



# Inquiring into Science Instruction Observation Protocol (ISIOP)

## USER'S MANUAL

Daphne Minner, PhD  
Jacqueline DeLisi, EdD  
2012

The ISIOP was created by Daphne Minner, PhD, and Jacqueline DeLisi, EdD, at Education Development Center, Inc. (EDC). The instrument development was supported through grants from the National Science Foundation (DRL #0535787 and DRL #0815364). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**For questions or comments regarding this EDC product, please visit <http://isiop.edc.org> or email Jackie DeLisi at [jdalisi@edc.org](mailto:jdalisi@edc.org). For communication with Daphne Minner, email [daphne\\_minner@abtassoc.com](mailto:daphne_minner@abtassoc.com).**

©2012 Education Development Center, Inc. All rights reserved.

### **Acknowledgments**

The ISIOP products—including this User’s Manual, the Codebook, the Data Collection Instrument, and all the training materials—represent the thinking of many people over many years. The authors wish to thank the EDC staff who were involved in this initiative at various points in time—Linda Hirsch, Ruth Krumhansl, Kathy Paget, Lisa Marco-Bujosa, Tzur Karelitz, and Neil Schiavo. Several consultants contributed to this protocol, including Jonathan Templin, Kenne Dibner, and Patricia Bills. There were also several advisors and approximately 50 researchers throughout the country who contributed their thoughts and observational data during a year-long field test of this instrument. While there are too many to name here, the insights they all provided were valuable for shaping the final instrument. This work could not have been completed without two master teachers who collaborated on the initial development of the tool—Michael McCarthy and Allison Friedman—and the many other teachers who offered our team and the many field-testing teams access to their classrooms. The product contained in these pages represents the variety and the effort across this group of classroom practitioners. And finally, the authors are extremely grateful for the support of Kerry Ouellet, who provided valuable insight into the look, style, and formatting of this product; and Denise Ethier and Marshall Goff, who assisted with the online media.

## Contents

<b>Introduction .....</b>	<b>1</b>
<b>ISIOP Conceptual Framework Overview.....</b>	<b>2</b>
<b>Use Considerations .....</b>	<b>3</b>
<b>ISIOP Training Procedure.....</b>	<b>5</b>
Overview of the Training Steps .....	5
Step 1: Locate and Familiarize Yourself with the ISIOP Products .....	5
Step 2: Coding Lesson Events .....	6
Step 3: Coding Verbal Practices and Investigation Experiences.....	7
Step 4: Coding Classroom Instructional Leadership Practices.....	9
Step 5: Coding the Entire Protocol .....	10
Step 6: Checking Inter-rater Reliability .....	10
Step 7: Additional Practice (if necessary).....	11
<b>Preparation for Live Classroom Observation .....</b>	<b>12</b>
<b>Scoring Guidelines .....</b>	<b>13</b>
Establishing Inter-rater Reliability .....	13
Analysis Procedures and Score Interpretation.....	17



## Introduction

The purpose of this classroom observation protocol (*Inquiring into Science Instruction Observation Protocol—ISIOP*) is to assist evaluators and researchers in determining the extent to which quality pedagogical practices, including instruction that integrates scientific practices and habits of mind, are present in secondary science classrooms. The ISIOP is a classroom observation protocol designed to quantitatively and comprehensively describe the extent of various aspects of science teaching, including the nature of a lesson’s discrete events, the varied teacher moves, and the Investigation Experiences. The ISIOP reflects a comprehensive view of classroom practice that is standards-based and inquiry-oriented, and includes teaching indicators that have been either theorized or demonstrated to be associated with student learning. The focus of the ISIOP is on the teacher, specifically those instructional practices that are exhibited during a given lesson. The observed practices range in grain size, from individual teacher utterances, through the types of activities, and up to variables, such as the teacher’s instructional style, which can be observed across an entire lesson. The indicators of teacher practice are organized to provide, for a given lesson, a quantitative metric of:

- the nature of the Verbal Practices (spoken instructional strategies) exhibited by the teacher,
- the overall organization and nature of the lesson’s discrete activities or events,
- the science content taught,
- the teacher’s Classroom Instructional Leadership (i.e., classroom management, instructional style), and
- the kinds of Investigation Experiences in which students were engaged.

These indicators are operationalized in the ISIOP using quantifiable data collection techniques. The ISIOP is designed to capture “teacher moves” (instructional decisions and verbal responses) that provide opportunities for students to learn independent of any particular curriculum guiding the instruction. It captures teacher-enacted curricula, not the intended curricula, and it focuses on the teacher’s practices in a given lesson rather than on student responses. We recognize that monitoring curriculum implementation and student engagement are possible ways in which the ISIOP may be used in evaluations, so for those applications to be possible, we have included a few questions in the Pre-Observation Questionnaire to give the context of the observed lesson within the larger instructional sequence, and have embedded a metric of student engagement within each lesson activity.

The ISIOP contains several materials, some of which are needed for training purposes only; others are required for observations. All of these products are located on the ISIOP website: <http://isiop.edc.org>:

- ISIOP User’s Manual (this document)
- ISIOP Codebook
- ISIOP Teacher Pre-Observation Questionnaire
- ISIOP Data Collection Instrument
- ISIOP training materials, including videos, transcripts, and coded examples

## ISIOP Conceptual Framework Overview

New national science education frameworks, such as those recently released by the National Research Council<sup>1</sup>, emphasize the importance of instruction that supports students' adoption of and engagement in scientific practices. The pedagogical approaches to teach about and through scientific investigation clearly play a significant role in science teaching and learning. However, to conduct research on what has traditionally been termed “inquiry-based” instruction and to illustrate the components of science instruction that are potentially associated with student outcomes, tools are needed to document such practice.

The development of this instrument began with an extensive review of the literature on inquiry-based instruction through the National Science Foundation-funded project, *Has Inquiry Made a Difference? A Synthesis of Research on the Impact of Inquiry*. This work demonstrated that the literature on inquiry-based practices indicates that emphasis on student active thinking and drawing conclusions from investigations results in greater student achievement. Yet, little emphasis had been placed on researching the specific strategies that teachers use to achieve these goals, and on creating methods for documenting the specific strategies that teachers use to encourage student thinking and responsibility. Further, it is clear that supporting students' engagement in scientific practices occurs through many types of teacher moves and not just through engaging students in the more traditional steps of an investigation. For example, through discussion or asking challenging questions, teachers can encourage students' reasoning and scientific argumentation skills.<sup>2</sup> The first conceptual framework for the ISIOP was shaped from this blending of various bodies of literature on the nature of scientific inquiry in addition to an examination of existing instruments. This conceptual grounding in the literature provided the groundwork for the development of items contained in the protocol and established one early line of evidence for content validity of the items. While a few tools have previously been developed for conducting classroom observations in science, none captures comprehensive yet detailed information about the specific teacher-student interactions and components of instructional practice that can facilitate understandings of science instruction and inquiry learning.

Instruction is a complex, dynamic, and creative interaction between a teacher and student. In an effort to capture and describe this phenomenon, we have developed a protocol that records the variety of aspects of teaching that can take place in secondary science classrooms. One of the unique aspects of the ISIOP is the data collection structure that distinguishes between what a teacher says in his/her interactions with students, the kinds of activities in which students engage, and the presence of scientific practices. The Verbal Practices and the Classroom Instructional Leadership Practices that are represented in the ISIOP are drawn widely from the literature on effective teaching across academic disciplines and grade levels, whereas the Investigation Experiences and Content Checklists are drawn from science education literature exclusively and are somewhat grade-level specific. However, underlying all of the protocol is a constructivist approach to teaching, which focuses on engaging students physically and cognitively in the act of learning rather than relying exclusively on passive transmission by a knowledgeable other, i.e., the teacher. Aspects of scientific inquiry are reflected in the Investigation Experiences and in the Nature of Scientific Inquiry subsection of the Science Content Checklist.

---

<sup>1</sup> National Research Council. (2012). *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K–12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

<sup>2</sup> Minner, D. M., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496.

## Use Considerations

The ISIOP is a tool for researchers to document the nature of instruction in scientific instruction, including the multiple ways in which teachers support the development of students' scientific habits of mind. This tool is built on the assumption that teaching is a very complex phenomenon, as there are many aspects of instruction that teachers consider and manage throughout the duration of any lesson. The ISIOP focuses on the aspects of Verbal Practices, Classroom Instructional Leadership, and Nature of Scientific Investigation Experiences. Given that the ISIOP was designed to comprehensively capture the nature of the teachers' support for scientific practices—including a heavy reliance on what a teacher says—there are several issues that should be considered before incorporating this protocol into an evaluation or research project.

1. The constructs and items within the ISIOP are based on aspects of instruction that have either been theorized or demonstrated to be associated with quality instruction. However, the tool is not meant to provide evidence of quality instruction for administrators; rather, its intended purpose is to be able to document for researchers the extent to which various aspects of instruction are present in classrooms.
2. Science lessons in which there are likely to be minimal teacher-student verbal interactions (e.g., assessment activity or test) will not provide much data on the ISIOP because the main source of data (i.e., Verbal Practices) will be limited. In selecting lessons to observe, this potential limitation should be discussed with the teacher prior to the observation. Additionally, in schools where the teaching philosophy may limit the kind of dialogue that takes place in the classroom, such as completely student-directed instruction (e.g., Montessori schools), this protocol may have limited utility.
3. To adequately capture instruction—including the activities, Verbal Practices, type of content, and level of engagement of students—an observer needs to be able to focus attention on the small details of interactions while also being able to maintain an awareness of the big picture, such as the types of class activities or the ways that the teacher establishes the classroom climate or organizes the lesson. Thus, raters who are finely attuned to details to the point of not being able to maintain the bigger picture, or raters who hold the larger picture to the point of not being able to keep track of the finer-grained interactions, will not be able to use the ISIOP to capture the full extent of classroom instruction.
4. The ways in which an observer codes a particular classroom practice and, thus, the data generated by any observation protocol are influenced by his/her perspective, which is informed by his/her background, experience, and personality. The ISIOP is no different, as the real instrument is really the observer; the protocol is a tool that captures the observer's perspective. For example, raters with classroom teaching experience might code a teacher's moves differently from one who has never taught. Additionally, raters who have previous teaching experience may also view the content and Classroom Instructional Leadership differently (e.g., by giving the teacher credit for more complexity) from those with less extensive backgrounds in this area. Therefore, users should be mindful of discrepancies that arise during training between coders who have varied prior experiences. The training procedures offer opportunities to probe raters' understandings of observed practices in order to reveal some of the differences in interpretation of events.
5. To complete the training process, coders should commit to spending approximately 20–25 hours learning and practicing using the protocol before doing observations in live settings. We found from our field test that the most efficient way to learn the protocol is to complete training in a short period of time (over a 2- to 4-week period) rather than extending the training over a longer duration such as an entire semester.
6. Research/evaluation team training should, if possible, be structured so that each team member pairs up with one partner for training. The reason for this is that an integral part of the training process consists of debriefing sessions where coding discrepancies are compared and discussed. If there are more than two people in these conversations, they take longer for everyone to participate, increasing the total amount of time required for training. If simultaneous training is going to take place with larger training teams, the time to complete training will also increase. For research teams greater than two people, we suggest rotating training partners after each step in the training so that everyone has an opportunity to build a

consensus understanding of the protocol by working with all team members over time, while still only having debriefing conversations in pairs.

7. Language is inherently cultural. The Verbal Practices section was designed to capture aspects of spoken teacher instruction that may be related to student outcomes. However, the ways in which teachers use language in their classrooms depends on the cultural backgrounds of the teacher, the students, and the culture of the classroom or school. As described above, the ISIOP captures the observer's perspective, which is also inherently cultural. To the extent possible, observers should be aware of this cultural relationship and code the teachers' instructional practices accordingly. During the training, discrepancies in two observers' ratings and perspectives of the same lesson can be discussed in debrief sessions.



## ISIOP Training Procedure

The ISIOP is based on a comprehensive framework of teachers' practices, different from that present in other protocols, such as those that focus on practices across entire lessons or on student responses. The work of training to use the ISIOP involves becoming familiar with the underlying framework and, at times, adopting new terminology and ways of thinking about teacher practice. The ISIOP coding system is complex and requires several practice sessions (about 10, totaling approximately 20 hours) before a coder can internalize the coding scheme well enough to code in real time in a classroom setting. Therefore, we have provided a staged procedure that begins with the least demanding parts of the coding scheme and builds to the more demanding aspects.

As stated in the Use Considerations section above, the training procedures assume that users are training to use the protocol as part of a **paired team**. As such, the goal of the training is to familiarize the users with the protocol to the point where two raters can reliably code the same segment of instruction. An integral part of the training process consists of debriefing sessions where coding discrepancies are compared and discussed. If research teams consist of more than two members, we suggest rotating pairs after each step in order to minimize the time required for debriefing sessions.

Details on the reliability calculations and thresholds are provided in Scoring Guidelines, which begins on page 13 of this document.

**NOTE:** This protocol can be used to describe either middle grades (5–8) or high school (9–12) science instruction. However, the Science Content Coding Rubric that the observer uses will differ between the two school settings; thus, we have created two versions of this checklist (pp. 10 and 11, respectively, of the Data Collection Instrument). The high school version has more items overall. Please choose the checklist that is appropriate for the grade level being observed.

## Overview of the Training Steps

ISIOP training involves a scaffolded approach that begins with an overview of the ISIOP products followed by a gradual introduction to the steps for coding each section of the protocol. Support for using the protocol is provided by our video library. After practicing coding the entire protocol, teams should check reliability using the spreadsheet located on the website. The steps for training include:

1. Locate and familiarize yourself with the ISIOP products
2. Coding Lesson Events
3. Coding Verbal Practices and Investigation Experiences
4. Coding Classroom Instructional Leadership Practices
5. Coding the Entire Protocol
6. Checking Inter-rater Reliability
7. Additional Practice (if needed)

### Step 1: Locate and Familiarize Yourself with the ISIOP Products

The first step in training is for you and your team to become familiar with the ISIOP Data Collection Instrument and Codebook by reading both carefully. These documents are found on the “Home” tab of the ISIOP website (<http://isiop.edc.org>). Also on the website are other resources that you will need for your training:

- Teacher Pre-Observation Questionnaire (for use in live observations)
- ISIOP training videos
- Sample observation transcripts, coded and uncoded
- Completed coding sheets for comparison
- A spreadsheet for calculating reliability
- Q & A document

Due to consent agreements with the students' parents, we are only able to disseminate the videos of classroom teaching via a password protected website. So, although the majority of ISIOP materials are located on an open website, **you will need to obtain an account in order to access the training videos.** You will see the place to obtain an account on the "Video" tab of the ISIOP website. You will be prompted to enter some basic information about yourself and a username. You will then be notified by e-mail when a member of our staff approves your account. At that time, you can create a password and will have access to the EDC training videos.

The majority of the training videos are of middle school instruction; however, three videos—Measuring Trees, Biochemistry Review, and Identifying Compounds—are of high school classrooms. Beginning with Step 5, when you and your team begin to code the Science Content Checklist, it will be important to know whether the video you are viewing is of middle or high school instruction. At that time, please ensure that you use the appropriate content checklist for the grade level of those videos.

Please familiarize yourself with the ISIOP materials, noting which materials are needed during the observations and which are needed immediately after observations have been completed. Specifically, the Codebook provides examples and guidance and should be frequently referred to during training. The Data Collection Instrument contains:

- Classroom observation pages and short summaries of the codes (pp. 1–5), which you should have during all observations
- Post-observation coding rubrics (pp. 6–11 of the instrument), which can be completed after the observation.

Throughout the training, we recommend that you have a hard copy of the Codebook and Data Collection Instrument available in order to record your own notes. Though our Codebook provides definitions and examples for each item, it is often helpful for observers to record their own examples of the indicators they observe in the training videos.

## Step 2: Coding Lesson Events

This step in the training will introduce you to the coding of Lesson Events. After you have familiarized yourself with all of the ISIOP products listed in Step 1, review carefully the Lesson Event sections in the Codebook (pp. 2–8) and Data Collection Instrument (pp. 2, 4).

Lesson Events are defined as identifiable segments within the lesson marked by shifts in the flow of the lesson. Lesson Event information is recorded on the top of the Classroom Observation Sheet. Lesson Event information includes the start time, class activity, student organization, and level of student disengagement. Pay particular attention to the definitions and examples of the codes for the Lesson Event information detailed in the Codebook.

### Preparing for Coding

1. Obtain an account to access training videos (see Step 1).
2. Initially, observers work closely with team members to code videos simultaneously. However, for many videos, each observer on a team should work independently to view and code videos. It is important that you communicate with your team members about when you will each be viewing and coding videos so that you can plan a time to debrief and review your coding within a short period of time after having viewed the videos.
3. Go to the ISIOP website: <http://isiop.edc.org>
4. Print off several copies of the Classroom Observation Sheet (p. 2) from the Data Collection Instrument.
5. Please note and have available the list of Lesson Event Codes (p. 4) from the Data Collection Instrument.
6. Have the Codebook available for reference.

7. Make sure you have the most recent version of QuickTime. If you do not, please go to <http://www.apple.com/quicktime/download/> to download and install the latest version for free.
8. Have your computer sound turned on so that you can hear the video.

### **Coding and Debriefing Lesson Events**

1. Find the Measuring Trees video in the video library.
2. View and code the video simultaneously with your partner. Use the top portion of the Classroom Observation Sheet to capture the Lesson Event information and note the end time of the lesson at the bottom of the Observation Sheet. Use the time codes on the video player (bottom left of screen).
3. You can pause the video when there are transitions in Lesson Events to discuss your coding with your partner and consult the coding rules and code definitions in the Codebook.
4. Discuss your coding and your interpretation of the definitions of the codes with your partner using the following questions as guidelines to hone in on sources of discrepancies, questions, and confusions.
  - a. **Lesson Events**
    - First, round each time to the nearest minute. Do you have the same number of Lesson Events?
    - If yes, do the start and end times match within 1 minute?
    - If not, which Lesson Events have differing start or end times? Why?
    - Are there rules in the Codebook that could help clarify when to begin or end a Lesson Event that is problematic?
  - b. **Lesson Event codes**
    - For matching Lesson Events (with matching start and end times), where did you agree on activity codes? Classroom organization codes? Disengagement codes?
    - For the disengagement codes, did you and your partner have any that were only one position apart on the scale (i.e., one of you coded “none” and one coded “few”)?
    - Were you coding disengagement based on the number of students who were displaying the behaviors noted on Codebook page 6, for the majority of the Lesson Event time?
5. Compare your codes with Lesson Event information for the coded Examples provided on the website. Review the Codebook to determine the reason for any discrepancies you notice with the coded examples.
6. Repeat the process of coding, comparing, and discussing Lesson Event information with the Heating and Cooling Air video. Use the time codes on the video player (bottom left of screen).
7. By the time you complete these coding experiences, you should have seen examples of several, but not all, of the class activity and student organization codes, as well as have a basic familiarity with coding Lesson Events. You should be able to agree on how to segment a lesson into Lesson Events and to agree most of the time on the Activity code to best describe each Lesson Event.
- 8. Keep the coding sheets for these two lessons; you will need them again in Step Three.**

### **Step 3: Coding Verbal Practices and Investigation Experiences**

During this step of the training you will become familiar with the Verbal Practices and Investigation Experiences sections of the Data Collection Instrument by coding both transcripts and videos. You will also be introduced to the Observations Notes page, where you can record any information about the context of the lesson or indicators that could help you think about the types of Investigation Experiences. Throughout this step of the training, please take notes in the Codebook of any examples you notice of indicators, particularly those indicators that you may find confusing.

Begin by reading the sections regarding Verbal Practices codes in the Codebook (pp. 9–15) and Data Collection Instrument (pp. 2, 5). Pay particular attention to the examples and the differences between codes. Think of examples of teacher utterances that you have heard in the past and how they might be coded.

### **Coding a transcript**

1. Go to <http://isiop.edc.org>.
2. Have available the list of Teacher Verbal Practices Codes from the Data Collection Instrument (p. 5).
3. Have the Codebook (pp. 9–15) available for reference.
4. Download and print the Uncoded Transcript for Sunscreen Planning.
5. In the spaces provided on the Uncoded Transcript, code the Verbal Practices for the highlighted utterances. Each utterance is highlighted so that you get a sense of “grain size” for different kinds of Verbal Practices. We have highlighted everything that should be coded as a Verbal Practice for this lesson.
6. Return to the website and compare your codes with the Coded Transcript.
7. Discuss your coding with your partner, noting any discrepancies, and refer back to the definitions and examples of the codes listed in the Codebook, as needed. In your discussion consider:
  - Were any utterances particularly difficult to code?
  - Were any codes used only by you or by your partner?
  - Were any codes particularly difficult to notice or distinguish from another code?

### **Preparing for coding video**

1. Have available the list of Teacher Verbal Practices Codes from the Data Collection Instrument (p. 5) and the Codebook for reference.
2. Print off several copies of the Classroom Observation Sheet from the Data Collection Instrument (p. 2).
3. Have your computer sound turned on so that you can hear the video.

### **Coding two video clips and debriefing**

1. During this step you will use the main body of the Classroom Observation Sheet to capture the teacher Verbal Practices counts in the lesson. For this first video only, put all of your counts in Lesson Event #1 so that you don’t have to worry about doing both Lesson Event and Verbal Practice coding simultaneously.
2. Go to <http://books.heinemann.com/shared/onlineresources/E00948/HammerClip.mov> and, with your partner, watch the 6-minute video clip about pendulums. You can pause the video as often as necessary to discuss your coding or utterances where you have questions.
3. Locate on the ISIOP website the video clip for Evolution Day 2.
4. This time you should try to code both Lesson Events and Verbal Practices for this 10-minute video clip. Therefore, you should be sure to put your Lesson Event counts in the proper column for each Lesson Event.
5. Discuss your coding and your interpretation of the definitions of the codes with your partner using the questions outlined above (item 4 in Step 2, Coding and Debriefing Lesson Events, and item 7 in Step 3, Coding a Transcript) to hone in on sources of discrepancies, questions, and confusions.
6. Compare your codes with coded examples provided on the website. Review the Codebook to determine the reason for any discrepancies you notice with the coded examples.

### **Coding two full-length lessons with accompanying transcripts**

Read over the sections regarding Investigation Experiences in the Codebook (pp. 16–19) and Data Collection Instrument (pp. 6–7) so that you will have a sense of what to look for in the lesson.

1. Retrieve your coding sheets from Step Two for the Measuring Trees and Heating & Cooling Air videos where you coded the Lesson Event information. Use these coding sheets to code the Verbal Practices within each Lesson Event, and Investigation Experiences for the lesson as a whole.
2. Print from the website or make two copies from the Data Collection Instrument of the Coding Rubric for Investigation Experiences (pp. 6–7 of the Data Collection Instrument).
3. There are accompanying uncoded transcripts for you to reference while you are watching the videos. You may want to code the Verbal Practices on the transcript directly and then transfer the tallies to the

Classroom Observation Sheet or you could just consult the coded transcript and Classroom Observation Sheet when you have completed your coding to determine where your discrepancies lie.

4. At this point it is still ok to pause the video and code it in shorter segments (10-minute chunks), checking your coding with the coded examples as you progress. You should use the Observation Notes section of the Data Collection Instrument (p. 3) to make notes of any anomalous practices or investigation activities as they occur.
5. After completing the viewing and coding of Verbal Practices in these two lessons, take some time to rate the Investigation Experiences that were present.
6. Discuss your coding with your partner, noting any discrepancies, and refer back to the definitions and examples of the codes listed in the Codebook, as needed. In your discussion consider:
  - a. **Verbal Practices**
    - Were any codes used only by you or by your partner?
    - Is there any place where your tallies differ dramatically?
    - Were any codes particularly difficult to notice or distinguish from another code?
    - Were codes encountered for the first time?
  - b. **Investigation Experiences**—Compare the individual items you each selected for agreement on what was present (i.e., when you both selected the same item) as well as what was not addressed in the lesson (i.e., when neither of you selected an item). Then, compare your overall ratings given for each category of Investigation Experience.
    - How far off were your ratings—by one or more rating categories?
    - Are there differences in the evidence (e.g., what the teacher said, how the class was structured, how the room was arranged, types of student activities) you were each using to rate particular items?
7. If you feel you need additional practice, you can use any of the resources listed on the website under Step 7, Additional Practice.

## Step 4: Coding Classroom Instructional Leadership Practices

In this step, you will continue to practice coding Lesson Events, Verbal Practices, and Investigation Experiences, while adding in observations of the teacher’s Classroom Instructional Leadership.

### Preparing for Coding

Read over the sections regarding Classroom Instructional Leadership in the Codebook (pp. 20–23) and Data Collection Instrument (pp. 8–9) so that you will have a sense of what to look for in the lesson.

1. Go to the ISIOP website and print off from the Data Collection Instrument several copies of the Classroom Observation Sheet (p. 2) and two copies of the Coding Rubric for Classroom Instructional Leadership (pp. 8–9).
2. Have available the Lesson Event Codes and the list of Teacher Verbal Practices Codes from the Data Collection Instrument (pp. 4–5).
3. Have the Codebook available for reference.

### Coding in Pairs and Debriefing in Real Time

1. On the video tab of the ISIOP website, find the Density video.
2. For this video, you and your partner will code the videos at the same time by calling out the codes for the Lesson Events and Verbal Practices that you observe. This real-time comparison will allow you to uncover questions or differences in your interpretations of the codes and to pause the video to resolve the discrepancies as they occur.
3. After you are done with this joint coding, take a few minutes to separately code the Coding Rubric for Classroom Instructional Leadership, then discuss together your coding and compare it with the coded example.

## Coding Separately and Debriefing

1. Locate the Biochemistry Review video on the ISIOP website.
2. Using the entire Classroom Observation Sheet, you and your partner will code the video separately.
3. In beginning to code instruction in real-time, you may find it necessary to pause the video to record your codes. However, you should soon be able to code without pausing the video.
4. After coding the Lesson Event and Verbal Practices, fill out the Coding Rubric for Classroom Instructional Leadership.
5. Discuss your coding and your interpretation of the definitions of the codes with your partner using the guidelines provided above and below to hone in on sources of discrepancies, questions, and confusions.
  - How far off were your ratings—by one or more rating categories?
  - Are there differences in the evidence you were each using to rate particular items?
  - Was one of you more than the other “reading into” the teacher’s intentions?
6. Compare your codes with coded example provided on the website. Review the Codebook to determine the reason for any discrepancies you notice with the coded examples.

## Step 5: Coding the Entire Protocol

In this step, you will put all of the sections of the protocol together by coding entire videotaped lessons. Begin by reviewing the remaining section of the Data Collection Instrument and Codebook—the rubric for emphasis on science content.

Code the following videos: Rust Lab and Exercise Lab. Code these on your own and with minimal pausing of the recording. Take notes of any questions you have about coding, either while coding or immediately afterwards. After coding each lesson, complete the protocol Section 3—Post-Observation Coding Rubrics (emphasis on science content, Investigation Experiences, and instructional leadership). Note that the two videos in this step—Rust Lab and Exercise Lab—are both of middle school instruction. Therefore, you will need to use the Content Checklist for Middle School, located on pages 10 and 11 of the Data Collection Instrument. After each lesson, you should compare your coding with your partner’s and the coded examples.

Make notes in your Codebook to help refine your interpretations of the codes. At this point, you should be able to code Lesson Events and Verbal Practices without pausing the video. You should also have internalized many, but possibly not all, of these codes so that you don’t need to refer to the codebook or list of codes very often. If you need additional practice, please see the videos located in Step 7. Please note that if you need specific practice coding high school content, you should view the video Identifying Compounds, which is one of the Step 7 videos.

## Step 6: Checking Inter-rater Reliability

You are now ready to check for reliability. You should code the entire protocol on the following videos without pausing or stopping in order to simulate a live observation: Microscopes, Testing for Starch, and Liquids. Please note that the videos used in this step are of middle school instruction. Therefore, you will need to use the Coding Rubric for Science Content—Middle School, located on page 10 of the Data Collection Instrument. Afterwards, you will determine each partner’s inter-rater reliability with the coded examples on the website, as well as the inter-rater reliability between you and your partner. To do this, follow the instructions on the reliability spreadsheet provided on the website. Further information on methods for determining inter-rater reliability is available on page 14 of this User’s Manual.

## **Step 7: Additional Practice (if necessary)**

If at any point in the coding process you need more practice, below are additional resources you can use.

- Volume of Gases video
- Wind video
- Evolution Day 1 video
- Identifying Compounds Video
- Sunscreen Planning video
- Rust Lab transcript
- Liquids transcript

## Preparation for Live Classroom Observation

In preparing to conduct live observations, please read these helpful hints and reminders to make your transition to real-time observation smooth.

1. Since this protocol captures teacher-talk, it is essential that you be able to hear the teacher in order to use it for data collection. Therefore, we have found it helpful to put a remote/portable microphone on the teacher so that we can listen through the receiver with an earpiece in one ear to hear the teacher as he/she is walking around the room and speaking with individual students. This is much less intrusive than following the teacher around the room. In the absence of a remote microphone, ensure that the observer(s) are located in a part of the classroom where the teacher will be audible through most of the lesson.
2. Send an electronic version of the Teacher Pre-Observation Questionnaire to the teacher several days in advance of the observation. Ask him/her to fill it out and have it ready for your visit. A quick read of the questionnaire upon your arrival to the classroom should give you a very good idea of the substance of the lesson to be observed as well as its place in the unit of study.
3. We recommend conducting two practice live observations prior to beginning research or evaluation data collection activities so that you become familiar with the coding process in situ.
4. Bring stop watches to record start and end times of the lessons. Begin the lesson at time 0:00.
5. If you are observing with a partner, establish a signal beforehand for when to begin the observation. In our experience, a simple head nod to begin the observation suffices.
6. Bring a clipboard or something to write on, extra Classroom Observation Sheets, and the lists of Lesson Event and Verbal Practices Codes.
7. Once you have completed the live observation, be sure to complete the post-observation sections (Science Content, Investigation Experiences, and Classroom Instructional Leadership Rubrics) as soon as possible.



## Scoring Guidelines

### Establishing Inter-rater Reliability

This section describes how to calculate inter-rater reliability for each section of the Data Collection Instrument. The instructions are intended to provide guidance on the process rather than statistical programming details. We will leave to you such decisions as how much to automate and how much to calculate manually, what software to use, how to lay out the data, and how many paired observations to store in the same file or worksheet. These are aspects of the analysis that you should consider based on your project needs and resources. The procedures outlined below may also be used to calculate inter-rater reliability for Step 6 of the training procedure (p. 11). However, if you are working on Step 6 and do not want to learn about the reliability calculations at this time, you may simply use the pre-formatted spreadsheet provided on the website.

For the purposes of these instructions, a *lesson* refers to one class session, either live or video recorded. An *observation* is the set of scores generated by one rater observing one lesson. The unit of analysis for inter-rater reliability is the *paired observation*, which means the set of scores generated by two raters observing the same lesson.

Usually it will be desirable to determine the average inter-rater reliability for a group of paired observations. For example, if a team of two raters were working together on Step 6 of the training procedure, and you wished to calculate average reliability across all three videos *between the team members and between each of them and the EDC examples*, you would consider six pairs of observations: three pairs of observers (Rater A vs. Rater B, Rater A vs. EDC, and Rater B vs. EDC) multiplied by three videos. There are five sections of the Data Collection Instrument—Lesson Events, Verbal Practices, Investigation Experiences, Classroom Instructional Leadership, and Science Content. Therefore, these comparisons are made and averaged for each section of the protocol. Although different combinations of raters are possible depending on your research design, the examples that follow will illustrate comparisons between two raters only, for simplicity. The remainder of this section describes the step-by-step procedure for calculating reliability for each section of the protocol. Your team may choose to follow the procedures below, or you may use the spreadsheet available on our website for calculating reliability for all sections except Verbal Practices, which requires use of SPSS.

#### Lesson Event Variables

1. For each paired observation, compare the alignment of Lesson Events recorded by the two raters. If both raters counted the same number of events in a lesson, and if the start and end times that they reported for each event (after rounding to the nearest minute) were no more than one minute apart, the lesson is considered to be similarly segmented with 100% matching Lesson Events. Paired observations that do not meet these criteria are considered to have differently segmented lessons and should be analyzed to discern which Lesson Events are matched and which are not, and to determine the source of disagreement. Table 1 illustrates the analysis of a hypothetical paired observation. Qualitatively, the major difference between the observations was that Rater 2 interpreted the time between 00:27 and 01:10 as one hands-on activity, whereas Rater 1 saw a hands-on activity followed by a writing activity. Consequently, Rater 1 has an “extra” Lesson Event. In this case, the source of the disagreement was the inclusion of writing into a hands-on activity. The two raters should discuss their observations and consult the definitions of the activity codes in order to consider whether the writing was a primary task, separate from the hands on activity, with a clear start time.

Table 1. Analysis of a paired observation with mismatched Lesson Events

Rater 1					Rater 2				
Lesson Event #	Start Time (mm:ss)	Activity Code	Organization Code	Disengagement Code	Lesson Event #	Start Time (mm:ss)	Activity Code	Organization Code	Disengagement Code
1	00:01	INST	W	NONE	1	00:01	INST	W	NONE
2	04:27	HAND S	G	NONE	2	04:27	HAND S	P	NONE
3	17:50	WRIT	G	NONE					
4	23:10	INST	W	[Missing]	3	23:10	DISC	W	FEW
5	31:13	READ	I	NONE	4	33:16	READ	I	FEW
END	45:20				END	46:20			

- Next, determine a reliability score. Each Lesson Event has five variables: Lesson Event number, start time, activity code, organization code, and disengagement code. The total number of possible points for a paired observation is five times the total number of different Lesson Events observed by the pair of raters, plus one point for the end time of the entire lesson. In the example in Table 1, the total number of Lesson Events is five, and the total number of points available is 26.
- For each paired observation, deduct five points for an unmatched Lesson Event (since one rater has one extra Lesson Event, you are deducting one point for each mismatched variable created by the extra Lesson Event). For matched Lesson Events, deduct one point for each corresponding start time, activity, organization, or disengagement code that is different.<sup>3</sup> For start times to be considered different, they must be more than one minute apart after rounding each to the nearest minute. Deduct one point if the end times for the overall lesson are different, using the same criterion. Also, deduct one point for each corresponding code or time where one or both raters have missing data. The example would be scored as follows:
  - Deduct five points for the unmatched Lesson Event (the writing activity that was observed by Rater 1 but not by Rater 2).
  - Deduct one point for the difference in the organization code in the second Lesson Event (Rater 1 coded G and Rater 2 coded P).
  - In the Lesson Event beginning at 23:10, deduct one point for the difference in activity code (Rater 1 coded INST and Rater 2 coded DISC) and one point for the missing disengagement code.
  - In the next lesson event, deduct one point for the difference in start time (Rater 1 noted 31:13 and Rater 2 noted 33:16; so these times are more than two minutes apart) and one point for the difference in engagement code (Rater 1 coded NONE and Rater 2 coded FEW).
  - Do not deduct anything for the lesson end times, because they are only one minute apart.
  - The total point count for the paired Lesson Event is 16 ( $26 - 5 - 1 - 2 - 2$ ).
- Convert the total point count into a percent of agreement. For the example, 16 divided by 26 equals 62%.
- To calculate the average inter-rater reliability across a group of paired observations, calculate the mean of their percents of agreement.

Raters should discuss sources of discrepancies. If average agreement is below the reliability standard for Lesson Event codes of 75%, more practice may be needed.

## Verbal Practices

Our analysis focused on the extent to which different observers noticed the same rank orderings of counts of different teacher Verbal Practices across an entire lesson. Within each observation, the frequency count for each Verbal Practice code was converted into a percentage of the total number of Verbal Practices logged by that rater. The set of Verbal Practices was then ordered according to frequency of occurrence. It is this order

<sup>3</sup> For a matched Lesson Event, do not deduct a point if the Lesson Event numbers are different, because this difference will already have been accounted for in a preceding unmatched Lesson Event.

of the code frequencies and not the actual number of utterances recorded, that are considered for reliability determinations. The Kendall’s tau-b, a test of rank-order correlation, was used to determine correlations between the ordering of the codes. A Bonferroni correction was used to determine statistical significance (alpha). In other words, for each pair, agreement was considered significantly greater than chance only if the p-value associated with the correlation was less than 0.05 divided by the total number of pairs in the sample ( $0.05 / 9 = 0.006$ ). If the correlation is high, that means that the counts of Verbal Practices by two raters match in relative terms—the raters may have counted a different number of practices, but overall they agreed on which practices were most frequent, least frequent, etc. The inter-rater reliability goal for this section was to achieve statistically significant and high (greater than 0.70) correlations on 80% of the pairs. Table 2 provides an example of Verbal Practices reliability calculations.

The step-by-step procedures are as follows:

1. First, find the total number of occurrences of each Verbal Practice for each observation. Consider each observation separately. Exclude the “missed utterance” code.
2. Add all 21 counts together to calculate the total number of utterances in the observation.
3. Calculate the proportion of the total utterances for each Verbal Practice code by dividing each count by the total. For example, if Rater A heard 120 total utterances in a lesson and 42 of them were coded as “Solicit” then the proportion of Solicit codes would be .35.
4. Match the 21 proportions for each paired observation in a dataset.

Table 2. Calculating Reliability of Verbal Practices.

Lesson	Verbal Practice Code	Proportion of total utterances	
		Rater A	Rater B
1	Solicit	.35	.40
1	Facts	.09	.09
1	Procedural	.04	.01
1	[etc. for remaining codes]		
2	Solicit	.24	.21
2	Facts	.20	.20
2	Procedural	.02	.01
2	[etc. for remaining codes]		

5. Run a separate Kendall’s tau-b test against the 21 paired proportions for each paired observation. (In the example above, you would get two Kendall’s tau-b values, one for Lesson 1 and one for 2.)
6. Report two metrics: the mean of the Kendall’s tau-b values for all of the paired observations in the sample and the percentage of the values that were statistically significant, given Bonferroni’s correction. For reliability purposes, the Kendall’s tau for each comparison should be at least .70, the average across lessons for a pair of observers should be at least .75, and the percentage of significant tau values should be 80%

### Science Content and Investigation Experiences Rubrics

The Science Content and Investigation Experiences sections contain two types of items: *checklists* of yes/no questions, which are discussed here, and *overall ratings*, which are covered in the next section (under Classroom Instructional Leadership). The Science Content and Investigation Experiences checklists are analyzed separately but use the same procedure. The goal of the reliability analysis is to determine the extent to which a pair of raters checks the same items. For this analysis, items are divided into three types: those that both raters agreed were present and checked off, those that one rater thought was present (checked) but the other rater did not, and those that both raters agreed were not present and left blank. Counts of each of these types of items are converted to a percentage in order to determine the percent of positive agreement (i.e., agreement that items were taught), disagreement (where only one rater thought it was present) and negative

agreement (where both raters agreed the item was not present). The inter-rater reliability goals for these sections are 70% overall agreement (the mean of the positive and negative agreement) averaged across all lessons and coding pairs in the sample.

The step-by-step procedure is as follows:

1. For each paired observation, count the number of items *checked by both raters*. This is the number of positive agreements (#POS).
2. Then, count the number of items checked by either one of the raters but not the other. This is the number of disagreements (#DIS).
3. Calculate the number of negative agreements (the number of items *left blank by both raters*) by subtracting the numbers of positive agreements and disagreements from the number of items in the checklist:

$$\#NEG = \text{total number of items} - \#POS - \#DIS$$

**NOTE:** The total number of items depends on the checklist being analyzed. The middle school version of the Science Content checklist has 13 items, the high school version has 18, and the Investigation Experiences checklist has 26.

4. Calculate the percent of positive agreement, which is the number of items checked by both raters divided by the number checked by at least one rater:

$$\%POS = \#POS / (\#POS + \#DIS)$$

5. Calculate the percent of negative agreement, which is the number of items checked by neither rater divided by the number checked by no more than one rater:

$$\%NEG = \#NEG / (\#NEG + \#DIS)$$

6. Average the percents of positive and negative agreement to get the percent of overall agreement for this paired observation:

$$\%OVR = (\%POS + \%NEG) / 2$$

7. To calculate the average inter-rater reliability across a group of paired observations, calculate the grand mean of their percents of overall agreement. Conceptually, data for a group of paired observations might look like those displayed in Table 3:

Table 3. Investigation Experiences Items (26 items)

Lesson	Raters	#POS	#DIS	#NEG	%POS	%NEG	%OVR
1	A vs. B	3	1	22	75	96	86
2	A vs. B	4	3	19	57	86	72
3	A vs. B	2	2	22	50	92	71
					Grand mean		76

8. The goal for the Science Content checklist is 70% average overall agreement. The goal for the Investigation Experiences checklist is 70% average overall agreement.

### Science Content and Investigation Experiences Overall Items, and Classroom Instructional Leadership Ratings

As mentioned in the previous section, the Science Content and Investigation Experiences sections contain two types of items: *overall ratings*, which are discussed here, and *checklists* of yes/no questions, which are covered in the previous section. The Classroom Instructional Leadership section includes ratings only. The Overall items from the Science Content and Investigation Experiences Rubrics, and the Instructional Leadership ratings are analyzed separately but use the same procedure. For these items, reliability is determined by calculating the percent of agreement across items.

The step-by-step procedure is as follows:

1. For each rubric, count the number of agreements (items where both raters assigned the exact same rating). For the Science Content and Investigation Experiences Rubrics, for missing items assume a rating of zero if the rater checked none of the corresponding yes/no checklist items; otherwise consider the rating to be missing. For Instructional Leadership, an absent rating is always considered missing.
2. Calculate the percent of agreement by dividing the number of agreements by the number of items with neither rating missing. The total number of rating items for Science Content is 4, Investigation Experiences is 9, and Instructional Leadership is 28.
3. To calculate the average inter-rater reliability across a group of paired observations, calculate the grand mean of their percents of agreement across observations. Conceptually, data for a group of paired observations might look like this:

Table 4. Classroom Instructional Leadership Items (28 items)

Lesson	Raters	#Agree	#Missing*	Calculation	%Agree
1	A vs. B	18	0	18 / 28	64
2	A vs. B	20	1	20/(28 - 1)	74
3	A vs. B	15	3	15 / (28 - 3)	60
				Grand mean	66

\*Number of pairs with at least one rating missing

4. The goal for the Science Content ratings is 80% average percent agreement. The goal for the Investigation Experiences ratings is 55% average percent agreement. The goal for the Classroom Instructional Leadership ratings is 60% average percent agreement.

## Analysis Procedures and Score Interpretation

The ISIOP was intended to enable researchers to observe and describe, in unambiguous and quantifiable ways, several aspects of science teaching that are considered relevant to instructional quality and/or student learning. The multiple types of data that can be collected with the ISIOP provide opportunities to explore relationships among different aspects of teacher practice, between this instrument and others, and between teacher practice and student outcomes. An initial analysis conducted by our team using ISIOP data consisted of summarizing scores into averages across lessons in order to create profiles of individual teachers<sup>4</sup>.

In addition, confirmatory factor analyses of the ISIOP field-test data provided evidence of underlying structure for the Verbal Practices and Classroom Instructional Leadership sections of the ISIOP. On these sections, researchers may consider using the ISIOP data to convert indicators into scale scores. More details about the items that comprise scales are found in the corresponding section of the Codebook (p. 15 for Verbal Practices and pp. 20–23 for Classroom Instructional Leadership). Data from the Lesson Event, Investigation Experiences, and Science Content sections of the ISIOP could be used to generate context variables (e.g., to see how Verbal Practices or Instructional Leadership techniques differ across different types of classroom activities) or counts (e.g., to test the relationship between the proportion of a teacher's lessons over time that include certain Investigation Experiences and the outcomes for his/her students). We expect that the use of the data generated from the ISIOP will vary according to the interest of the individual research team and is likely to change over time as the instrument is used by the field.

<sup>4</sup> DeLisi, J., McNeill, K., Minner, D. (2011). *Illuminating the relationship between inquiry science instruction and student learning: Results from three case studies*. Paper presentation. Annual meeting of the National Association for Research in Science Teaching (NARST), April 3-6, 2011.