The Development of Instruments for Assessment of Instructional Practices in Standards-Based Teaching

Camille L. Wainwright, PhD
College of Education
Pacific University, 2043 College Way
Forest Grove, OR 97116
503-352-2963 (Office); 503-352-2907 (Fax)
wainwric@pacificu.edu

Lawrence Flick, PhD
Mathematics and Science Education, 263 Weniger Hall
Oregon State University; Corvallis, OR 97331
541-737-3664; flickl@ucs.orst.edu

Patricia Morrell, PhD
School of Education, University of Portland
5000 N. Willamette Blvd., Portland, OR 97203
503-943-8013; morrell@up.edu

Abstract:

We provide a description and rationale for the development of two instruments – 1) a classroom observation protocol, and 2) a teacher interview protocol – designed to document the impact of reform-based professional development with undergraduate mathematics and science faculty and its impact on the resultant preparation of teachers. Constructed upon review of the research on teaching and standards documents in mathematics and science, these instruments form the basis for data collection in a three-year longitudinal study of teaching practice among early career teachers as well as undergraduate college faculty. In addition, we suggest further applications of the observation protocol beyond the original purpose of our research study.

Introduction:

In 1997 the Oregon Collaborative for Excellence in the Preparation of Teachers (OCEPT) was established with funding from the National Science Foundation (DUE/9996453). It was designed to improve the preparation of science and mathematics teachers in elementary, middle and high schools and to attract a more diverse group of students to the teaching profession.

College level mathematics and science courses tend to promote the success of those who major in the subject and find the subject matter intrinsically interesting, thus limiting the number of students who enroll in these courses. Elementary and middle level teachers are expected to teach mathematics/science to all students at crucial points in our educational system. Thus, pre-service teachers form an important population who ought to enroll in numerous content courses and should, ideally, enjoy these valuable mathematics and science experiences.

Making content courses both more effective and more inviting for a broader range of students is an important goal in the development of a mathematics/science literate society; this is especially critical in the preparation of future teachers. More effective teaching and assessment methods that will motivate and challenge students who are not majoring in mathematics/science and may not find these content areas intrinsically interesting have a research base in the literature of mathematics and science education (Sternberg, 1997; Tobias, 1992; Holton & Horton, 1996). However, the educational research literature in higher education is at an early stage of development, and includes studies which are qualitative in design and diverse in perspective. Methods for more effective teaching and assessment have also been highlighted in recent reform documents in mathematics and science education which are primarily focused on K-12 education. [NSF (1996); AAAS (1989, 1993); NRC (1996, 2000); NCTM (2000)]

OCEPT was designed to foster innovations in the teaching and assessment of college level mathematics and science courses. Prospective elementary, middle level and secondary teachers taking these courses will have firsthand experience in learning mathematics and science through the modeling of strategies and technologies that should not only benefit them as learners, but should also support more effective pedagogy when they begin their own teaching. They should view these courses as a valuable component in their preparation for classroom teaching.

As OCEPT approached its fifth and final year, a variety of evaluation strategies were developed in order to determine its effectiveness. Numerous methods were implemented, including a) the development of case studies at core institutions, b) tracking Faculty Fellows' professional development, and c) collecting data on supply and demand trends within the state as well as quantitative data on the number of teachers entering the profession from underrepresented groups. In addition, an Outcomes Research Study was designed to determine the impact OCEPT Fellows and their courses have had on the quality of newly-licensed Oregon teachers.

The Outcomes Research Study:

The specific research study questions sought to be answered by the Outcomes Research Study are:

1. What is the relationship between student teachers' instructional practices and their undergraduate preparation?

2. How did Faculty Fellows' participation in OCEPT contribute to their instructional design and practice?

3. How do student teachers'/Faculty Fellows' teaching practices reflect those recommended by current research in mathematics/science education?

4. What is the relationship between student teachers'/Faculty Fellows' perceptions of their own instruction and the observed classroom practice?

Our goal was to document and describe standards-based practices in college courses taught by OCEPT Faculty Fellows. In addition, we wanted to study classrooms of student teachers who had been enrolled in those courses. In both settings, we wanted to compare teacher instructional intentions (as described during the interviews) with observations of actual classroom teaching.

Purpose

The purpose of this study is to develop the instruments considered necessary for conducting the Outcomes Research Study. In preparing to engage in this study, we faced a classic problem of research in relatively undeveloped fields of study. There are few accepted methods and a dearth of good data from which to build. New approaches and new instruments are necessary to address the meaningful questions posed by scholars in the field. Jenks and Riesman (1968) expressed the problem in the preface to their analysis of higher education over three decades ago. "…responsible scholarship must invent methods and data appropriate to the important problems of the day. To reverse this process, choosing one's problems to fit the methods and data that happen to be most satisfactory, strikes us as an invitation to triviality…" (p. xii). Consequently this is the first of a series of reports designed to describe our efforts to study reformed teaching at the college level and its impact on new teachers. In so doing, we hope to avoid another long-standing and contrasting criticism of scholars and innovators in educational reform – that past work is ignored as though there is nothing on which to build (Jackson, 1983). Between these two critical positions we hope to develop innovative methods while maintaining a clear connection to past scholarship.

Existing Protocols:

Choosing an observation protocol for this study involved thinking about the context of teaching both in college courses as well as in K-12 classrooms. From the perspective of college instructors, educational reforms are intended to improve understanding and use of subject matter. From the perspective of K-12 teachers, the purpose is similar, but reform goals give a greater emphasis to improving student-teacher interactions. Further reflection on these two contexts suggests that they are more similar than they are different; this is especially true for college science and mathematics courses designed for non-majors such as elementary and middle level teachers. In these courses, reform advocates have stressed the need for significant improvement not only in translation of content into instruction but also about the necessity of positive and encouraging student-teacher interactions (NSF, 1996). For these reasons, protocol design proceeded under the assumption that the same observation tool would be used in classrooms from the elementary level through undergraduate college level.

The broad use of such an observation tool came with obvious caveats. We knew from the outset, for example, that we would not see the same constellation of behaviors in an undergraduate mathematics class as we would see in a mathematics lesson in an elementary school classroom. There was no *a priori* expectation that all K-12 teachers and college instructors would be meeting the same criteria. Further, we knew that when observing college lecture classes, the kinds of student-teacher interactions afforded by that setting would be significantly different from what is possible and desirable in a recitation section. There are numerous other differences that became a matter of reflection as we put the instruments to use. This will be discussed in more detail in the Implementation section of this paper.

Several scholars have attempted to design classroom observation protocols that assessed standards-based teaching practice. Methods of validation have tended to be *ad hoc* in nature. For example, Sawada and Piburn (March 2000) worked from personal expertise to design an observation protocol (RTOP) of 25 items in three categories supplemented by observational field notes. Reliability data was derived primarily from observer training and inter-rater reliability. They have achieved some correlation with RTOP ratings and student achievement. These interesting results provide no methods for isolating intervening variables. The problem is that there is no agreed-upon set of practices that represent the mathematics and science standards. Even the expected standards-based outcomes are open to wide interpretation. What does it mean, for instance, for a student to engage in problem solving in mathematics or inquiry in science? Other observational protocols have proceeded with significantly different assumptions about the nature of reformed teaching. See, for example, the Lederman and Schwartz (2001) procedure based on teaching about the nature of science and the Dana (2000) protocol based on instructional activity in laboratory settings.

The literature base also lacks clarity when it comes to determining what is going on in classrooms when standards-based instruction is taking place. There is often confusion in research reports between learning theory and instructional theory. For example, a researcher conducts a study and describes what students are doing and assesses what they are learning. From this, the researchers may inappropriately infer what teachers should do, when in fact no data were collected on the actions of the teacher (Roth, 1994; Carter & Jones, 1994). Data on how students learn and conditions for learning do not translate directly into teaching practices. Instructional design theory is concerned with what a teacher does and must include specific instructional method variables. Learning theory is concerned with mental representation, memory, reasoning, and other inferred mental processes. The distinction is important because instructional design theory directs teachers to emphasize particular variables that have been operationalized in research. Operationalizing learning theory research for the classroom, however, is much more subtle and challenging for the teacher (Reigeluth, 1999).

After examining the published instruments and protocols, we decided that none of the existing tools and methods met our needs. We determined that we needed to develop our own tools to carry out the Outcomes Research Study.

Development of New Protocols

We examined two decades of research on explicit teaching for initial guidance on the development of an observation protocol (Rosenshine & Furst, 1973). This work has produced a reliable set of observable instructional principles (Rosenshine & Stevens, 1986) relative to a defined perspective of teaching:

- Review previous and prerequisite learning.
- Clearly state learning goals.
- Present new material in small steps.
- Give clear and detailed instructions and explanations.
- Provide high levels of active practice for all students.
- Ask large numbers of questions and obtain responses from all students.

- Guide students during initial practice.
- Provide systematic feedback.
- Provide explicit instruction for independent practice and continually check for understanding.

Research on explicit teaching has provided a productive background for researchers and teachers interested in developing constructivist teaching approaches. More recent research has learned that high school and college age students have trouble using logical competence in scientific reasoning despite their presumed attainment of the Piagetian level of formal thought. Examining ninth graders through adults, Kuhn's (1992) results show broad problems in argumentation skills. These problems include confusing co-occurrence of events with cause and effect, preference for confirming rather than disconfirming evidence, and failure to consider potentially important factors by judging them irrelevant. A critique of this work by Koslowski and Maqueda (1993) suggested that Kuhn's evaluation may be overly restrictive. However, Koslowski and Maqueda emphasized that these capabilities require purposeful practice involving reflection on the relationships between theory and evidence and how they mutually constrain possible conclusions. In their review of these issues, Driver, Newton, and Osborne (2000) emphasize the significance of explicit teacher support in modeling and providing practice in thinking through various interpretations of evidence. The message is that relevant cognitive skills are not developed ready for use in classrooms or daily experience, but must be prompted, exercised, coached, and reinforced.

We also relied on the existing observation protocols in helping in our design. We appreciated the observational categories of Sawada and Piburn (March, 2000). Dana's (2000) laboratory observation protocol presented two useful dimensions: the student's role and the teacher's role. We reviewed studies of the Social Science Education Consortium (Fall, 1994) which utilized the 5E model (Bybee, 1997) and provided descriptions of teacher and student actions consistent with the model. The Lederman and Schwartz study (2001) described relevant characteristics of the nature of science and scientific inquiry appropriate for classroom teaching, identifying reform practices by specific statements delivered by the teachers in class. The Horizon Research Corporation (19999) observation protocol provided valuable descriptive categories. Finally, we examined the protocol designed by Lawrenz, Huffman, Appledoorn, and Sun (2001) for use in National Science Foundation Collaborative projects such as ours.

Building primarily on the work of Sawada and Piburn (2000), Lawrenz, *et al* (2001) and the Social Science Education Consortium (Fall, 1994), the authors designed the OCEPT Classroom Observation Protocol (O-TOP) (Appendix A). As we each reviewed and revised the instrument, it was circulated repeatedly among the three of us for feedback. Further review of the initial instrument suggested that observations of teaching should consider what is happening to include not just teacher actions but also student behaviors. As noted by Good and Brophy (1997) "…observers often try to reduce the complexity of classroom coding by focusing their attention exclusively on the teacher…but it is misplaced emphasis. The key to thorough classroom observation is student response. If students are actively engaged in worthwhile learning activities, it makes little difference whether the teacher is lecturing, using discovery techniques, or using small-group activities for independent study" (p. 51).

During the revision phase, the authors reviewed the instrument with respect to personal background and expertise in science education reform-based practices. In addition, the team reviewed the instrument for

a) limiting the observation categories to a number that an observer can remember and reflect upon during a class period

b) developing examples so that trained observers experienced in classroom teaching could reach agreement on meaning, and
c) setting a scale for each category that could be reliably applied.

The resultant instrument was examined by the entire research team, consisting of four science and/or mathematics education faculty and three graduate students. As a group we discussed the meaning of each item and the wording used as prompts. The team proposed revisions and additions to the instrument wording. When we felt there was sufficient agreement, we viewed a videotape of classroom teaching and individually rated the observed instruction on each of the ten items. Table 1 shows the percent agreement among the seven observers for rating each item with the same score as well as for rating each item within one point difference. For eight of the ten items, more than half of the research team agreed on the same score. For the same eight items, all seven observers were within a one point differential.

Two of the items initially caused a problem in interpretation. For Item #2 (Metacognition) and Item #5 (Student Preconceptions), there was a 57% and 71% agreement within one on the five point scale. The graduate students on the team had less experience with the topic of metacognition than the college faculty, and less experience in applying the research on misconceptions/preconceptions as well. Through discussion, the group reflected on personal classroom experience and related this to the meaning of reform standards. In the end, we were able to identify specific changes warranted in the instrument as a whole and for Items #2 and #5 in particular to ensure reliability in the use of the instrument. Further validation and reliability checks were carried out by pairs of researchers observing actual classrooms at the elementary, middle, high school, and college levels.

Table 1

Percent Agreement in Using the O-TOP

Item	Same Score	Within One
1	100%	100%
2	29%	57%
3	57%	100%
4	57%	100%
5	43%	71%
6	57%	100%
7	71%	100%
8	86%	100%
9	71%	100%
10	57%	100%

We felt the resultant instrument captured what needed to be observed and did so in a way that was manageable with a reasonable amount of training. In addition, the authors also designed an interview protocol (OCEPT Teacher Interview Protocol-O-TIP) based directly on the O-TOP (see Appendix B). The four open-ended questions prompt broad discussion within the ten categories of the classroom observation protocol. The process of reviewing and refining the O-TIP was considerably shorter given that the major categories had already been validated. Using the O-TIP along with the O-TOP acts to further validate the observational data and adds an in-depth description of the instructor's perspective.

The interview and observation protocols were further examined and evaluated by various expert groups. For example, the team presented the instruments at the OCEPT summer institutes and Oregon Academy of Science conference. Feedback from all groups was readily accepted and applied in strengthening the instruments.

Pilot Study

A Pilot Study to field test the instruments was implemented at three institutions (Oregon State University, University of Portland, and Pacific University). For this process, students were identified who were currently accepted into a teacher education program, working toward initial licensure, and had taken at least two courses from OCEPT Fellows. Twelve student teachers and six Faculty Fellows were involved in the pilot study. Most student teachers were observed teaching on three occasions; the Faculty Fellows were observed twice. Global scan field notes were taken during each observation, and the O-TOP instrument was completed following each class. As noted above, the initial observations were done by two members of the research team to check for inter-rater reliability in the use of the instrument. After the series of observations, the student teachers/Faculty Fellows were individually interviewed using the interview protocol. The interviews (typically 30 minutes in length) were audio-taped and later transcribed.

Data Analysis

The amount of data collected during the Pilot Study was daunting. We had 48 sets of observational field note, 48 completed O-TOP instruments, and 18 interview transcripts. We realized that when we applied these tools to our actual study, where we hoped to have a sample of 20 student teachers and 15 Faculty Fellows, the amount of data would be even larger.

To assist in analyzing this volume of data, the observers wrote a composite for each participant summarizing data from the field observations, the O-TOP instruments, and the O-TIP transcribed interview. The composites specifically included:

1. A table listing the student teacher's O-TOP rating for each item for each observation
2. A graph showing the sets of O-TOP ratings for comparisons
3. A description of the context
    - class type/methodology (e.g. lecture, lab, demonstration)
    - subject content/topic
    - place in sequence of unit (e.g. introduction, on-going, review) and/or relationship of observations (3 consecutive days, etc.)
    - description of students and make up of the class (e.g. sophomore and juniors in an elective class)
    - size of class
    - institution (public v. private, etc.)
    - important constraints (e.g. room set up, equipment limitations)
4. A description of the observed behaviors that led to the O-TOP scores for each observation
5. Patterns and interpretations of the total data set, relying on observations, O-TOP ratings and interview data
6. Additional pertinent comments/concerns not captured above.

The authors then analyzed all the composite case studies—referring to primary documents when necessary—to see if any generalizable patterns emerged. We are hopeful this method of analysis will be manageable as we continue with an expanded three-year longitudinal study.

Results

We were pleased with the actual application of the protocols. We were able to reliably gain the data we needed to answer the questions posed for the Outcomes Research Study. It should be noted, however, that the broad use of the observational tool came with obvious caveats. We knew from the outset, for example, that we would not see the same constellations of behaviors in an undergraduate mathematics class as we would see in an elementary mathematics class. There was no a priori expectation that all K-12 teachers and college instructors would be meeting the same criteria. Further, we knew that when observing college lecture classes, the kinds of student –teacher interactions afforded by that setting would be significantly different from what is possible ad desirable in a recitation section.

Additionally, unlike several other observation protocols (for example, see MacIsaac & Falconer, 2002) that rate the teaching experience and then total the numerical ratings, the O-TOP is meant to be a descriptive tool. We designed the O-TOP to generate a profile of what was happening across instructional settings rather than to assign a score to a particular lesson. In other words, we treat the ratings on the O-TOP items as categorical rather than interval data. This differs from the way the R-TOP has been used in recent reports (Piburn et al., 2000). We see the O-TOP results in combination with interviews and field notes from classroom visits as a prelude to theory building. Our understanding of how the items of the O-TOP performed in classroom observations had to be informed by the class context as well as the teacher's perspective.

Implications for Future Research

A great deal of interest in the observation instrument has developed from various sources suggesting applications of the O-TOP tool beyond its original intent in the Outcomes Research

Study. Several School of Education University Supervisors have reported using the instrument to provide feedback to their student teachers while observing in the field. Higher education faculty members have adopted the O-TOP as the protocol for implementing peer reviews within their departments. New teachers have indicated that the O-TOP provides a user-friendly checklist of good practices to consider during lesson planning, while experienced teachers have utilized the observation protocol as a component of their ongoing professional development. Some teachers have asked their principals to use the O-TOP during the annual evaluation process, especially principals who are unfamiliar with standards-based teaching in mathematics and science. Even college faculty and teachers outside of mathematics and science education have commented on the O-TOP's ability to describe effective teaching in their own content areas. For each of these applications, a preference has been expressed for the non-numerical version of the O-TOP, in which the 'scoring' is recorded on a continuum rather than on a "0 to 4" scale. (Appendix C)

The program of research stimulated by OCEPT that generated the instruments described here asks the broad question, "How does the whole of the college experience develop teacher knowledge and skill?" Specifically, we are interested in the higher education experiences that influence K-12 teaching in mathematics and science. It was a new concept for many faculty in mathematics and science departments to think of themselves as part of the teacher education process. Another broad implication from our work is the need to address the question, "How can we design tools that help higher education faculty evaluate their curriculum and instruction to better meet the needs of future teachers (as well as their non-education students)?" When considering the needs of elementary teachers, as compared to high school teachers, this implication has an even greater impact. Elementary teachers are an important subset of a much larger population of students taking content coursework who are non-majors. Therefore investigations that lead to an improvement in the academic experience of prospective elementary teachers will also improve the experience of the majority of all other students taking those mathematics and science content courses.

Discussions among science, mathematics, engineering and technology (SMET) faculty often focus on the expectation that teachers need additional subject matter courses, despite the fact that the courses available to non-majors are often taught in lecture-dominated formats where content is unconnected to familiar situations. Meetings with SMET faculty often confront the fact that about half of prospective elementary teachers take fewer than six semesters of science and almost half of those will not take any physics or chemistry at all. The mathematics faculty are only mildly appeased by the fact that virtually all students (96%) take a "mathematics for elementary teachers" sequence, but most will take no additional college mathematics courses. Education faculty are aware that only about half of future elementary teachers will meet NSTA's course background standards (National Science Teachers Association, 1998).

The O-TOP instrument is the kind of tool that can provide a common language for higher education faculties to use when discussing the structure and delivery of courses for teachers. Increasing faculty interest in new approaches to upgrading the content knowledge of future and practicing teachers holds the promise of promoting collaborative research efforts between SMET and Education faculties. The O-TOP tool is a starting point for research in designing data-based feedback to professors and graduate teaching assistants for the improvement of teaching. It provides a positive response to glaring shortcomings that have been identified in mathematics and science curriculum and instruction (National Science Foundation, 1996).

One outcome from the OCEPT project has been the development of a set of indicators to assist faculty in designing and evaluating their course revisions with respect to their value for prospective teachers. The Indicators for Selection of Mathematics and Science Content Courses Appropriate for Future Teachers (Appendix D) were evaluated by SMET and Education faculties of various institutions and organizations before they were employed as a self-evaluation tool for course modifications supported by OCEPT. These broad recommendations are consistent with recommendations for changes in science education at the collegiate level (Committee on Undergraduate Science Education, 1999).

The demands of teaching for higher order outcomes, such as promoting understanding of problem solving or scientific inquiry, is resulting in an increased awareness of teachers' interactions with students. The O-TOP instrument provides a starting point for K-12 teacher reflection on instructional practices. As higher education faculties become more aware of the impact of student-teacher interactions on student outcomes, they too have cause for reflection on their instructional practices. In a recent analysis of her own teaching, for example, Parsons (2001) outlined the implicit emphasis on reflection in teaching. She cites a large body of research dealing with a) defining reflection, b) developing curriculum to facilitate reflection, and c) examining the developmental process associate with reflection. She notes that the literature is rich in K-12 in-service and pre-service teaching, but sparse concerning reflection in college and university teaching. Not only can O-TOP provide a valuable tool for feedback that will support reflection for college and K-12 teachers, it can also be a starting point for a common dialogue on teaching that spans K-16 instruction.

Summary

Our research team has developed instruments for classroom observations and interviews which have a variety of applications at multiple levels of instruction. Through the use of these protocols, we hope to report on the relationship between beginning teachers' instructional strategies and the courses/instruction they experienced as an undergraduate. These instruments are appropriate for encouraging reflection and self-evaluation among K-12 teachers and college-level instructors alike.

Appendices:

A. The O-TOP (numerical range)

B. The O-TIP

C. The O-TOP (continuum, non-numerical)

D. Indicators for Selection of Mathematics and Science Content Courses Appropriate for Future Teachers

References:

American Association for the Advancement of Science (1989). *Science for all Americans.* New York: Oxford University Press.

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

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices.* Portsmouth, N.H.: Heinemann.

Carter, G. & Jones, M. G. (1994). Relationship between ability-paired interactions and the development of fifth graders' concepts of balance. *Journal of Research in Science Teaching, 31*, 847-856.

Committee on Undergraduate Science Education (1999). *Transforming Undergraduate Education in Science, Mathematics, Engineering and Technology.* Washington, DC: National Academy Press.

Dana, L. (Fall 2000). *The situated laboratory activity instrument (SLAI): A user's handbook.* Unpublished paper.

Driver, R., Newton, P., Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*. 287-312.

Good, T. L. & Brophy, J. E. (1997). *Looking in Classrooms, 5th edition.* New York: Longman.

Holton, B. E. & Horton, G. K. (1996). The Rutgers physics learning center. *The Physics Teacher, 34*, 138-143.

Horizon Research Corporation, Inc. (1999) *Local systemic change core evaluation data collection manual.* Chapel Hill, NC: Author.

Jackson, P. W. (1983). The reform of science education: A cautionary tale. *Daedalus, 112*, 143-166.

Jencks, C. & Reisman, C. (1968). *The academic revolution*. Garden City, NY: Doubleday and Co.

Koslowski, B. & Maqueda, M. (1993). What is confirmation bias and when do people actually have it? *Merrill-Palmer Quarterly, 39*, 104-30.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155.

Lawrenz, F., Huffman, D., Appeldoorn, K. & Sun, T. (2001). *Classroom observation handbook.* Minneapolis, MN: University of Minnesota.

Lederman, N. G. & Schwartz, R. (2001). Nature of science and scientific inquiry: Operational definitions and teaching approach as promoted in Project ICAN. Unpublished paper.

MacIsaac, D. & Falconer, K. (November, 2002). Reforming physics instruction via RTOP. *The Physics Teacher (40),* 479-485.

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

National Research Council (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

National Science Foundation (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology.* NSF 96-139. Washington, DC: The Author.

National Science Teachers Association (1998). *Standards for Science Teacher Preparation.* In collaboration with the Association for the Education of Teachers in Science. Available: http://www.nsta.org.

Parsons, E. C. (2001). Reflecting upon philosophy and practice correspondence: Guidelines for college and university educators. Paper presented at the meeting of the Association for the Education of Teachers in Science, Costa Mesa, CA.

Reigeluth, C. M. (Ed.) (1999). *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Roseman, J. E., Kesidou, S., & Stern, L. (1996). Identifying curriculum materials for science literacy. Online: www.project2061.org/research/roseman/roseman2.html.

Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching.* Chicago: Rand McNally.

Rosenshine, B. & Stevens, R. (1986). Teacher behavior and student achievement (pp. 376-391). In M. C. Wittrock, (Ed.) *Handbook of research on teaching, 3rd edition.* New York: MacMillan Publishing Company.

Roth, W-M. (1994). Experimenting in a constructivist high school physics laboratory. *Journal of Research in Science Teaching, 31*, 197-223.

Sawada, D. & Piburn, M. (March 2000). *Reformed teaching observation protocol (RTOP): Training Guide*. Phoenix, AZ: Arizona Board of Regents.

Social Science Education Consortium (Fall 1994). *Teaching about the history and nature of science and technology: Teacher's resource guide; field test edition.* Colorado Springs, CO: Author.

Sternberg, R. J. (1997). *Teaching introductory psychology: Survival tips from experts*. Washington, DC: American Psychological Association.

Tobias, S. (1992). *Revializing undergraduate education: Why some things work and most don't*. Tucson, AZ: Research Corporation.