2005–06 Local Systemic Change
Annotated Guide to the
Professional Development Observation Protocol

General Comments

The 2005–06 Local Systemic Change Professional Development Observation Protocol is designed to be completed by the person conducting LSC core evaluation professional development observations. The instrument is intended to reflect current standards for exemplary practice, but not to prescribe particular instructional strategies. Observers should refer to the following documents for background information about current standards for professional development and classroom instruction: NCTM's *Principles and Standards for School Mathematics* (2000) and NRC's *National Science Education Standards* (1996).

Observers should assess the professional development session in light of the particular purposes of the session being observed. It is important to remember that ratings are to be criterion-referenced, not norm-referenced. The professional development session should be rated based on its match to current standards, not rated highly because it is better than most other sessions you have seen in particular kinds of settings.

Completed Professional Development Observation Protocols must be submitted to HRI on the core evaluation website (**http://www.horizon-research.com/LSC**). (See the 2005–06 LSC Professional Development Observations, Guidelines for Evaluators for more details.)

The purpose of this annotated version of the Professional Development Observation Protocol is to clarify points about completing the form and to more explicitly define some of the terms used in the protocol. Each comment is keyed to the section or item number to which it refers. Please note that throughout the protocol the symbol "/" should be read "and/or."

BACKGROUND INFORMATION

Project: Enter the project name.

Location: Enter the city and state where the project is located.

Observer: If a team is observing the professional development session, select one name to enter on the form, typically the lead evaluator.

Approximate Duration of Observation: Note that the segment of professional development included on a single observation protocol should not exceed one-half day.

SECTION ONE: CONTEXTUAL BACKGROUND AND ACTIVITIES

I.B. 2.   "Teacher leaders" (previously referred to as "lead teachers") refers to teachers who serve in that role with the *LSC project*. They may be released from the classroom full-time, part-time, or not at all; and their responsibilities in the LSC may be very extensive or quite limited.

I.C.1.&2. Leave these items blank if there was no designated facilitator. For example, participants in a study group might share facilitation responsibilities, such as setting the agenda or making organizational/substantive decisions.

I.C.2.   Note that any teachers facilitating the observed session should be documented in one of the first two categories, depending on their current affiliation. Teacher leaders for the LSC project could fit into either of these categories depending on whether or not they are released full-time from classroom responsibilities.

III.A.   Note the reference to "primary"—referring to those purposes that best characterize the session.

Typically, the intended purpose(s) of the session will be determined based on the pre-interview with the project staff and/or the session facilitator. (See 2005–06 Professional Development Observation Guidelines for Evaluators for additional information about the Pre-Observation Interview.) Sometimes, purposes may shift in the course of the session and you would need to rate based on the revised purposes. For example, if the facilitator said the purpose was to increase teachers' content knowledge as well as increase familiarity with instructional materials, but you saw only the latter, it would be important to verify the purposes with the facilitator after the session.

III.A. 1. Note that "content knowledge" includes both concepts and process skills in science and mathematics.

III.A.2. *Explicit* attention requires a direct discussion/presentation about classroom pedagogical strategies and/or the designated instructional materials. If the plan is to simply model effective pedagogical strategies, but not to devote substantial attention to classroom applications, this purpose should not be selected.

In this, and other items, when choosing an umbrella category (e.g., explicit attention to classroom pedagogy) be sure to indicate any applicable sub-categories, as well. Any time a sub-category is selected, the umbrella category should also be marked.

III.A.3. As in III.A.2., *explicit* attention (e.g., through direct discussion/presentation) to strategies, issues or roles for teacher leaders, principals, or others in leadership positions is necessary to select this item as a purpose. If the plan is to model effective leadership strategies, but not to devote substantial attention to leaders' applications of those strategies, this purpose should not be selected.

III.A.4.e. If participants are learning about technology for their own use (e.g., using e-mail for communication with other teachers) it should be coded here. If participants are exploring technology for classroom instructional purposes, it should be documented as III.A.2.d., "Learning how to use technology in the classroom."

III.B.   Mathematics/science content can be addressed in a session either to increase participants' understanding of content or as a vehicle to accomplish other goals. In this item, document any content addressed in either manner.

While the left column primarily lists areas of mathematics and the right column primarily lists areas of science, you should feel free to mark items in both columns if they are major parts of a particular session. For example, data collection and analysis may be appropriate to describe the content of science sessions.

The check-box is provided for professional development sessions that focused on general pedagogy (e.g., cooperative learning) or leadership issues (e.g., peer mentoring) that did not relate directly to mathematics/science content.

III.B.1. "Numeration and number theory" also include rational numbers (fractions, decimals, percents).

III.B.16. Please write in the specific area in the category of life sciences.

III.B.17. "Physical science" includes concepts from both chemistry and physics. Please write in the specific area within these disciplines.

IV.A.   This question refers to materials designated in the project proposal or subsequently approved by the NSF Program Officer. Note that the observer needs to be aware of which instructional materials have been designated by the LSC for classroom use and whether any of those materials are being used in the observed session. (A question included on the Pre-Observation Interview addresses whether the session will focus on any of the designated materials.)

Note: This item is simply a documentation of what materials are used. Any comments regarding quality of materials themselves may be written in the space provided for comments (Section One: IV.D.) or may be reflected in the ratings in Section Two. Keep in mind that if the materials themselves have a particular weakness, that weakness is likely to be reflected in the assessment of the session.

The observer's ratings should reflect the *quality* of what was implemented. In some cases, the use of weak or flawed materials may result in a poor session and, therefore, poor ratings; on the other hand, if the facilitator adjusts the session to compensate for any flaws in the materials, the rating of the session should reflect that change.

IV.B.   Note the reference to "major" activities (i.e., the key activities that comprised the session). For example, a 15–20 minute discussion or presentation would probably warrant documentation, whereas a five-minute discussion or presentation would most likely not be recorded as a "major" activity. If you find you are filling in most items to describe the activities, you may need to be more discriminating in your interpretation of "major."

IV.B.3. The term "investigation" refers to activities in which participants are working toward finding the answers to meaningful questions. There is no restriction on who posed the question (i.e., it could have been either the participants or the facilitator).

IV.C.   The major approaches to professional development are defined as follows:*
Workshop/institute/course/seminar:
Using structured opportunities outside of the classroom to focus
intensely on topics of interest (including science or mathematics
content or the use of designated instructional materials) to learn from
others with more expertise.
Receiving formal professional development via technology:
Using telecommunications/videoconferencing/distance learning to
learn content and/or pedagogy.
Study groups/"kit clubs"/discussion groups/school-based meetings:
Engaging in regular, structured, and collaborative interactions
regarding topics identified by the group , with opportunities to
examine new information, reflect on practice, or assess and analyze
outcome data.
Coaching/mentoring:
Working one-on-one with an experienced teacher to improve teaching
and learning through a variety of activities, including classroom
observation and feedback, problem-solving and trouble-shooting, and
co-planning.

IV.D.   This item is intended for comments that are critical to understanding the
ratings that follow in Section Two. It is perfectly okay to leave it blank; there
is ample opportunity to document important features of the session in the
space provided for supporting evidence following each rating category in
Section Two.

_____

* Description of strategies is adapted from *Designing Professional Development for Teachers of Science
and Mathematics* by S. Loucks-Horsley, P.W. Hewson, N. Love, and K.E. Stiles. Thousand Oaks, CA:
Corwin Press, Inc., 1998.

SECTION TWO: RATINGS

This section relies heavily on the observer's synthesis and interpretation of what was observed in the professional development session. The key indicators are a guide to elements of the session that should be considered in each of the six rating categories.

Throughout this section, a key question to ask is: "Was this session likely to help move participants in the desired direction?" Keep in mind that building the capacity of teachers (as well as teacher leaders, administrators, and others) is an incremental process over time. Individual sessions cannot address every nuance of what is expected of an entire professional development program. For example, a session may focus heavily on having teachers experience inquiry as learners, with little, or no attention directed *in that session* to reflection about classroom implications. It is critical that observers have a clear understanding of the purposes of the session so that ratings may reflect the specified purposes.

Each key indicator in each of the six rating categories may be rated 1–5, 6 (don't know), or 7 (not applicable). Ratings of "not applicable" should be made when the attribute described in the key indicator is not present, but its absence did not limit the effectiveness of the session. For example, if the mathematics/science content did not include attention to real-world contexts, but the lack of those examples was inconsequential, indicator #8 in mathematics/science content would be rated "not applicable." If the attribute is one whose absence negatively impacts the effectiveness of the session, it should receive a low rating. For example, if opportunities for reflection were not designed into the session, and there clearly needed to be some time devoted to in-depth thinking then indicator #9 in design would receive a relatively low rating.

Note that even an excellent session would likely not rate highly on every applicable indicator; nor are the synthesis ratings intended to be an "average" of the individual indicator ratings. It is the observer's responsibility with knowledge of the purposes and context of the session and the experience of being there, to make a decision about how to weight these indicators to arrive at a synthesis rating for each category.

What Categories of Section Two Need To Be Rated?

While Section One of the Professional Development Observation Protocol needs to be completed in its entirety, there may be categories in Section Two that are not applicable to a particular session. Evaluators should use the following guidelines in choosing the appropriate categories to rate.

1. In all instances, Design (I), Implementation (II), Professional Development Culture (VI), and the Capsule Description (VII.C.) should be completed by the evaluator.

2. Categories corresponding to a stated purpose by the facilitator should always be completed unless the purposes are modified during the session. If you think the purposes may have changed – e.g., the facilitator said the lesson was aimed at increasing content knowledge and exploring pedagogy, but you saw no attention to pedagogy—check with the facilitator after the session to verify that the stated purposes remained in effect. If the facilitator reiterates that both were purposes, rate both Categories III and IV; if s/he indicates a revision to address content only, skip Category IV (Exploring Pedagogy/Instructional Materials).

3. In some sessions, categories **in addition** to those explicitly stated as purposes will need to be completed. There are two instances when this may occur: (1) rating mathematics/science content when it was used primarily as a vehicle for accomplishing other purposes in the session, and (2) rating either Mathematics/ Science Content or Exploring Pedagogy/Instructional Materials when the lack of coverage in either area acted as a barrier to other stated purposes.

   - Example #1: The stated purpose of the session was to increase teachers' familiarity with instructional materials. Throughout the session, teachers experienced the science kits on electricity just as the students would in their classrooms. In this example, the category Exploring Pedagogy/Instructional Materials should be rated because it was a stated purpose. In addition, Mathematics/Science Content should be rated because electricity was a vehicle for the purpose of exploring the instructional materials and whether or not that content was sound, developmentally appropriate, etc. makes a difference in the effectiveness of the session.
   - Example #2: The facilitator indicated that the purpose of the session was to explore leadership issues with teacher leaders. The session focused on strategies for coaching and mentoring peers in using inquiry in their classrooms. During the session, discussions among teacher leaders revealed confusion about what did and did not "count" as inquiry, but this confusion was never addressed in the session. The evaluator should rate Exploring Pedagogy/Instructional Materials (as well as Leadership Content) because inadequate coverage of pedagogy hindered the achievement of the stated purposes of leadership development.

4. For most sessions involving teachers and/or teacher leaders (except those with the sole purpose of exploring strategies/issues/roles of teacher leaders, principals, or others in leadership positions), evaluators should rate the likely impact of the session on participants' capacity to provide high quality mathematics/science education (Category VII.A). Ratings of the likely impact on participants' leadership capacity (Category VII.B.) should be completed if teacher leaders or others in leadership positions (e.g., administrators) were included as participants in the session, whether or not leadership issues were dealt with explicitly, and whether or not there were other participants as well as teacher leaders.

I. Design

The "design" of the session refers to the structure of the observed session: what happened first, second, third, etc., and how much time was allowed for each of these activities. This category generally encompasses the activities, the strategies, the assigned roles, and the resources of the session. A given design can be replicated by multiple facilitators.

The key indicators provide detailed information on what aspects of the session to consider in rating the design. As noted in the overview to Section Two, individual sessions may not include every component of what would be considered sound professional development. Ratings for these items should reflect the specified purposes of the session.

Some notes on individual key indicators:

1. Including "tasks, roles, and interactions consistent with a spirit of investigation" does not require investigative activities, *per se*; rather this rating should reflect the extent to which the types of tasks and roles and the nature of the interactions in the session encouraged the collaboration, intellectual rigor and curiosity, and independent thinking that characterizes the investigative process.

2. & 3. The ability to rate these indicators may be limited by the observers' knowledge of the targeted participants. Data from a variety of sources could contribute to rating the extent to which the session reflected attention to participants' experience, preparedness, knowledge, etc.: comments made by the facilitator in the session or in the pre-interview; apparent fit between the planned activities and the participants' capacities; types or levels of questions directed to particular participants, etc. Without sufficient information, observers should not hesitate to rate this indicator "don't know."

7. The term "collaborative" refers to a design in which the participants share/use each other's knowledge, contributions and experiences. It does not necessarily imply that group work was a part of the session.

9. In this indicator, "sense-making" is broadly defined to include time for thought and processing. It may occur in a variety of contexts: individually, in groups, or as a whole group activity, and at any time in the course of a lesson: before, during, or after activity, as part of the wrap-up, etc. Document the extent to which the session included well-planned opportunities for "sense-making" here; ratings of quality and effectiveness may also occur elsewhere (e.g., in category III., Mathematics/Science Content, Item 9.)

10. In this indicator, document the extent to which the session included well-planned opportunities for participants to share experiences and insights. The quality of this component may impact ratings in other categories, as well (e.g., the mathematics/science content; the exploration of pedagogy and instructional materials; and the culture of the professional development session).

II.   Implementation

This category refers to how the facilitator(s) carried out the basic structure of the session and is unique to each facilitator. The critical questions are "How effectively do the facilitator(s) implement the design?" and "Is the implementation of the session likely to move the participants forward in their capacity as teachers and/or leaders?"

Note that some activities are more "sensitive" to implementation issues than others. For example, the effectiveness of a small-group task is tightly linked to the soundness of its design; its quality would likely be assessed in the ratings for design. On the other hand, the quality of some activities may be highly dependent on how the facilitator carries them out (such as formal presentations or informal comments/contributions). In those cases, ratings of quality would be reflected in both the soundness of the design and its implementation.

Some notes on individual key indicators:

3. In rating questioning strategies, attend to whether the facilitator(s)' questions were open-ended, as opposed to focused solely on one answer, and the extent to which the facilitator probed for participants' reasoning and misconceptions. Key questions to guide the rating are "Does the questioning help the facilitator understand participant conceptions?" and "Does the questioning help develop the conceptual understanding of participants and move their thinking forward?" Note that the "conceptual understanding" could be in relation to mathematics/science content, pedagogical content, and/or leadership content.

7. Keep in mind that indicators, such as this one, that may be absent from the session should be rated "not applicable" unless their absence hindered the effectiveness with which the session was likely to move participants' thinking forward.

III. Mathematics/Science Content

This category should be rated if "increasing mathematics/science content knowledge of participants" was selected as a purpose of the session in Section One, Part III.A., if mathematics/science content was a vehicle for accomplishing other professional development purposes, or if inadequate coverage in this area acted as a barrier to accomplishing other stated purposes. If not, check the box that indicated none of these apply and skip to Category IV, Exploring Pedagogy/Instructional Materials.

This content category includes both concepts and process skills in science and mathematics. Keep in mind that there is no one "right" formula—the balance of concepts and processes needs to be appropriate for the purposes of the session.

Some notes on individual key indicators:

3. In rating this indicator, observers should be attuned to the qualitative connotation of "intellectually engaged" that goes beyond simply doing the task. In other words, evaluators should assess the extent to which a "minds-on" intellectual rigor characterized the session. A rating of "5" should be made only if most participants were deeply engaged with the mathematics/science content of the session.

7. If elements of mathematical/scientific abstraction were included when it was appropriate to do so, this item should be rated highly. If abstractions were included, but inappropriate for the participants or the session, the item should be rated 1 or 2. If mathematical/scientific abstractions were lacking and their absence was detrimental to the session, the item should be rated 1 or 2. If mathematical/scientific abstractions were absent and their absence was not detrimental, the item should be rated "N/A."

8. Connections to other areas of mathematics/science, to other disciplines, and/or to real-world contexts are an element of instruction that may facilitate participant learning, but may not be necessary or even appropriate for particular sessions. Use the general rating guidelines: "not applicable" if it was not a part of the session and its absence was not a detriment to participant learning; a low rating if the absence of connections was a likely hindrance to participant learning.

9. Rate the appropriateness of "sense-making" considering where the participants were in their exploration of the content and where this session fit in the overall sequence of professional development. Information provided by the facilitator in the pre-interview may be useful here.

IV. Exploring Pedagogy/Instructional Materials

This category should be rated if "explicit attention to classroom pedagogy/ designated instructional materials" was selected as a purpose of the session in Section One, Part III.A., or if lack of adequate coverage in this area acted as a barrier to accomplishing other stated purposes. If not, check the box that indicates neither of these apply and skip to Category V, Leadership Content.

Some notes on individual key indicators:

1., 2., & 3. Use the general rating guidelines: "not applicable" if each is not a part of the session and its absence was not detrimental to participant learning; assign a low rating if the absence was an obstacle to participant learning.

5. In rating this indicator, observers should be attuned to the qualitative connotation of "intellectually engaged" that goes beyond simply doing the task; i.e., evaluators should assess the extent to which a "minds-on" intellectual rigor characterized the session. A rating of 5 should be made only if most participants were deeply engaged with ideas relevant to classroom practice.

6. Rate the appropriateness of "sense-making" considering where the participants were in their exploration of the particular classroom practice(s) and where this session fit in the overall sequence of professional development. Information provided by the facilitator in the pre-interview may be useful here.

V.   Leadership Content

This category should be rated only if "explicit attention to strategies/issues/roles of teacher leaders, principals, or others in leadership positions" was selected as a purpose of the session in Section One, Part III.A. If not, check the box that indicates this was not a purpose of the session and skip to Category VI, Culture of the Professional Development Session.

Indicators 1, 2, and 3 may well be rated "not applicable" if the area specified in the indicator was not included as a focus of the session.

Some notes on individual key indicators:

5.   In rating this indicator, observers should be attuned to the qualitative connotation of "intellectually engaged" that goes beyond simply doing the task; i.e., evaluators should assess the extent to which a "minds-on" intellectual rigor characterized the session. A rating of "5" should be made only if most participants were deeply engaged with the leadership content of the session.

VI. Culture of the Professional Development Session

This category includes indicators directed toward the extent and nature of the engagement of participants in the session. Observers also have the opportunity to comment on issues of equity and diversity that may have impacted the quality of the session by responding to item A.2. with specific examples that reflect particular sensitivity or insensitivity toward diversity.

Note that the anchors for the culture synthesis rating are worded differently from anchors in other component areas, which referred to "best practice" or "current standards." This synthesis rating should reflect the extent to which the culture of the professional development session *interfered with* or *facilitated* engagement of participants as members of a professional learning community. A rating of "3" has the connotation of a "neutral" rating.

Observers should also keep in mind as they assess the impact of culture on the overall ratings of the session, that a positive culture may contribute to the accomplishment of session goals, but it cannot compensate for deficits in other aspects of the session. Regardless of how positive the culture, the session must still be assessed on the extent to which it was likely to help participants gain important knowledge and skills. In other words, a positive culture is a desired, but not sufficient ingredient for an effective professional development session.

Some notes on individual key indicators:

7. Ratings of this indicator should reflect the character of participants' intellectual engagement with each other and the facilitator; i.e., the extent to which ideas were challenged and challenging; constructive criticism was offered; and rigor, openness, and honesty characterized the exchange of ideas and interactions.

VII. Overall Ratings of the Session

Parts A. and B. of Category VII are intended to serve as "on-ramps" to the final capsule description of the session. The indicators are designed to prompt the observer to synthesize the previous ratings along with related interviews and knowledge of the professional development program to assess the likely impact of the session on participants' capacity to provide high-quality mathematics/science education and/or participants' leadership capacity. The important question is "Was this session likely to help move participants in the desired direction?" (Remember that a variety of types of sessions may build capacity and "move thinking forward"; no one professional development strategy is ideal in every setting.)

The key indicators in this section are rated on a different scale, one that focuses on the likely *effect* (negative, neutral, or positive) of the session. The proportion of participants impacted as well as the various components of the session itself should be reflected in these ratings.

If the session was likely to have a markedly positive effect on most of the participants, fill in the bubble under "positive effect." A session likely to have a highly positive effect for some participants, but neutral for others might be averaged to a rating between the two columns labeled "positive effect" and "neutral." Sessions in which there is likely to be either no discernible effect (neutral) or an unequal, but relatively acceptable level of impact on participants might be rated as having a "mixed or neutral effect."

A. Likely Impact on Participants' Capacity to Provide High-Quality Mathematics/Science Education

Complete this part for any session that focused on building the capacity of teachers or teacher leaders for classroom instruction. This would include most sessions involving teachers; the only exception would be sessions whose sole purpose was exploring strategies/issues/roles of teacher leaders, principals, or others in leadership positions.

Some notes on individual key indicators:

1.–5. Note that items 1., 2., & 3. target specific types of mathematics/science and pedagogical content knowledge; where as items 4. & 5. are much broader in scope. Ratings of items 4. & 5. may well be impacted by the quality with which a variety of other aspects of professional development were addressed.

7. "Professional networking" should be interpreted well beyond simply providing the opportunity for participants to interact and get to know each other within a particular professional development session. This item refers to establishing/maintaining ongoing teacher networks for enhancing classroom instruction.

B. Likely Impact on Participants' Leadership Capacity

Complete this part for any session that included, as participants, teacher leaders, principals, or others who are in leadership positions, whether or not leadership issues were dealt with explicitly, and whether or not the session included participants in addition to teacher leaders/administrators.

C. Capsule Description of the Quality of the Professional Development Session

This final rating encapsulates the observer's overall assessment of the quality and likely impact of the session. As with previous synthesis-type ratings, it is not intended to be an average of the various categories and "on-ramp" ratings, but should reflect the observer's assessment of the relative importance of the available information, taking into consideration the purposes of the session.

If Level 1 (Ineffective Professional Development) is selected, the observer should also designate one of two sub-categories: Passive "Learning" or Activity for Activity's Sake. Note that Passive "Learning" is an oxymoron of sorts. Little, if any, learning is expected to take place in sessions characterized as either passive "learning" or activity for activity's sake.

If Level 3 (Beginning Stages of Effective Professional Development) is selected, the observer should also designate whether it is a low, solid, or high 3. These sub-categories were added to "fine-tune" the Level 3 category since it includes a broad range of professional development sessions.

Please note that observers should provide a rationale for their final capsule rating of the session.