment must be integrated into instructional and learning experiences, oftentimes becoming indistinguishable from them. Such integration happens most productively when it occurs as a self-reflective process engaged in by students as a natural critique and verification of their own thinking done alone or in the setting of other students engaged in similar reflection.

The Technology Principle: "Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning" (p. 24). Current research strongly supports the view that students can learn mathematics more deeply when technology is used appropriately. Proper use includes enriching "the range and quality of investigations by providing a means of viewing mathematical ideas from multiple perspectives" (p. 25).

In addition to the six principles, Principles and Standards articulate five generic standards. These are very similar to the generic standards proposed in the Curriculum and Evaluation Standards (1989).

Generic Standards

Problem Solving: "Teachers play an important role in the development of students' problem-solving dispositions by creating and maintaining classroom environments, from prekindergarten on, in which students are encouraged to explore, take risks, share failures and successes, and question one another" (p. 53). Principles and Standards go on to say that, "In such supportive environments, students develop confidence in their abilities and a willingness to engage in and explore problems, and they will be more likely to pose problems and to persist with challenging problems" (p. 53).

Reasoning and Proof: "By developing ideas, exploring phenomena, justifying results, and using mathematical conjectures in all content areas and – with different expectations of sophistication – at all grade levels, students should use and expect that mathematics makes sense" (p. 56).

Communication: "Listening to others' explanations gives students opportunities to develop their own understanding. Conversations in which mathematical ideas are explored from multiple perspectives help the participants sharpen their thinking and make connections" (p. 60). Teachers must be aware that in supporting and encouraging student participation in mathematical discourse, it is important to avoid a premature and oftentimes heavy-handed rush to impose formal mathematical language. Patience to allow students to frame the ideas in their mode of thinking is paramount.

Connections: "When students can connect mathematical ideas, their understanding is deeper and more lasting. They can see mathematical connections in the rich interplay among mathematical topics, in contexts that relate mathematics to other subjects, and in their own interests and experience" (p. 64).

Representation: "The importance of using multiple representations should be emphasized throughout students' mathematical education . . . As students become mathematically sophisticated, they develop an increasingly large repertoire of mathematical representations as well as a knowledge of how to use them productively" (p. 69). Significant in the use of multiple representations is the move toward abstraction that brings out the powerful role of mathematics in revealing and operationalizing pattern.

Principles and Standards for School Mathematics begins with a vision of a classroom. It is presented here as a summary: " . . . imagine a classroom . . . Students confidently engage in complex mathematical tasks chosen carefully by teachers. They draw on knowledge from a wide variety of mathematical topics, sometimes approaching the same problem from different mathematical perspectives or representing the mathematics in different ways until they find methods that enable them to make progress. Teachers help students make, refine, and explore conjectures on the basis of evidence and use a variety of reasoning and proof techniques to confirm or disprove those conjectures. . . . Alone or in groups and with access to technology, they work productively and reflectively . . . Orally and in writing, students communicate their ideas and results effectively". (p. 3)

Test Development

Evaluation during the first two years of the ACEPT project was sub-contracted to a group from another state. At the end of the third year, a national search was undertaken for a person to fill a full-time position as ACEPT Project Evaluator. A mathematics educator from another university was identified and hired.

While the search was underway, a small team began an internal evaluation. An ASU faculty member was identified as Internal Evaluator, and two other members of the ACEPT staff were assigned to this group. It met throughout the fall semester, and was engaged in this activity when the new External Evaluator arrived at ASU. The evaluation group was then expanded by assigning graduate assistants to it who were also working with ACEPT collaborators. The term Evaluation Facilitation Group (EFG) was coined by the External Evaluator to name and describe the team that had been created.

Shortly after his arrival, the External Evaluator and other members of his team attended a meeting in Washington in December 1998, sponsored by the National Science Foundation, of CETP evaluators. One of the most salient topics of that meeting seemed to be the question of how to identify “reformed” mathematics and science classes. This question became an important agenda item for subsequent EFG meetings.

A decision was made to develop an observational instrument that would allow the EFG to characterize any classroom on a quantitative scale of reform. Among the many instruments examined was one developed by the Horizon Research Corporation that had been highly recommended at the December 1998 NSF meeting for the consideration of CETP evaluators. Another was a check-sheet contained in a text authored by an ACEPT collaborator (Lawson, 1995). A number of other instruments were also reviewed. None of these focused exclusively on the reformed nature of the classroom – all had other components reflective of “good” teaching more generally such as “lesson closure” or adequate “wait time”.

A first draft the Reformed Teaching Observation Protocol (RTOP) was written in 1998. Because the language of the items in the first draft was particularly referenced toward science teaching, the RTOP was presented to mathematicians in the ACEPT project for review. Major critique and suggestions for revision included suggestions for additional items reflecting mathematical modes of thinking, and an unequivocal request to overhaul the science-dominated language. A member of EFG who was a mathematics educator incorporated the feedback from the mathematicians, devising new items reflecting the mathematics standards as well as the inquiry-based priorities of ACEPT. Additionally, he rewrote each item to be more inclusive of mathematical modes of expression and thinking. Thus began the journey of revision that eventually resulted in an observation instrument with very special qualities.

However, the original structure of the RTOP did not changed. It still consists of 25 items divided into three subsets: *Lesson Design and Implementation (5), Content (10),* and *Classroom Culture (10)*. The second and third subsets are each divided into two smaller groups of five items. The first subset, was designed to capture what had become the ACEPT model for reformed teaching. It describes a lesson that begins with recognition of students’ prior knowledge and preconceptions, that attempts to engage students as members of a learning community, that values a variety of solutions to problems, and that often takes its direction from ideas generated by students. The second subset was directed at content, and was divided into two parts. The first assessed the quality of the content of the lesson, and the second attempted to capture the ACEPT understanding of the process of inquiry. The final subset, consisting of ten items, was directed at the climate of the classroom. It was the authors' intention to capture the full range of ACEPT reformed teaching with these 25 items.

As part of its effort to stimulate interest among undergraduates, an ACEPT competition was created for student teachers. Awards were given for the best teaching of mathematics and/or science at the elementary and secondary levels. Part of the competition required the teaching of videotaped lessons for evaluation. These tapes were used by the EFG for the first formal evaluation of the RTOP.

All members of the evaluation group met together and reviewed tapes using the RTOP. After viewing the tapes, inter-rater reliabilities were computed and the judgments of the reviewers were discussed. This process was continued for three semesters, with new videotaped lessons, and the RTOP items were continuously revised. As this was happening, the External Evaluator began writing a Training Manual that could be used to convey the developing interpretive consensus underlying the increasing reliability estimates.

During the spring of 1999, members of the EFG began visiting university classrooms to make further observations designed to 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 when plans for a summative evaluation were put in place for the Fall Semester. The July 1999 version of the RTOP marked the end of the developmental process. The items in later versions are identical to those in the July 1999 version.

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. A new component to these studies was a very detailed and time-consuming analysis, using the RTOP, of the teaching employed in all of these classes. The sample consisted of mathematics and science classes in middle school, high school, community colleges and universities. A group of nine observers completed 287 RTOP forms over a sample of 153 classrooms. This is the data set upon which this report is based.

Psychometric Properties

Reliability

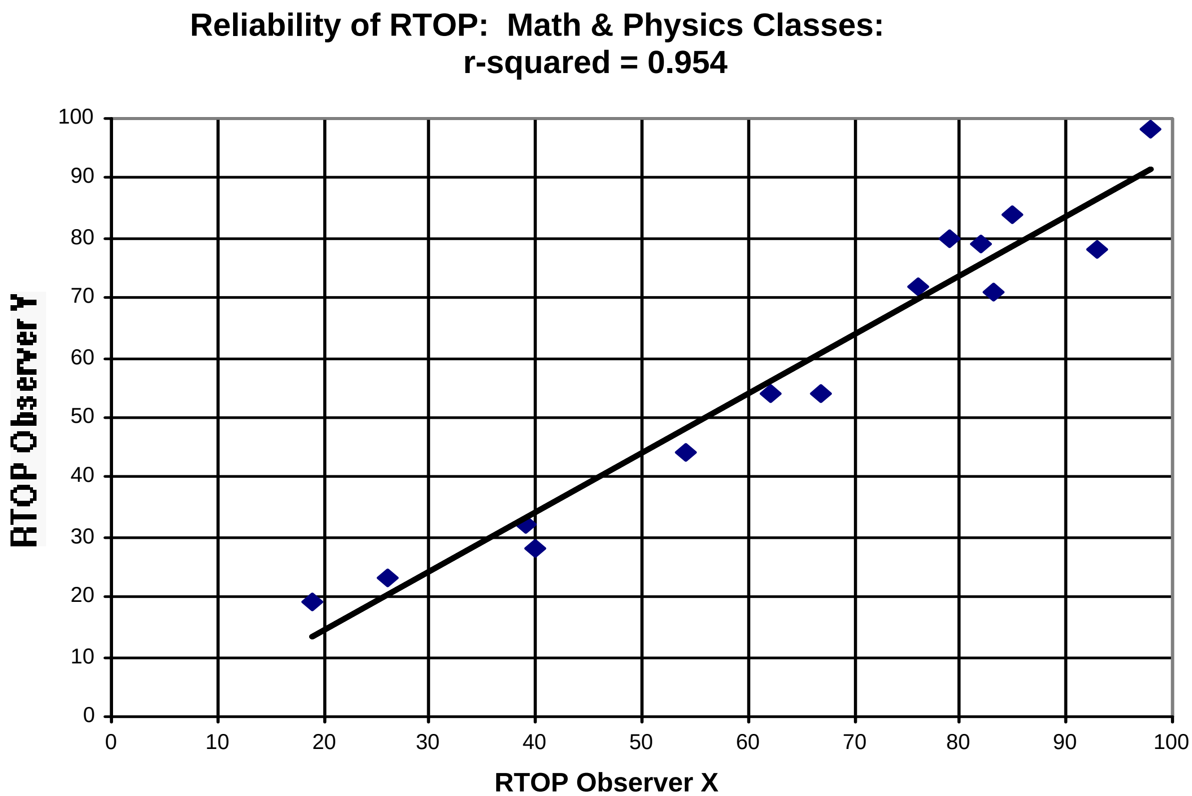
Data Set 1

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 observation toward the close of the course. In order to get an early reading of inter-rater 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 reliability. Two items of technical significance should be noted: (1) 17 pairs were available for analysis but one of the lessons was so strikingly unique it prompted discussion between the two observers. The ratings could no longer be considered "independent" and the observations were excluded from the reliability data; and (2) for three of the paired observations, the instructor was the same but the paired observation was of a lesson taught on a different day but 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 reliability were obtained by doing a best-fit linear regression of one set of observations on the other. Figure 1 shows a scatter plot of the 32 data points (some data points fall on each other). The equation for the best fitting line and the proportion of variance accounted for by that line ($R^2 = 0.954$) are shown. This estimate of reliability, 0.954, is exceptionally high for an instrument of this type.

Figure 1. Reliability estimate of RTOP based on Physics/Math observations



In a similar manner, reliabilities were also estimated for the five subscales that constitute RTOP. Because each subscale consists of only 5 items, it was anticipated that the reliabilities for the subscales would be substantially lower than for the total score. While this was true for Subscale Two, it was not true for the other subscales as shown in Table 1.

Table 1. Reliability Estimates of RTOP Subscales

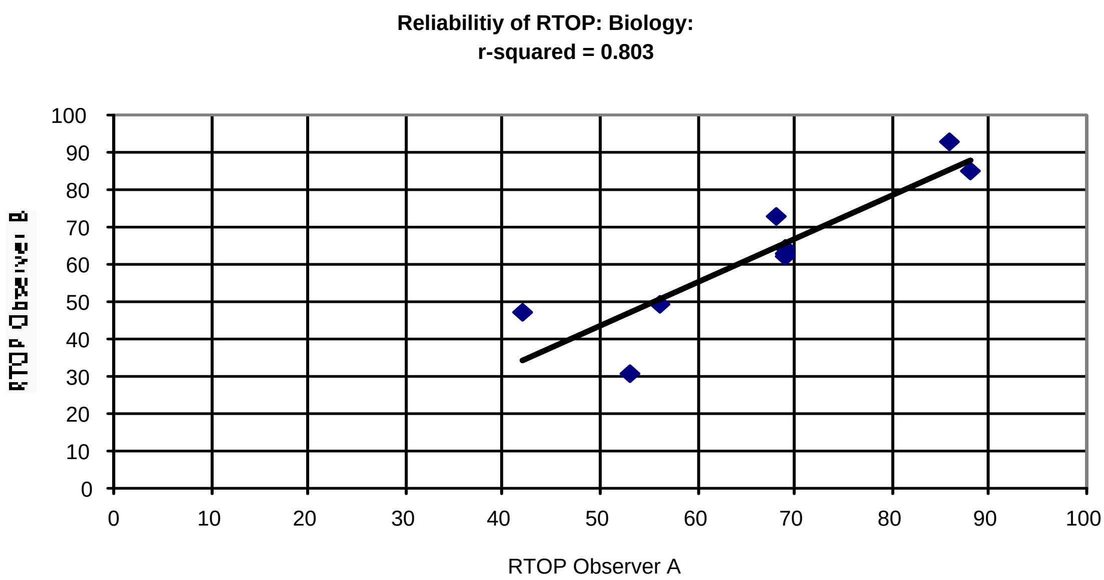
Name of Subscale	R-Squared
Subscale 1: Lesson Design and Implementation	0.915
Subscale 2: Content – Propositional Pedagogic Knowledge	0.670
Subscale 3: Content – Procedural Pedagogic Knowledge	0.946
Subscale 4: Classroom Culture – Communicative Interactions	0.907
Subscale 5: Classroom Culture – Student/Teacher Relationships	0.872

One of the subscales, Subcale 3 ($R^2 = 0.946$) had almost as high a reliability estimate as did the total score (0.954).

Data Set 2

Further data suitable for estimating reliability became available in the fall of 1999 when as part of the Biology evaluation, two members of EFG different from those participating in Data Set 1, gathered RTOP observations on eight biology instructors. While the number of paired observations is not high the correlation coefficient was 0.90. The graph below shows the scatter plot of the observations and the best-fitting line (Figure 2).

Figure 2. Reliability estimate of RTOP based on Biology observations



This second data set appears to confirm the very high reliabilities that paired observers who have received training are able to obtain with the RTOP.

Validity

Face Validity

The Face Validity of RTOP draws on three major sources:

- National Council Teachers of Mathematics. Curriculum and Evaluation Standards (1989), Professional Teaching Standards (1991), Assessment Standards (1995), and Principles and Standards (2000).
- National Academy of Science, National Research Council. National Science Education Standards (1995) and Inquiry and the National Science Education Standards (2000).
- American Association for the Advancement of Science, Project 2061. *Science for All Americans* (1990) *Benchmarks for Scientific Literacy* (1993).

While face validity is a helpful characteristic, it is by no means sufficient. Indeed, it can sometimes be misleading as was revealed during factor analytic studies of the instrument (see below). Construct validity is the critical kind of validity an instrument such as RTOP must possess. Without high construct validity, high reliabilities can be meaningless as well as misleading.

Construct Validity

Construct validity refers to the theoretical integrity of an instrument. The inter-relationships among the contructs in the instrument should give rise to empirical correlations that mirror those theoretical coherences. Because RTOP is a quantitative measure of the degree to which a classroom is in accord with science and mathematics reforms as embodied in the ACEPT project, the theoretical relationships of interest are those underlying ACEPT reform.

The first principles of ACEPT reform are two in number.
1. Inquiry-based, and
2. Standards-based.

These two principles are somewhat different in the way they are usually represented. On the one hand, the standards are diverse, with well over 100 individual content specifications in science and mathematics. On the other hand, "Inquiry-Orientation" is a much more singular notion providing a coherent approach to all subject matter. Inquiry orientation is always said in the singular; standards are always said in the plural.

Thus it would be expected that the RTOP would span several standards but that underlying all these would be a single dimension of "inquiry-orientation". If each of the 25 items in RTOP were independently accessing a different standard, then there could be as many as 25 factors underlying RTOP. However, the RTOP was not designed to represent 25 different standards. It was designed to span a range of standards within the breadth of its five subscales, all the while acknowledging the priority of "inquiry-orientation".

To test the hypothesis that "Inquiry-Orientation" is a powerful integrating force in the structure of RTOP, a correlational analysis was performed on the five subscales. Each subscale was used to predict the total score. High R-squares would support the hypothesis. Low R-squares would serve to reject it. Support for the hypothesis is support for the construct validity of RTOP.

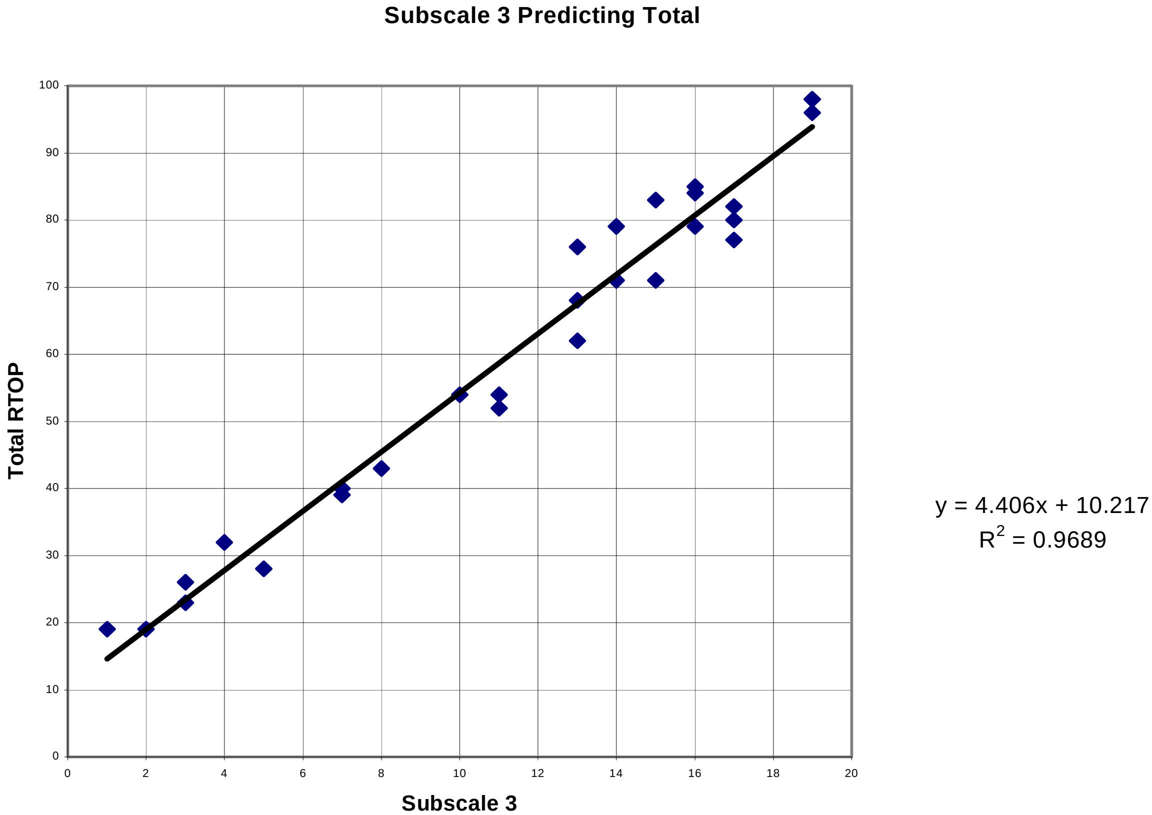
Table 2 provides the R-squares for each subscale as a predictor of the total score. As can be seen, the R-squares approach the reliabilities of each subscale. This offers very strong support for the construct validity of RTOP.

Table 2. Subscales as Predictors of the RTOP Total Score

Predictor	R-squared as Predictor of Total
Subscale 1	0.956
Subscale 2	0.769
Subscale 3	0971
Subscale 4	0967
Subscale 5	0.941

Because Subscale 3 has the highest R-squared as a predictor of the Total RTOP score, a scatter plot of the prediction is shown in Figure 3 to provide a visual feel for the coherence (Inquiry-orientation) underlying the relationship between Subscales (Standards-based) and the total (ACEPT-based reforms).

Figure 3. Subscale 3 as a predictor of the total score.



It is safe to say that four of the five subscales are very good if not excellent predictors of the total score. The other subscale is a good predictor. The construct "Inquiry-Orientation" produces a strong integrative coherence across the many standards. This analysis is presented in support of the construct validity of RTOP.

Predictive Validity

A great deal of evidence has been collected confirming the predictive validity of RTOP in four different instructional settings on Community College and University campuses. In the evaluation of introductory biology, 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). As well, content pre and posttests were given in math, physical science, and physics and a scientific reasoning test was given in biology.

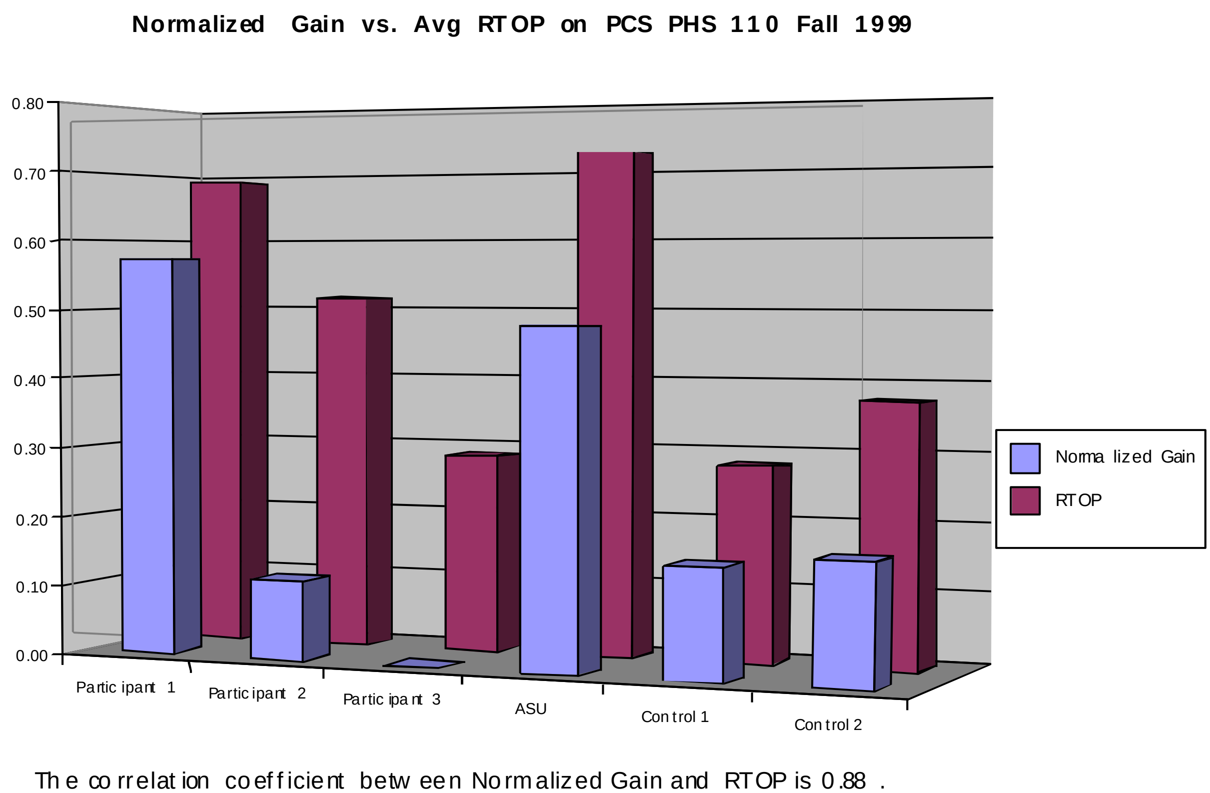
Predicting Gains in Content Achievement

In mathematics, physical science, and physics, multiple instructors were involved. Each instructor was observed a minimum of two times during the fall semester 1999. There were 6 instructors in mathematics, 6 in physical science and 4 in physics. 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 gain as a proportion of the potential gain. It is a score without a unit.

As an example, the RTOP and normalized gain scores for Physical Sciences 110 are presented in Figure 4. It can be seen that the normalized gain falls or increases very much in the same manner as the RTOP score of the instructor of the class.

Figure 4. Covariation of RTOP with Normalized Gain Scores in Physical Sciences 110.



The correlation between RTOP scores and normalized gain scores for these 6 classrooms was 0.88. Despite small sample size a correlation of this magnitude is significant at the 0.01 level. Similar graphs and correlations were obtained in mathematics and physics as shown in Table 3.

Table 3. Correlation Between RTOP and Normalized Gains in Three Subject Areas

Content Area	Correlation or RTOP with Normalized Gain
Mathematics (n = 6)	
Conceptual Understanding	0.94
Number Sense	0.92
Physical Science 110 (n = 6)	0.88
Physics 121 (n = 4)	0.97

Exploratory Factor Analysis

The 25 item RTOP protocol was analyzed using a database containing observations from 153 classrooms. The principal components extraction method and the principal axes extraction method were both performed resulting in very similar analyses (to be expected given the very high reliability estimates). Because the sample size was adequate, the principal components analysis followed by a Varimax rotation is reported here.

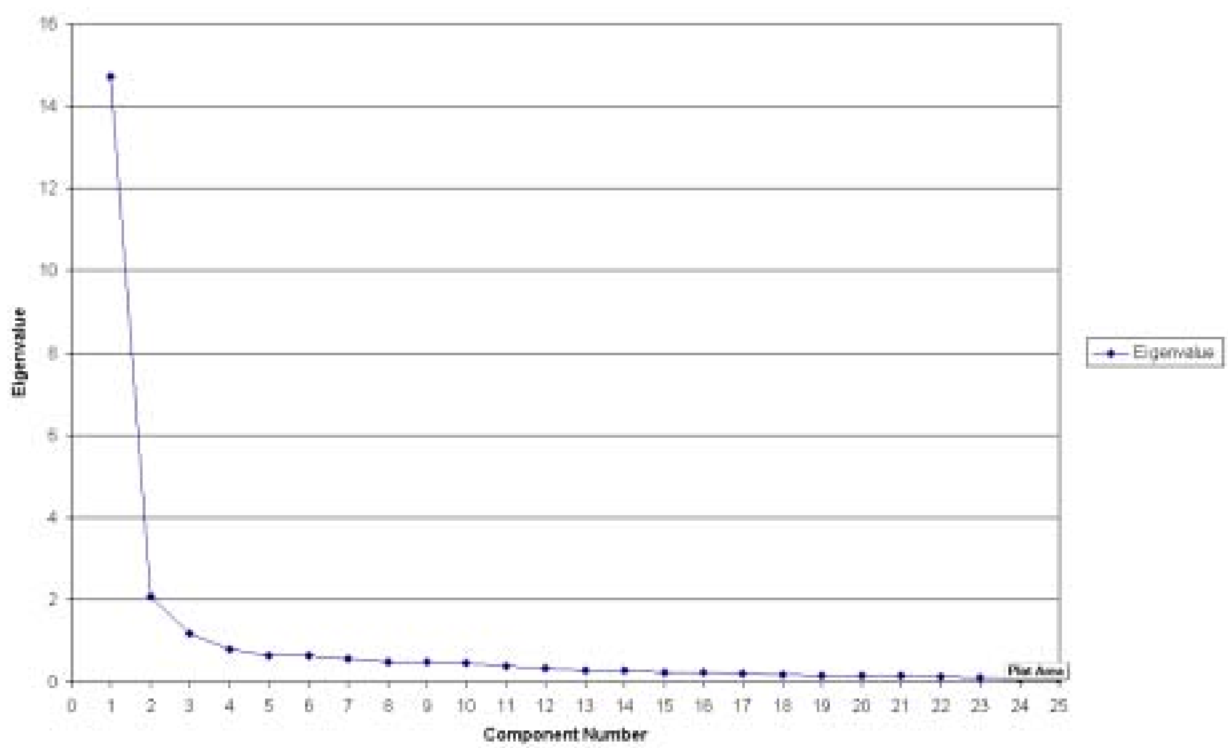
The reliability studies done earlier indicated that the number of components was likely to be small. Solutions asking for two, three, and four principle components to be extracted were run on SPSS resulting in two strong factors and a borderline third factor as shown in Table 4.

Table 4. 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

To confirm whether the third factor with eigenvalue 1.18 was a "legitimate" component, a Scree test was also performed (see Figure 5). It shows that the third component is definitely located in the curvilinear region thus justifying it as a legitimate component. Three factors were therefore retained and interpreted.

Figure 5. Scree Plot



A First Level “Simple Structure” Analysis of the Factor Pattern

To visually and numerically simplify the factor pattern a simple iconic coding was imposed on the coefficients in factor pattern (see Appendix I for the coefficients). Using strings of asterisks to signify the magnitude of a coefficient, a visually more parsimonious pattern is revealed in Table 5. The coding scheme, which only included coefficients equal to or greater than 0.50, is indicated at the bottom of the Table. Given the high magnitude of many of the coefficients, such a high cut-off seemed warranted. The high cut-off also allowed a visual “simple structure” to emerge.

Table 5. Level One Interpretation of the Factor Pattern

RTOP Item	Item No.	Factor 1	Factor 2	Factor 3
The instructional strategies and activities respected students’ prior knowledge and the preconceptions inherent therein.	1	**		
The lesson was designed to engage students as members of a learning community.	2	****		
In this lesson, student exploration preceded formal presentation.	3	****		
This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	4	****		
The focus and direction of the lesson was often determined by ideas originating with students.	5	***		
The lesson involved fundamental concepts of the subject.	6		****	
The lesson promoted strongly coherent conceptual understanding.	7		***	
The teacher had a solid grasp of the subject matter content inherent in the lesson.	8		**	
Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	9		*	
Connections with other content disciplines and/or real world phenomena were explored and valued.	10		**	
Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	11	**		
Students made predictions, estimations and/or hypotheses and devised means for testing them.	12	****		
Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	13	***		
Students were reflective about their learning.	14	***		
Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	15	***		
Students were involved in the communication of their ideas to others using a variety of means and media.	16	***		
The teacher’s questions triggered divergent modes of thinking.	17	**		
There was a high proportion of student talk and a significant amount of it occurred between and among students.	18	***		
Student questions and comments often determined the focus and direction of classroom discourse.	19	**		
There was a climate of respect for what others had to say.	20	*		**
Active participation of students was encouraged and valued.	21	**		*
Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	22	**		
In general the teacher was patient with students.	23			****
The teacher acted as a resource person, working to support and enhance student investigations.	24	****		
The metaphor “teacher as listener” was very characteristic of this classroom.	25	***		

* (0.5 - 0.59), ** (0.60 – 0.69), *** (0.70 – 0.79), **** (0.80 – 0.99)

Factor 1

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

Factor 2

Factor 2, on the other hand, draws exclusively on subscale 2, a subscale that in the instrument is labeled “content propositional knowledge”. Because all five items of the subscale load on this factor, the same label seems appropriate for this factor.

Factor 3

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 original variance, it met both the eigenvalue and scree criterion for inclusion. However, its occurrence 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 relationship”. However, not all of the items in that section loaded on the third factor.

Factor 3 is interpreted here as embodying a concern for “fairness” or “justice” or “democratic rights” or “equity” in the classroom. The student’s voice is recognized as a legitimate source; the student has a role in “agenda-setting”. It is a way of acknowledging value in the preconceptions that students bring with them. If one word had to be used to name Factor 3, it might be “collaboration”.

A Second Level “Finer Structure” Analysis of the Factor Pattern

Initial examination of the RTOP revealed three factors that characterize that instrument. For many purposes, such as interpretation of individual results or computing factor scores for multivariate studies, this level of analysis is adequate.

However, as is usually the case, many items are not uniquely identified with a single factor. It is often useful, after the initial interpretation, to examine the finer structure of the instrument by grouping items into subsets on the basis of factor loadings. This yields smaller groups of items that, although not uniquely identified with a single factor, do add to the interpretive power of the instrument.

The objective of such an analysis is to create groups of items that are similar in the way that their loadings distribute across factors. That is, a group might load most heavily on only one factor, or relatively equally on two factors, or relatively the same on all three factors. Such patterns are ignored when simple structure is the goal. However, just as factors can be identified, characterized and named, so can groupings of items.

A decision-rule for such groupings generally involves a comparison of loadings on separate factors to see if they are similar or different. Although a statistical test for differences between factor loadings is possible, that approach is cumbersome, and was not used here.

Instead, a decision-rule was adopted for this analysis that accepts as meaningful any factor loading greater than 0.30. (Recall that the cut-off for the simple structure analysis was 0.50). A coefficient of 0.30 reflects

an amount of variance shared between item and factor of approximately 10%. That is the same level that is often used as a "rule-of-thumb" criterion for deciding whether a correlation coefficient is meaningful.
A third kind of decision-rule, in which the variances shared between item and factor are compared across factors, was also used. Although the procedure did not produce a substantially different grouping of items than the simpler one that has just been described, it does provide additional information about the strength of a grouping. Thus, in this section, the difference between loadings of items on separate factors is described in terms of multiples of variance shared with each factor. For example, if an item loading on one factor is 0.60 and on another is 0.40, then the variance shared with the former (36%) is about 2 1/2 times the variance shared with the latter (14%). This kind of a comparison reveals the degree to which a particular item should be interpreted as belonging to more than one factor.

Using the above procedures, the most factorially distinct group contains seven items, all with loadings greater than 0.68 on Factor 1, and less than 0.30 on Factors 2 or 3 (Table 6). The smallest amount of variance that any of these items share with Factor 1 (>45%) is more than five times as great as the largest amount of variance that any of them shares with any other factor (8%). Such differences are very large, and the items should be interpreted as strongly uni-factorial.

Table 6. Group 1: Items loading only on Factor 1

FACTOR	1	2	3
3. In this lesson, student exploration preceded formal presentation.	.86	.13	.09
4. This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	.84	.19	.16
11. Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	.68	.16	.19
12. Students made predictions, estimations and/or hypotheses and devised means of testing them.	.83	.27	.03
13. Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	.78	.29	.27
14. Students were reflective about their learning.	.78	.25	.20
16. Students were involved in the communication of their ideas to others using a variety of means and methods.	.75	.22	.27

These items are not from a single subscale of the RTOP. Two (3, 4) come from *Lesson Design and Implementation,* four (11-14) come from *procedural knowledge,* and one (16) comes *from Classroom Culture.* All of them characterize activities of individual students or characteristics of the lesson that typify inquiry in its purest form. Students used "a variety of means to represent phenomena", were engaged in " making predictions, estimations or hypotheses" and " thought-provoking activities," communicated their ideas to others, and were "reflective about their learning." In the lesson, "exploration preceded formal presentation," and students were encouraged to "seek and value alternative modes of investigation or of problem-solving. This group of items is strongly suggestive of a pedagogy of inquiry.

The next set of items consists of three with loadings of 0.64 or greater on Factor 2, and loadings of 0.25 or less on Factors 1 or 3 (Table 7). These items are also factorially distinct. The smallest amount of variance that any shares with Factor 2 (>40%) is more than six times the greatest amount of variance shared with Factors 1 or 3.

Table 7. Group 2: Items loading only on Factor 2

FACTOR	1	2	3
6. The lesson involved fundamental concepts of the discipline.	.06	.82	-.17
7. The lesson promoted strongly coherent conceptual understanding.	.19	.76	.12
10. Connections with other content disciplines and/or real world phenomena were explored and valued.	.25	.64	.25

All three items (6, 7 and 10) are from the *propositional knowledge* portion of the RTOP. They seem to tap the lesson's attention to "fundamental concepts", "conceptual understanding," and "connections" with other contexts. Taken together, this group supports the definition of Factor 2 as predominantly one concerned with the scientific knowledge base contained in the lesson.

A group of four items load on both Factors 1 and 2, although always more heavily on Factor 1 (Table 8). Two of these (1 and 5) came originally from *Lesson Design and Implementation*, one (15) from *Procedural Knowledge,* and one (22) came from *Student/Teacher Relationships.* In this group, the difference between the variance shared between the items and factors 1 and 2 is much smaller, ranging from two-and-one-half to four times.

Table 8. Group 3: Items loading on both Factors 1 and 2

FACTOR	1	2	3
1. The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	.60	.37	.29
5. The focus and direction of the lesson was often determined by ideas originating with the students.	.72	.38	.29
15. Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	.79	.37	.28
22. Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	.69	.43	.24

There seem to be two unifying themes among the items. The first involves a respect for "students' prior knowledge" and the "ideas originating with the students." The second embodies the extent to which the lesson stimulates "criticism and the challenging of ideas," and encouraged students to "generate conjectures, alternative solution strategies, and ways of interpreting evidence. This set of items is best interpreted as representing the intersection between Factors 1 and 2. The implications of this will be discussed further in the summary of this section of the report.

A very strongly related set of six items loads moderately to very heavily both on Factors 1 and 3 (Table 9). This consists of item 3 from *Lesson Design and Implementation* and items 18-25 from *Classroom Culture.* Two of the items (3 and 24) have loadings on Factor 1 of almost seven times those on Factor 3, and so could as reasonably be included in the cluster of items associated with Factor 1 alone. However, the remainder (18, 20, 21 and 25) have loadings that are within one-and-one-half and three times each other. Because all but one of the items come from the same sub-test of the RTOP, they are included together here.

Table 9. Group 4: Items loading on both Factors 1 and 3

FACTOR	1	2	3
2. The lesson was designed to engage students as members of a learning community.	.83	.08	.30
18. There was a high proportion of student talk and a significant amount of it occurred between and among students.	.76	.02	. 46
20.There was a climate of respect for what others had to say.	.50	.16	.69
21. Active participation of students was encouraged and valued.	.66	.24	.57
24.The teacher acted as a resource person, working to support and enhance student investigations.	.82	.02	.32
25.The metaphor "teacher as listener" was very characteristic of this classroom.	.73	.12	.48

All of these items reflect the central notion of a classroom as a place where students work together to learn. This is distinct from the content of a lesson, and goes beyond a more simplistic notion of inquiry. In such a classroom, students are encouraged to participate, to talk among themselves, and to respect what others say. The role of the teacher is to act as a "resource person" and to serve as a "listener."

A final set of four items is very similar to those just mentioned, except that in this case there is a strong tendency for the items to load at very similar weightings across all Factors (Table 10). Two of these items (8 and 9) came from *Content* and two (17 and 19) came from *Classroom Culture.* The largest difference for any single item for variance shared with any two factors is only twice (# 19).

Table 10. Group 5: Items loading on all three Factors

FACTOR	1	2	3
9. Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	.38	.56	.41
17. The teacher's questions triggered divergent modes of thinking.	.60	.46	.43
19. Student questions and comments often determined the focus and direction of classroom discourse.	.65	.40	.42

These items are similar to one-another in describing a divergence of thinking that is triggered by teachers and uses student comments to re-focus the direction of a lesson, while always encouraging elements of abstraction that might maintain some central focus to the lesson.

To this point in the analysis, two items remain ungrouped (Table 11). Because they stand alone, they do not constitute a group together. Both are difficult to interpret. One (#25 from *Classroom Culture)* loads uniquely on Factor 3. It refers only to the "patience" of the teacher. Another (#8 from *Content*) loads on Factors 2 and 3.

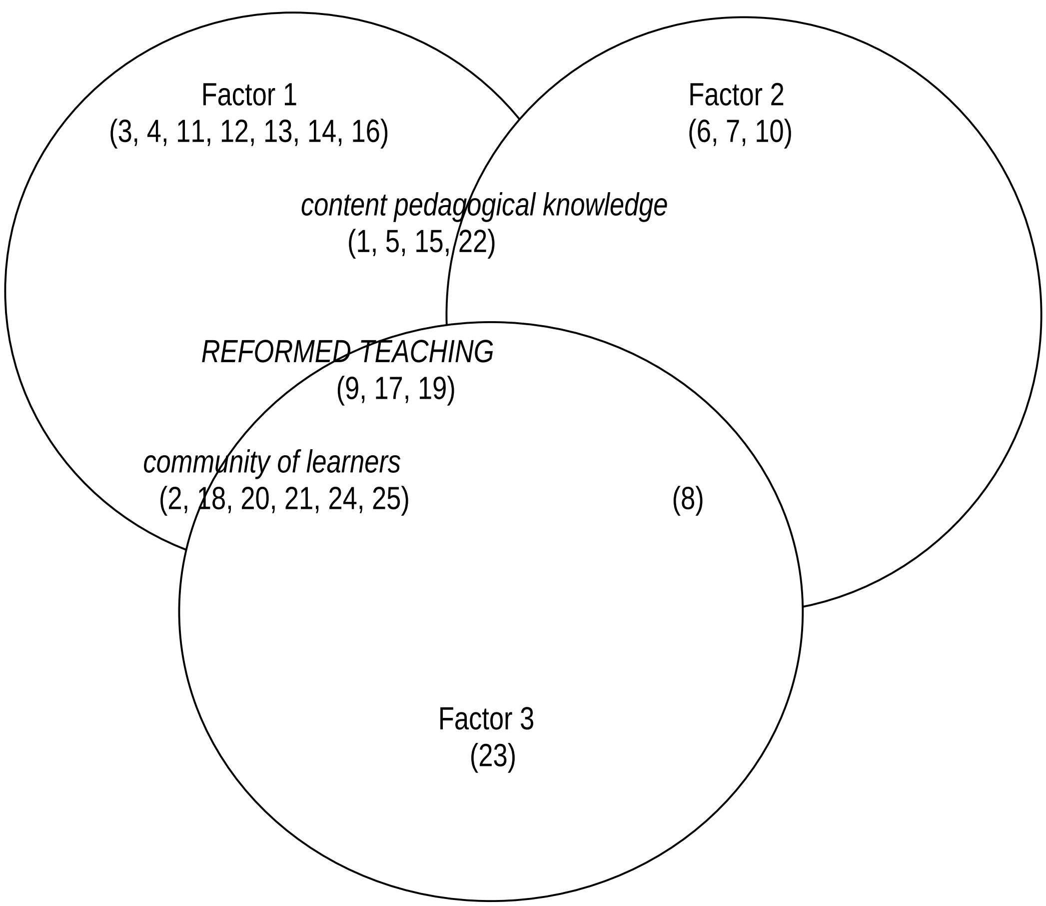
Alternatively, if groups with only 1 item are entertained, each of these items might be interpreted as signaling the existence of potential groups. Taking this stance prompts the question: Why are these groups so sparse? This question is addressed following the presentation of Figure 6 in which these potential groups are indicated.

Table 11. Items not grouped

FACTOR	1	2	3
5. The teacher had a solid grasp of the subject matter content inherent in the lesson.	.08	.64	.44
23. In general, the teacher was patient with students.	.26	.19	.85

A closer examination of the finer structure of the RTOP, accomplished by grouping items with similar patterns of factor loadings, has revealed a set of groups that seems to speak as much to the structure of science classrooms as it does to the RTOP itself. The relationships among items can be displayed most appropriately by the use of a Venn Diagram (Figure 6).

Figure 6. Venn diagram showing relationships among RTOP items



The circle that represents the unique part of Factor 1 is defined by seven items. Taken as a group, these seem to characterize the *pedagogy of inquiry teaching* that is so prominently tapped in the RTOP. There is a separate set of three items which form a group heavily loaded on Factor 2. However, this group contains only those items from the *propositional knowledge* portion of *the Content* subscale of the RTOP. This group of items appears to represent a cluster that could, therefore, be identified as characterizing *propositional knowledge*.

These two groupings (Tables 1 & 2) reveal a particularly important message about the finer structure of science lessons. Although propositional knowledge and procedural knowledge are both contained within the *Content* sub-test of the RTOP, they do not separate the same way in this analysis. In fact, procedural knowledge is intimately tied in with a number of items from *Lesson Design and Implementation,* probably through an underlying construct of inquiry.

However, there is group of four items that exists in the intersection of Factors 1 and 2. Shulman (1986) spoke about a kind of knowledge held by experienced teachers that somehow fused their understanding of content and pedagogy. Insofar as such knowledge is represented in the RTOP, it reflects the teacher's ability to understand students' preconceptions and prior knowledge, and to respect that when designing a lesson. This knowledge allows the teacher to create a lesson with focus and direction that originates with the ideas of students. But it also entails a value for "intellectual rigor, constructive criticism, and the challenging of ideas." However, the respect of the teacher for the ideas of students, as well as a deep understanding of the nature of the propositional knowledge that is the structure of the lesson, can result in the encouragement of conjectures, alternative solutions, and a variety of ways of interpreting evidence. These four items define the meaning of *content pedagogical knowledge* operationally within the RTOP.

There is only one item on the RTOP that loads uniquely on Factor 3. The Item (#23) refers to the "patience" of the teacher. However, there is a subset of six items that loads on both Factor 1 and Factor 3. This is represented in Figure 6 as located at the intersection of the two, and has been named *community of learners.* There seems to be a very intimate relationship between classrooms that can be characterized as learning communities and those that foster inquiry learning. It is possible for inquiry learning to exist in isolation, but apparently the converse is less likely. Learning communities seem of necessity to require an inquiry orientation.

Finally, there is a group of three items that exists at the intersection of Factors 1, 2 and 3. These appear to describe a classroom that is relatively divergent, with the teacher encouraging exploration by students while also structuring the lesson by insisting on abstractions and other organizing devices. Tapping, as they do, all elements of the behaviors described within the RTOP. They appear to define a cluster that could be called *REFORMED TEACHING*.

The above interpretation based upon groups of items has little to say about either the unique part of Factor 3 or about the intersection of Factors 2 and 3. These two locations each have only one item in them and therefore hardly constitute "groups". We return now to the question raised earlier: "Why are these groups so sparsely populated? One interpretation is that these regions are not important to inquiry. For example, if, as suggested earlier, Factor 3 is essentially about "willingness to collaborate", then a pure form of collaboration (collaboration alone) might be seen as inappropriate to reformed teaching. Realizing that “pure” collaboration is much more likely to be found in early childhood classrooms (and almost never in colleges and universities) and noting that the sample for this analysis contains no elementary classrooms, it should not be surprising that this region is almost devoid of items. On the other hand, if the sample had included classrooms from the lower grades, this group might have had more items. This prediction can be tested in further research.

Similar thinking can to applied to the single item (8) identifying the intersection between Factor 2 and 3. The near emptiness of this location indicates that a classroom with strong emphasis on propositional content knowledge but no emphasis on inquiry (pure factor 2) will rarely support collaboration. The two (collaboration and strong emphasis on propositional knowledge) do not mix well at grade levels beyond elementary school. Said another way, the lack of items in the intersection of Factor 2 and 3 might suggest that instructors with a high priority on content don’t feel a need to collaborate unless they have a concern for inquiry. Inquiry is what brings these two priorities together as indicated by the three items in the intersection of all three factors. Again, the relationship might have been different if elementary classrooms had been sampled.

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, norms from the sample used to create the factor analysis for this report are given here (Table 12).
The sample consists of 153 classes. These include 38 classes in mathematics, 51 in science, and 12 in education (methods courses). Among these, 62 were taught at the university level, 26 at community colleges, 37 in high schools and 28 in middle schools. Science and mathematics classes are presented separately in Table 12.

Table 12. Norms for RTOP scores in mathematics and science classrooms by subject and educational level

	Mathematics				Science				Total		
	n	mean	s.d		n	mean	s.d.		n	mean	s.d.
University	10	63.9	22.0		40	58.25	21.3		50	59.4	21.3
C. 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

RTOP scores for this sample ranged from a high of 98 to a low of 18. The mean for the entire sample of 141 classes 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 14).

Table 14. 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	S.D.
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

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

Summary

The Reform Teaching Observation Protocol (RTOP) has proven highly worthwhile in the study of mathematics and science classrooms in middle and high schools, colleges and universities. With appropriate training, it is possible to achieve very high inter-rater reliabilities using this instrument. RTOP scores predict improved student learning in mathematics and science classrooms at all levels. Analysis of the RTOP suggests that it is largely a uni-factorial instrument that taps a single construct of inquiry. A finer-scale analysis lends new meaning to the phrases “pedagogical content knowledge” and “community of learners.” The instrument seems amply able to measure what it purports to measure reformed teaching.

References

AAAS (1989). *Project 2061: Science for All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics, and Technology.* Washington, D.C.: American Association for the Advancement of Science.

Bereiter, C. & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, Ill: Open Court.

Bransford, J.D., Brown, A.L. & Cocking, R.R. (1999). *How People Learn: Brain, Mind, Experience, and School.* Washington, D.C.: National Academy Press.

Cobern, W. (1993). Contextual constructivism: The impact of culture on the learning and teaching of science. Chapter 4 *In* K. Tobin (Ed.). *The Practice of Constructivism in Science Education.* Hillsdale, NJ: Lawrence Erlaum.

Driver, R., Asako, H., Leach, J., Mortimer, E. & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*, 5-12.

Lawson, A. E. (1995). *Science teaching and the development of thinking.* Belmont, CA: Wadsworth Publishing. Pp. 206-208.

Lawson, A., Abraham, M. & Renner, J. (1989). *A theory of instruction: Using the learning cycle to teach science concepts and thinking skills.* NARST Monograph #1. National Association for Research in Science Teaching.

NCTM (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: NCTM

NCTM (1991). *Professional Standards for Teaching Mathematics.* Reston, VA: NCTM

NCTM (1995). *Assessment Standards for School Mathematics.* Reston, VA: NCTM

NCTM (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM

NRC (1996). *National Science Education Standards.* Washington, D.C.: National Academy Press.

NRC (2000). *Inquiry and the National Science Education Standards.* Washington, D.C.: National Academy Press.

Sawada, D., Piburn, M., Falconer, K., Turley, J., Benford, R., & Bloom, I. (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.

von Glasersfeld (1989). Cognition, construction of knowledge, and teaching. *Synthese,* 80(1), 121-140.

Appendix 1. Matrix of Factor Pattern Coefficients

RTOP Item	Item No	Factor1	Factor2	Factor 3
The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	1	.60	.37	.29
The lesson was designed to engage students as members of a learning community.	2	.83	.08	.30
In this lesson, student exploration preceded formal presentation.	3	.86	.13	.09
This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	4	.84	.19	.16
The focus and direction of the lesson was often determined by ideas originating with students.	5	.72	.38	.29
The lesson involved fundamental concepts of the subject.	6	.06	.82	-.17
The lesson promoted strongly coherent conceptual understanding.	7	.19	.76	.12
The teacher had a solid grasp of the subject matter content inherent in the lesson.	8	.08	.64	.43
Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	9	.38	.56	.41
Connections with other content disciplines and/or real world phenomena were explored and valued.	10	.25	.64	.25
Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	11	.68	.16	.19
Students made predictions, estimations and/or hypotheses and devised means for testing them.	12	.83	.27	.03
Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	13	.78	.29	.27
Students were reflective about their learning.	14	.78	.25	.20
Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	15	.79	.37	.28
Students were involved in the communication of their ideas to others using a variety of means and media.	16	.75	.22	.27
The teacher's questions triggered divergent modes of thinking.	17	.60	.46	.43
There was a high proportion of student talk and a significant amount of it occurred between and among students.	18	.76	.02	.46
Student questions and comments often determined the focus and direction of classroom discourse.	19	.65	.40	.42
There was a climate of respect for what others had to say.	20	.50	.16	.69
Active participation of students was encouraged and valued.	21	.66	.24	.57
Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	22	.69	.43	.22
In general the teacher was patient with students.	23	.26	.19	.85
The teacher acted as a resource person, working to support and enhance student investigations.	24	.82	.02	.32
The metaphor "teacher as listener" was very characteristic of this classroom.	25	.73	.12	.48

Appendix II
Reformed Teaching Observation Protocol (RTOP)

Daiyo Sawada — External Evaluator
Michael Piburn — Internal Evaluator

and

Kathleen Falconer, Jeff Turley, Russell Benford and Irene Bloom
Evaluation Facilitation Group (EFG)

Technical Report No. IN00-1
Arizona Collaborative for Excellence in the Preparation of Teachers
Arizona State University

I. BACKGROUND INFORMATION

Name of teacher ______________________ Announced Observation? ______________________ (yes, no, or explain)

Location of class ____ (district, school, room)

Years of Teaching ______________ Teaching Certification ______________________ (K-8 or 7-12)

Subject observed ______________________ Grade level ______________________

Observer ______________________ Date of observation ______________________

Start time ______________________ End time ______________________

II. CONTEXTUAL BACKGROUND AND ACTIVITIES

In the space provided below please give a brief description of the lesson observed, the classroom setting in which the lesson took place (space, seating arrangements, etc.), and any relevant details about the students (number, gender, ethnicity) and teacher that you think are important. Use diagrams if they seem appropriate.

Record here events that may help in documenting the ratings.

Time	Description of Events

III. LESSON DESIGN AND IMPLEMENTATION

		Never Occurred				Very Descriptive
1)	The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	0	1	2	3	4
2)	The lesson was designed to engage students as members of a learning community.	0	1	2	3	4
3)	In this lesson, student exploration preceded formal presentation.	0	1	2	3	4
4)	This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	0	1	2	3	4
5)	The focus and direction of the lesson was often determined by ideas originating with students.	0	1	2	3	4

IV. CONTENT

Propositional knowledge

6)	The lesson involved fundamental concepts of the subject.	0	1	2	3	4
7)	The lesson promoted strongly coherent conceptual understanding.	0	1	2	3	4
8)	The teacher had a solid grasp of the subject matter content inherent in the lesson.	0	1	2	3	4
9)	Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	0	1	2	3	4
10)	Connections with other content disciplines and/or real world phenomena were explored and valued.	0	1	2	3	4

Procedural Knowledge

11)	Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	0	1	2	3	4
12)	Students made predictions, estimations and/or hypotheses and devised means for testing them.	0	1	2	3	4
13)	Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	0	1	2	3	4
14)	Students were reflective about their learning.	0	1	2	3	4
15)	Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	0	1	2	3	4

Continue recording salient events here.

Time	Description of Events

V. CLASSROOM CULTURE

Communicative Interactions		Never Occurred				Very Descriptive
16)	Students were involved in the communication of their ideas to others using a variety of means and media.	0	1	2	3	4
17)	The teacher's questions triggered divergent modes of thinking.	0	1	2	3	4
18)	There was a high proportion of student talk and a significant amount of it occurred between and among students.	0	1	2	3	4
19)	Student questions and comments often determined the focus and direction of classroom discourse.	0	1	2	3	4
20)	There was a climate of respect for what others had to say.	0	1	2	3	4

Student/Teacher Relationships

21)	Active participation of students was encouraged and valued.	0	1	2	3	4
22)	Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	0	1	2	3	4
23)	In general the teacher was patient with students.	0	1	2	3	4
24)	The teacher acted as a resource person, working to support and enhance student investigations.	0	1	2	3	4
25)	The metaphor "teacher as listener" was very characteristic of this classroom.	0	1	2	3	4

Additional comments you may wish to make about this lesson.

Reformed Teaching Observation Protocol (RTOP) TRAINING GUIDE

Daiyo Sawada External Evaluator
Michael Piburn Internal Evaluator
and

Jeff Turley, Kathleen Falconer, Russell Benford, Irene Bloom, and Eugene Judson

The Evaluation Facilitation Group

Arizona Collaborative for Excellence in the Preparation of Teachers
Arizona State University

ACEPT Technical Report No. IN00-2

The Reformed Teaching Observation Protocol (RTOP) is an observational instrument that can be used to assess the degree to which mathematics or science instruction is “reformed.” It embodies the recommendations and standards for the teaching of mathematics and science that have been promulgated by professional societies of mathematicians, scientists and educators.

The RTOP was designed, piloted and validated by the Evaluation Facilitation Group of the Arizona Collaborative for Excellence in the Preparation of Teachers. Those most involved in that effort were Daiyo Sawada (External Evaluator), Michael Piburn (Internal Evaluator), Bryce Bartley and Russell Benford (Biology), Apple Bloom and Matt Isom (Mathematics), Kathleen Falconer (Physics), Eugene Judson (Beginning Teacher Evaluation), and Jeff Turley (Field Experiences).

The instrument draws on the following sources:

- National Council for the Teaching of Mathematics. *Curriculum and Evaluation Standards* (1989), *Professional Teaching Standards* (1991), and *Assessment Standards* (1995).
- National Academy of Science, National Research Council. *National Science Education Standards* (1995).
- American Association for the Advancement of Science, Project 2061. *Science for All Americans*(1990), *Benchmarks for Scientific Literacy*(1993).

It also reflects the ideas of all ACEPT Co-Principal Investigators, but especially those of Marilyn Carlson and Anton Lawson, and the principles of reform underlying the ACEPT project. Its structure reflects some elements of the *Local Systemic Change Revised Classroom Observation Protocol* , by Horizon Research (1997-98).

The RTOP is criterion-referenced, and observers’ judgments should *not* reflect a comparison with any other instructional setting than the one being evaluated. It can be used at all levels, from primary school through university. The instrument contains twenty-five items, with each rated on a scale from 0 (not observed) to 4 (very descriptive). Possible scores range from 0 to 100 points, with higher scores reflecting a greater degree of reform.

The RTOP was designed to be used by trained observers. This *Training Guide* provides specific information pertinent to the interpretation of individual items in the protocol. It is intended to be used as part of a formal training program in which trainees observe actual classrooms or videotapes of classrooms, and discuss their observations with others. The *Guide*, in its present form, is also designed to solicit trainee thoughts and concerns so that they feel comfortable in using the instrument. For that reason, a space is provided after each item for trainee comments. Such input helps all those being trained to achieve a higher degree of consistency in using the instrument. Please keep this in mind in making comments.

I. BACKGROUND INFORMATION

This section contains space for standard information that should be recorded by all observers. It will serve to identify the classroom, the instructor, the lesson observed, the observer, and the duration of the observation.

comments:

II. CONTEXTUAL BACKGROUND AND ACTIVITIES

Space is provided for a brief description of the lesson observed, the setting in which the lesson took place (space, seating arrangements, etc.), and any relevant details about the students (number, gender, ethnicity, etc.) and instructor. Try to go beyond a simple description. Capture, if you can, the defining characteristics of this situation that you believe provide the most important context for understanding what you will describe in greater detail in later sections. Use diagrams if they seem appropriate.

comments:

The next three sections contain the items to be rated. Do not feel that you have to complete them during the actual observation period. Space is provided on the facing page of every set of evaluations for you to make notes while observing. Immediately *after the lesson*, draw upon your notes and complete the ratings. For most items, a valid judgment can be rendered only after observing the entire lesson. The whole lesson provides contextual reference for rating each item.

Each of the items is to be rated on a scale ranging from 0 to 4. Choose "0" if in your judgment, the characteristic *never* occurred in the lesson, not even once. If it did occur, even if only once, "1" or higher should be chosen. Choose "4" only if the item was very descriptive of the lesson you observed. Intermediate ratings do not reflect the number of times an item occurred, but rather the degree to which that item was *characteristic* of the lesson observed.

The remainder of this Training Guide attempts provides a clarification of each RTOP item and the subtest (there are five) of which it is a part.

III. LESSON DESIGN AND IMPLEMENTATION

1) The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.

A cornerstone of reformed teaching is taking into consideration the prior knowledge that students bring with them. The term "respected" is pivotal in this item. It suggests an attitude of curiosity on the teacher's part, an active solicitation of student ideas, and an understanding that much of what a student brings to the mathematics or science classroom is strongly shaped and conditioned by their everyday experiences.

comments:

2) The lesson was designed to engage students as members of a learning community.

Much knowledge is socially constructed. The setting within which this occurs has been called a "learning community." The use of the term community in the phrase "the scientific community" (a "self-governing" body) is similar to the way it is intended in this item. Students participate actively, their participation is integral to the actions of the community, and knowledge is negotiated within the community. It is important to remember that a group of learners does not necessarily constitute a "learning community."

comments:

3) In this lesson, student exploration preceded formal presentation.

Reformed teaching allows students to build complex abstract knowledge from simpler, more concrete experience. This suggests that any formal presentation of content should be preceded by student exploration. This does not imply the converse...that all exploration should be followed by a formal presentation

comments:

4) This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.

Divergent thinking is an important part of mathematical and scientific reasoning. A lesson that meets this criterion would not insist on only one method of experimentation or one approach to solving a problem. A teacher who valued alternative modes of thinking would respect and actively solicit a variety of approaches, and understand that there may be more than one answer to a question.

comments:

5) The focus and direction of the lesson was often determined by ideas originating with students.

If students are members of a true learning community, and if divergence of thinking is valued, then the direction that a lesson takes can not always be predicted in advance. Thus, planning and executing a lesson may include contingencies for building upon the unexpected. A lesson that met this criterion might not end up where it appeared to be heading at the beginning.

comments:

IV. CONTENT

Knowledge can be thought of as having two forms: knowledge of what is (Propositional Knowledge), and knowledge of how to (Procedural Knowledge). Both are types of content. The RTOP was designed to evaluate mathematics or science lessons in terms of both.

Propositional Knowledge

This section focuses on the level of significance and abstraction of the content, the teacher's understanding of it, and the connections made with other disciplines and with real life.

6) The lesson involved fundamental concepts of the subject.

The emphasis on "fundamental" concepts indicates that there were some significant scientific or mathematical ideas at the heart of the lesson. For example, a lesson on the multiplication algorithm can be anchored in the distributive property. A lesson on energy could focus on the distinction between heat and temperature.

comments:

7) The lesson promoted strongly coherent conceptual understanding.

The word "coherent" is used to emphasize the strong inter-relatedness of mathematical and/or scientific thinking. Concepts do not stand on their own two feet. They are increasingly more meaningful as they become integrally related to and constitutive of other concepts.

comments:

8) The teacher had a solid grasp of the subject matter content inherent in the lesson.

This indicates that a teacher could sense the potential significance of ideas as they occurred in the lesson, even when articulated vaguely by students. A solid grasp would be indicated by an eagerness to pursue student's thoughts even if seemingly unrelated at the moment. The grade-level at which the lesson was directed should be taken into consideration when evaluating this item.

comments:

9) Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.

Conceptual understanding can be facilitated when relationships or patterns are represented in abstract or symbolic ways. Not moving toward abstraction can leave students overwhelmed with trees when a forest might help them locate themselves.

comments:

10) Connections with other content disciplines and/or real world phenomena were explored and valued.

Connecting mathematical and scientific content across the disciplines and with real world applications tends to generalize it and make it more coherent. A physics lesson on electricity might connect with the role of electricity in biological systems, or with the wiring systems of a house. A mathematics lesson on proportionality might connect with the nature of light, and refer to the relationship between the height of an object and the length of its shadow.

comments:

Procedural Knowledge

This section focuses on the kinds of processes that students are asked to use to manipulate information, arrive at conclusions, and evaluate knowledge claims. It most closely resembles what is often referred to as mathematical thinking or scientific reasoning.

11) Students used a variety of means (models, drawings, graphs, symbols, concrete materials, manipulatives, etc.) to represent phenomena.

Multiple forms of representation allow students to use a variety of mental processes to articulate their ideas, analyze information and to critique their ideas. A "variety" implies that at least two different means were used. Variety also occurs within a given means. For example, several different kinds of graphs could be used, not just one kind.

comments:

12) Students made predictions, estimations and/or hypotheses and devised means for testing them.

This item does not distinguish among predictions, hypotheses and estimations. All three terms are used so that the RTOP can be descriptive of both mathematical thinking and scientific reasoning. Another word that might be used in this context is "conjectures". The idea is that students explicitly state what they think is going to happen before collecting data.

comments:

13) Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.

This item implies that students were not only actively doing things, but that they were also actively thinking about how what they were doing could clarify the next steps in their investigation.

comments:

14) Students were reflective about their learning.

Active reflection is a meta-cognitive activity that facilitates learning. It is sometimes referred to as "thinking about thinking." Teachers can facilitate reflection by providing time and suggesting strategies for students to evaluate their thoughts throughout a lesson. A review conducted by the teacher may not be reflective if it does not induce students to *re-examine* or *re-assess* their thinking.

comments:

15) Intellectual rigor, constructive criticism, and the challenging of ideas were valued.

At the heart of mathematical and scientific endeavors is rigorous debate. In a lesson, this would be achieved by allowing a variety of ideas to be presented, but insisting that challenge and negotiation also occur. Achieving intellectual rigor by following a narrow, often prescribed path of reasoning, to the exclusion of alternatives, would result in a low score on this item. Accepting a variety of proposals without accompanying evidence and argument would also result in a low score.

comments:

V. CLASSROOM CULTURE

This section addresses a separate aspect of a lesson, and completing these items should be done independently of any judgments on preceding sections. Specifically the design of the lesson or the quality of the content should not influence ratings in this section. Classroom culture has been conceptualized in the RTOP as consisting of: (1) Communicative Interactions, and (2) Student/Teacher Relationships. These are not mutually exclusive categories because all communicative interactions presuppose some kind of relationship among communicants.

Communicative Interactions

Communicative interactions in a classroom are an important window into the culture of that classroom. Lessons where teachers characteristically speak and students listen are not reformed. It is important that students be heard, and often, and that they communicate with one another, as well as with the teacher. The nature of the communication captures the dynamics of knowledge construction in that community. Recall that communication and community have the same root.

16) Students were involved in the communication of their ideas to others using a variety of means and media.

The intent of this item is to reflect the communicative richness of a lesson that encouraged students to contribute to the discourse and to do so in more than a single mode (making presentations, brainstorming, critiquing, listening, making videos, group work, etc.). Notice the difference between this item and item 11. Item 11 refers to representations. This item refers to active communication.

comments:

17) The teacher's questions triggered divergent modes of thinking.

This item suggests that teacher questions should help to open up conceptual space rather than confining it within predetermined boundaries. In its simplest form, teacher questioning triggers divergent modes of thinking by framing problems for which there may be more than one correct answer or framing phenomena that can have more than one valid interpretation.

comments:

18) There was a high proportion of student talk and a significant amount of it occurred between and among students.

A lesson where a teacher does most of the talking is not reformed. This item reflects the need to increase both the amount of student talk and of talk among students. A "high proportion" means that at any point in time it was as likely that a student would be talking as that the teacher would be. A "significant amount" suggests that critical portions of the lesson were developed through discourse among students.

comments:

19) Student questions and comments often determined the focus and direction of classroom discourse.

This item implies not only that the flow of the lesson was often influenced or shaped by student contributions, but that once a direction was in place, students were crucial in sustaining and enhancing the momentum.

comments:

20) There was a climate of respect for what others had to say.

Respecting what others have to say is more than listening politely. Respect also indicates that what others had to say was actually heard and carefully considered. A reformed lesson would encourage and allow every member of the community to present their ideas and express their opinions without fear of censure or ridicule.

comments:

Student/Teacher Relationships

21) Active participation of students was encouraged and valued.

This implies more than just a classroom full of active students. It also connotes their having a voice in how that activity is to occur. Simply following directions in an active manner does not meet the intent of this item. Active participation implies agenda-setting as well as "minds-on" and "hands-on".

comments:

22) Students were encouraged to generate conjectures, alternative solution strategies, and/or different ways of interpreting evidence.

Reformed teaching shifts the balance of responsibility for mathematical of scientific thought from the teacher to the students. A reformed teacher actively encourages this transition. For example, in a mathematics lesson, the teacher might encourage students to find more than one way to solve a problem. This encouragement would be highly rated if the whole lesson was devoted to discussing and critiquing these alternate solution strategies.

comments:

23) In general the teacher was patient with students.

Patience is not the same thing as tolerating unexpected or unwanted student behavior. Rather there is an anticipation that, when given a chance to play itself out, unanticipated behavior can lead to rich learning opportunities. A long "wait time" is a necessary but not sufficient condition for rating highly on this item.

comments:

24) The teacher acted as a resource person, working to support and enhance student investigations.

A reformed teacher is not there to tell students what to do and how to do it. Much of the initiative is to come from students, and because students have different ideas, the teacher's support is carefully crafted to the idiosyncrasies of student thinking. The metaphor, "guide on the side" is in accord with this item.

comments:

25) The metaphor "teacher as listener" was very characteristic of this classroom.

This metaphor describes a teacher who is often found helping students use what they know to construct further understanding. The teacher may indeed talk a lot, but such talk is carefully crafted around understandings reached by actively listening to what students are saying. "Teacher as listener" would be fully in place if "student as listener" was reciprocally engendered.

comments:

VI. SUMMARY

The RTOP provides an operational definition of what is meant by "reformed teaching." The items arise from a rich research-based literature that describes inquiry-oriented standards-based teaching practices in mathematics and science. However, this training guide does not cite research evidence. Rather it describes each item in a more metaphoric way. Our experience has been that these items have richly intuitive meaning to mathematics and science educators .

Further information about the underlying conceptual and theoretical basis of the RTOP, as well as reliability and validity data and norms by grade-level and context, can be found in the *Reformed Teaching Observation Protocol MANUAL* (Sawada & Piburn, 2000).