

2005–06 Local Systemic Change

Annotated Guide to the Classroom Observation Protocol

General Comments

The 2005–06 Local Systemic Change Classroom Observation Protocol is designed to be completed by the certified observer conducting LSC core evaluation classroom 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 classroom instruction: NCTM’s *Principles and Standards for School Mathematics* (2000) and NRC’s *National Science Education Standards* (1996).

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

Completed classroom observation protocols must be submitted to HRI on the core evaluation website (<http://www.horizon-research.com/LSC>). (See the 2005–06 LSC Classroom Observations, Guidelines for Evaluators for more details.)

The purpose of this annotated version of the Classroom 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.

LSC ID: Enter the ID number of the teacher, indicated in the Classroom Observation Sample provided by HRI.

Subject Observed: Choose “science” or “mathematics,” depending on the discipline area targeted by the LSC. In projects that target both science and mathematics, select the subject for which the teacher was sampled.

Observer’s Role in Project: Select the description that *best* describes the observer’s role. Note: Only certified individuals are approved to conduct classroom observations for the LSC Core Evaluation. Classroom observations conducted by non-certified observers cannot be accepted in the core evaluation database.

SECTION ONE: CONTEXTUAL BACKGROUND AND ACTIVITIES

- I.B. Note that in this item you are to record the percentage of *white non-Hispanic* students. (This is different from forms, which often ask for percentage of “minority students.” The term “minority” was ambiguous since in some districts “minority students” actually comprise a majority of the student population.)
- I.C and D. If the classroom you observe is team-taught, describe the teacher selected for the random sample in item I.C. and describe the second teacher in item I.D. Comments about the teaching arrangement could be noted in item V.D.

II. Note that you must use information provided by the teacher to describe where this lesson fits in the overall unit of study.

III.A. “Major” content refers to content that was a substantial focus of the lesson; you should not document *every* nuance of the lesson. If you were describing the lesson to someone, the topics you designate are ones you would be sure to mention.

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

In the rare event that the lesson focused on mathematics/science process—e.g., problem-solving strategies, the scientific method, etc.—but did *not* target or use mathematics/science concepts, indicate “None of the above” and explain in the space provided.

III.A.1. “Numeration and number theory” also includes rational numbers (fractions, decimals, percents).

III.A.16. Please indicate the specific area in the category of life science.

III.A.17. “Physical science” includes concepts from both chemistry and physics. Please indicate the specific area within these disciplines.

III.B. Note the reference to “primary”—referring to those purposes that best characterize the lesson.

Typically the intended purpose(s) of the lesson will be determined based on the pre-interview with the teacher. Sometimes, purposes may shift in the course of the lesson. If you think that may have happened, you can ask the teacher about it in the post-observation interview and use that information in completing the purposes section. The point here is to document what the teacher was intending to accomplish in this lesson, *not* what you think was actually accomplished.

IV.A.–D. It is important for the observer to be thoroughly conversant with the instructional materials designated for use in the project. You may need to review the specific lesson in the unit prior to completing these items.

IV.C. “Teacher manual” is used to refer to any instructions to the teacher accompanying the instructional materials.

V.A.–C. In all of these sections, “major” has a similar interpretation as described in III.A. above. If you find you are filling in *large numbers* of items to describe the class resources and activities you may need to be more discriminating in your interpretation of “major.”

V.A. & B. Students working in groups on the *same* activities will be documented by the combined responses to items V.A. and V.B.; i.e., “small groups” marked in item V.A. and the first option marked in item V.B. (entire class was engaged in same activities at the same time).

V.C. Please note that these are broad *descriptive* categories to document the type of activities that occurred. Observers will provide an assessment of the *quality* of the class activities in Section Two of the protocol.

When choosing an umbrella category (e.g., listened to a presentation), be sure to indicate any applicable sub-categories as well. Any time a sub-category is selected the umbrella category should be marked, as well.

V.C.3.d. The sub-category “had some latitude in designing an investigation” should be used when students have the opportunity to make choices about the design and/or implementation of an investigation. Note that this category is appropriate for activities in which students design objects within constraints, i.e., toothpick bridges, aluminum foil boats, egg drop containers.

V.C.4. “Reflection” should be marked when the structure of the lesson provides a substantial amount of time for students to think individually or within groups about a topic, activity, problem, etc.

V.C.5. Sub-categories under the umbrella category “used technology/audio-visual resources” allow the observer to document *how* computers, videotapes, and other such resources were used in the lesson.

REMINDER: Throughout the entire protocol, the symbol “/” should be read “and/or.”

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

SECTION TWO: RATINGS

This section relies heavily on the observer's synthesis and interpretation of what was observed in the classroom. The key indicators are a guide to elements of the instruction that should be considered in each of the four rating categories. Throughout this section, a key question to ask is "Was this lesson likely to help move students in the desired direction?" Keep in mind that "moving thinking forward" does not mean that students must have a *full* understanding of the targeted concept or that the lesson necessarily comes to complete resolution. Students' understanding may be moved forward in different ways through different types of lessons. For example, lessons might be designed for students to gain experiential awareness; to engage in open exploration; to practice skills; to review, reinforce, and apply emerging understanding; or to develop curiosity and questioning about a particular concept or phenomenon.

Each key indicator in each of the four 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 lesson. If the attribute is one whose absence negatively impacts the effectiveness of the lesson, it should receive a low rating. For example, if opportunities for "sense-making" were not designed into the lesson, and there clearly needed to be some time devoted to in-depth thinking (e.g., why different groups came up with different solutions to a particular problem; the implication of findings for subsequent work) then indicator 7 in design would receive a relatively low rating.

Note that even an excellent lesson 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 lesson 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.

Some concern has been expressed about the potential bias toward "active and collaborative learning" conveyed by the key indicators of this form. While current national standards emphasize an active role for students, based on the belief (and some research evidence) that such strategies lead to greater student understanding, NSF's overriding goal is to have students learn important science and mathematics concepts. You may observe a fairly didactic, teacher-centered lesson that is geared to the needs and interests of the students, clearly engages them, and is likely to move their conceptual understanding forward; such a lesson could be rated a 4 or even a 5.

A number of indicators throughout Section Two require ratings that consider the purposes of the lesson. Please refer back to the notes from III.B. in Section One of this annotated guide regarding documentation of the purposes of the lesson. If the purpose of the lesson is modified in the course of the lesson, indicator ratings should be based on that modified purpose.

You may encounter situations where, for example, the strategies and materials are appropriate for the intended purposes, but in your judgment, the purposes of the lesson are not reflective of current standards. In such circumstances, indicator ratings should be made based on the *teacher's* intended purposes. Synthesis ratings and your supporting evidence would reflect the fact that the purposes were not consistent with current standards.

I. Design

The “design” of the lesson is not a reference to the “lesson plan,” but rather the *structure* of the observed lesson: what happened first, second, third, etc.; and how much time was allowed for each of these activities. This category generally encompasses the activities, the instructional strategies, the assigned roles, and the resources of the lesson. The key indicators provide detailed information on what aspects of the lesson to consider in rating the design.

Some notes on individual key indicators:

1. The term “investigative” refers to instruction in which students 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 students or the teacher.
2. See the above clarification of the term “design.”
3. While classes may exhibit variations in extent of attention toward individual students’ experience, preparedness and/or learning style, this rating should reflect the predominant manner in which the strategies and activities addressed student needs.
6. The term “collaborative” refers to a design in which the students and teacher share/use each other’s knowledge, contributions and experiences.
7. In this indicator, “sense-making” is broadly defined to include time for thought/processing. It may occur in a variety of contexts: individually, in groups, or as a whole group activity, and at any time in the lesson: before, during, or after an activity, as part of the wrap-up, etc. Document, here, the extent to which the lesson included well-planned opportunities for “sense-making”; rating of quality and effectiveness may also occur elsewhere (e.g., in category III. Mathematics/Science Content, item 9).
9. “Formal assessments” might include tests or performance tasks. The point is to look at the message “tests” are sending to students about what learning is important. Tests that assess deeper understanding would be rated highly on this indicator; tests that assess superficial knowledge of terminology and computation would often receive lower ratings.
10. While you may be able to rate this indicator based only on the classroom observation, in most cases a rating other than 6 or 7 would need to be based on information from the post-observation interview.

II. Implementation

This category refers to how the teacher carries out the basic structure of the lesson. The critical questions are “How effectively does the teacher implement the design?” and “Is the implementation of the lesson likely to move the students in that particular class forward in their understanding?”

Some notes on individual key indicators:

1. Note that in Section One, Part IV, observers are asked to document whether LSC-designated materials are used. If so, observers indicate the extent to which the teacher adhered to the instructional materials, and how adaptations affected the quality of the lesson design. In rating the indicator in this section, the observer should assess the extent to which the implementation reflected the spirit of the LSC-designated materials *whether or not* those particular materials were used in the lesson.
2. Instructional strategies that are most consistent with “investigative science/mathematics” portray the disciplines as dynamic bodies of knowledge generated and enriched by investigation. Strategies that are rated highly might build in and actively demonstrate those characteristics; those rated lower might negate or misrepresent the investigative nature of the disciplines.
3. Indicators of confidence may well depend on indirect or subjective cues. This indicator should be rated based on the activities/behaviors/interactions of the teacher *in the class* rather than information provided in pre- or post-interviews.
6. Data from a variety of sources could contribute to rating the extent to which the teacher was able to “read” the students’ level of understanding and adjust accordingly i.e., types or levels of questions directed to particular students, etc. Without sufficient information, observers should not hesitate to rate this indicator “don’t know.”
7. In rating questioning strategies, attend to whether the teacher’s questions are open-ended, as opposed to focused solely on one answer, and the extent to which the teacher probes for student reasoning and misconceptions. Key questions to guide the rating are “Does the questioning help the teacher understand student conceptions?” and “Is the questioning likely to help develop the conceptual understanding of students and move their thinking forward?”
8. Modifications of the lesson may or may not be apparent. Use both the observation and pre-post-interviews to make this rating, but keep in mind that some of the best teachers may not even be consciously aware of how they are modifying lessons. Follow the general guidelines established for ratings: “N/A” if the lesson was not modified and didn’t need to be; “don’t know” if it is not clear whether it was modified or not; or ratings from 1–5 as appropriate. For example, if it was clear (e.g., from students’ questions, comments, performance) that students were not getting it and the teacher proceeded as planned with no change and had no appropriate rationale for sticking to the plan, you might rate it a 1 or 2; if the teacher clearly modified the lesson in process, you might rate it between 3 and 5 depending on the effectiveness of the modification.

III. Mathematics/Science Content

The 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 lesson.

Some notes on individual key indicators:

1. When rating this indicator, consider the importance of the mathematics/science content independent of other factors, such as appropriateness for the particular students in the observed class.
2. Focus in this indicator on the developmental appropriateness of the mathematics/science content for *this* class regardless of its inherent importance in K-12 science/mathematics.
3. In rating this indicator, observers should be attuned to the qualitative connotation of “intellectually engaged” that goes beyond simply doing the task. Attention should also be directed to the proportion of students who appeared to be intellectually engaged. A rating of 5 should be made only if most students were deeply engaged.
4. Rate this indicator “N/A” if there was no teacher-provided content information.
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 students or lesson, the item should be rated 1 or 2. If mathematical/scientific abstractions were lacking and their absence limited the effectiveness of the lesson, the item should be rated 1 or 2. If mathematical/scientific abstractions were absent and their absence was inconsequential, the item should be rated “N/A.”
8. Connections to other areas of mathematics/science, to other disciplines, and/or to real-world contexts is an element of instruction that may facilitate student learning, but may not be necessary or even appropriate for particular lessons. Use the general rating guidelines: “N/A” if it was not a part of the lesson and its absence was not a detriment to student learning; a low rating if the absence of connections was a likely hindrance to student learning.
9. Rate the appropriateness of “sense-making” of mathematics/science content in this lesson, considering where the class was within the sequence of lessons in a particular unit. Information provided by the teacher in the post-interview may be useful here.

IV. Classroom Culture

This category combines indicators directed toward the extent and nature of the engagement of students in the class, and provides an opportunity for an observer to comment on issues of equity and diversity that may have impacted the culture of the classroom.

Key indicators 1–4 should always be rated on the 1–5 scale; for indicators 5 and 6, the full rating scale (1–7) should be considered. Note that the synthesis rating should reflect the extent to which classroom culture *interfered* with or *facilitated* student learning.

Some notes on individual key indicators:

1. When rating this indicator, attend to the *extent* to which active participation was encouraged/valued. The nature and quality of the participation is addressed in other indicators, such as 5 and 6.
6. This indicator often carries a lot of “weight” in the synthesis rating since participation without intellectual rigor is not likely to contribute meaningfully to student learning.

V. Overall Ratings of the Lesson

A. Likely Impact of Instruction on Students' Understanding of Mathematics/Science

Part A of Category V is intended to serve as an “on-ramp” to the final capsule description of the lesson. The indicators are designed to prompt the observer to synthesize the previous ratings to assess the likely impact of the instruction on students' understanding of mathematics/science. The important question is “Was this lesson likely to help move students in the desired direction?” (Remember that a variety of types of lessons may “move thinking forward”; refer back to the beginning of Section Two for more detailed information.)

The key indicators in this section are rated on a scale that focuses on the likely *effect* (negative, neutral, or positive) of the lesson. If the lesson was likely to have a markedly positive effect on most of the students in the class, fill in the bubble under positive effect. A lesson likely to have a highly positive effect for some students but neutral for others might be averaged to a rating in between these two columns.

A mixed or neutral rating could indicate the lesson was likely to have either no discernible effect (neutral) or an unequal, but relatively acceptable level of impact on the students.

B. Capsule Description of the Quality of the Lesson

This final rating encapsulates the observer's overall assessment of the quality and likely impact of the lesson. 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 lesson.

If Level 1 (Ineffective Instruction) 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 lessons characterized as either passive "learning" or activity for activity's sake.

If Level 3 (Beginning Stages of Effective Instruction) 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 instruction.

Please note that evaluators should provide a rationale for their final capsule description of the lesson.