HAS INQUIRY MADE A DIFFERENCE?
A SYNTHESIS OF RESEARCH ON THE IMPACT OF INQUIRY SCIENCE INSTRUCTION ON STUDENT OUTCOMES

Technical Report 5:
Operationalizing the Inquiry Science Instruction Coding Process

The Inquiry Synthesis Project

Center for Science Education

Education Development Center, Inc.

Final Update APRIL 2006

Please use the citation below when referring to this document:

The Inquiry Synthesis Project, Center for Science Education, Education Development Center, Inc. (EDC) (2006, April). *Technical report 5: Operationalizing the Inquiry Science Instruction Coding Process*. Retrieved [date retrieved], from http://cse.edc.org/products/inquirysynth/pdfs/technicalReport5.pdf.

This research is supported by a grant from the National Science Foundation (NSF) (#ESI-0101766). 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 NSF.

INTRODUCTION

This technical report is the fifth in a series of reports that introduce and explain structures and processes used for the Inquiry Synthesis Project. *Technical Report 1: Generating the Synthesis Sample of Studies* describes **Phase I** of the project (Report Collection) and the criteria used for conducting the search for research reports to include in the synthesis. **Phase II** of the project is the Coding Process, which has three stages:

1. Inclusion/exclusion coding
2. Inquiry-science-instruction coding
3. Research rigor, context, and findings coding

Technical Report 2: Conceptualizing Inquiry Science Instruction focuses on stage 2 of Phase II. It introduces and explains a framework that accommodates the great variability in definitions of "inquiry science" commonly used by practitioners, academics, and the public for describing the treatment—inquiry science instruction—and the rationale for and process by which the research team used that framework to develop a codebook for guiding systematic, detailed description of that treatment. This report is the second on stage 2 of Phase II—inquiry-science-instruction coding—and explains the operationalization of that codebook to determine the presence, extent, and nature of the inquiry in the instructional treatment in each research study. Those that were determined to contain some inquiry in the science instruction were eligible to continue on to stage 3 of coding.

CODING FOR THE INSTRUCTIONAL TREATMENT

Overview

The framework for describing the instructional treatment detailed in *Technical Report 2* was operationalized into the Inquiry Science Instruction Codebook. The codebook included specific items to assist coders in identifying the presence of science content, describing the nature of the students' engagement with that science content, and identifying whether and to what extent elements of inquiry—student responsibility, student active thinking, and student motivation—were present in five components of instruction—question, design, data, conclusion, and communication.

A threshold was set for a minimum amount of inquiry that must be present in order for a treatment to pass through to stage 3 of the coding—research rigor, context, and findings. Of the 364 studies coded for inquiry instruction, 82 were excluded because there was no evidence of the presence of any of the elements of inquiry in any component of the instruction.

The Coding Process

Prior to the beginning of the coding process, the codebook was tested and revised. During this phase, all members of the research team used the codebook to code the same study and then met as a group to review ratings and make revisions that addressed and accommodated differences in understandings. The development of definitions and examples for each item in the codebook was a critical component of the testing and refinement process.

After three rounds of codebook testing and refinement, it was determined that the coding process needed to proceed by consensual coding teams rather than individual coders in order to ensure appropriate levels of inter-rater agreement. Thereafter, each study was coded independently by two researchers on a coding team whereby each coder clearly documented their evidence to support their coding choice in the study text itself and on the coding sheet. Each coding pair then met to consensually code and to reconcile any differences in ratings. If differences could not be reconciled through discussion, a third, senior researcher was called in to resolve the question based on the evidence presented. Once agreement was reached, reconciled data were entered into the dataset. Due to this consensual coding process, inter-rater agreement was 100%. To ensure against "coder drift" between coding teams, the specific team members were regularly rotated.

OPERATIONALIZING THE CODEBOOK

Coding for the Presence of Science Content

When determining criteria for studies to be included in this synthesis, the research team decided that in order for any treatment to be considered "inquiry science instruction," it must include science content. The *National Science Education Standards* were used as the guideline for what would qualify as "science content." For the purposes of this study, an instructional treatment had to include at least one of the following four types of science content: (1) physical science; (2) life science; (3) earth and space science; and/or (4) inquiry science content. The reason for limiting the instructional science content of studies in this synthesis was a pragmatic one. These were the science content areas of most interest to the field, and if other science content areas were included, it would have expanded the scope of this synthesis beyond the resources available.

If an instructional treatment addressed some other aspect of science articulated in the content standards (i.e., science in personal and social perspectives, history and nature of science, and science and technology) but none of the four identified above, it was excluded at this point. For example, one study at first appeared to focus on life science, but during the coding process, it was determined to actually focus on the ethical questions around organ transplantation. Another study that at first appeared to focus on exploration in outdoor environments was determined to actually focus on the learning benefits of fishing. An additional issue was the categorization of inquiry science content. The team decided that inquiry content must be an explicit focus of the instruction in order to be considered the content of the lesson. For example, an intervention that focuses on life science *and* explicitly states that students will learn about manipulating variables in an experiment would be considered to have inquiry content. On the other hand, an intervention that focuses on life science by having students do experiments but doesn't *explicitly* teach students about how to manipulate variables would not be considered to have inquiry content as a focus of the lesson.

Coding for the Nature of Student Engagement with the Science Content

The next section of the codebook focused on the nature of the students' engagement with the content. This section was designed with the recognition that many people consider the manipulation of materials equivalent to inquiry, and/or that manipulation of materials is a prerequisite for inquiry to be present. This project does not take that position. Rather, the view of this project is that inquiry is a function of the items we have called "the elements of the inquiry domain": student active thinking, responsibility for learning, and motivation. As long as one or more of these elements are present in instruction, inquiry is present, with or without manipulation of materials (students' direct, firsthand experience with materials that illustrate or demonstrate an aspect of science content); with or without observation of scientific phenomena (observations of phenomena that can be controlled, such as chemical reactions, and those that cannot, such as the movement of the sun); and/or with or without the use of secondary sources (sources that range from textbooks to scientist experts to maps and data tables). At least one of these types of engagement with science content is necessary for inquiry science instruction to be present, but they alone are not sufficient. They must be accompanied by some level of expectation that students will take some responsibility for their learning and/or thinking actively about their learning.

Coding for the Presence of Inquiry

The core of this stage of coding comprised coding for the elements of inquiry. The Inquiry Framework (see *Technical Report 2*) organizes instruction into a matrix of the five components of instruction and the three elements of the inquiry domain. Organizing the coding around the matrix allowed the coders to be specific about the nature of the inquiry instruction present and the placement of that inquiry in the instruction so that in analysis of the data, we could explore whether or not the elements of inquiry, or components of the instruction, predicted differences in student content learning.

Coding for the Components of Instruction

Five components of instruction were considered in this coding process: questioning, design, data gathering, conclusions, and communication. Coding captured the presence or absence of each of these components in

an instructional treatment. Coders did not consider the extent to which a particular component was present—only its presence or absence. When a component was present, coders went on to identify the level of student responsibility for learning and student active thinking relative to that component.

The five components were defined in the following way for this study:

- Questioning—establishment, identification, or generation of a **question** to guide student work. We inferred that studies that reached this stage of coding included a question for which students were seeking an answer. Thus, this component was always coded as present.
- Design— establishment, identification, or generation of a framework, plan, or **design** to gather data that will answer an investigation question.
- Data—gathering, recording, and/or structuring **data.**
- Conclusions—generation of **summaries, interpretations, explanation, inferences,** or **implications** from the data.
- Communication—oral, visual, or written **communication** about data, results, or other aspects of an investigation.

Coding for the Inquiry Domain Elements

Student Responsibility and Student Active Thinking: The codebook lists examples of evidence that a coder should consider before making a final rating about the extent to which the *instruction demonstrated the expectation* that students would take part in the actions and/or behaviors described in these lists. These lists were non-exhaustive, so if a coder encountered other manifestations of responsibility for learning or active thinking, they were noted on the coding form and considered along with the other examples for the final rating. The final rating options in each inquiry domain element were *no emphasis*, *some emphasis*, *a lot of emphasis* in the instruction, or *not reported*.

A rating of *no emphasis* indicates that there is sufficient evidence of only a token or minimal expectation that students will take part in the actions and/or behaviors characterized by the inquiry domain element. A rating of *some emphasis* indicates that there is sufficient evidence of the expectation that students will take part in the actions and/or behaviors characterized by an inquiry domain element, but that expectation does not significantly shape and/or dominate that component of instruction. A rating of *a lot of emphasis* indicates that there is sufficient evidence of the expectation that students will take part in the actions and/or behaviors characterized by an inquiry domain element and that expectation significantly shapes and/or dominates that component of instruction. A rating of n*ot reported* indicates that although a component of instruction is present, there is no evidence of any kind with regard to the nature of the instruction in that component.

Two additional operating principles guided this aspect of coding. First is the principle that the ratings of *no emphasis*, *some emphasis*, and *a lot of emphasis* are in no way related to the *amount* of evidence. Thus, it is possible to have a very small quantity of evidence that there is "a lot" of emphasis on an inquiry domain element in a component and thus determine that the most appropriate rating for that component is *a lot of emphasis*. Second is the principle that the instruction may exhibit a high level of expectation for student responsibility and/or active thinking, and still be significantly shaped or facilitated by a teacher. For example, an instructor may craft the conclusion component of instruction to include a teacher-facilitated whole-class discussion with the intent that *students* will, in the end, decide how to analyze the data. Thus, even though the teacher plays a significant role in this component of the instruction, such a clear emphasis on "student responsibility" warrants a rating of "a lot."

Motivation: Originally, student motivation was treated like the other elements of the Inquiry Domain, with coders having to code for it within each component of the intervention. However, during the initial phase of coding, it became clear that motivation differed from the other two elements of the Inquiry Domain for two reasons. First, it is more difficult to align with a single component of instruction, and second, because it was seldom explicitly described in the studies. Therefore, the team decided that there should be a single rating for motivation for the whole intervention rather than for each component of instruction. Due to the

limitations of the reporting, even a single explicit indication of motivation was sufficient to code *some emphasis*. *No emphasis* was chosen only if there was evidence that motivation was clearly not an intended part of the instruction.

The Intersection of Components of Instruction and Inquiry Domain Elements
The ways in which the elements of the inquiry domain were considered for each component of instruction are discussed more specifically below.

Questioning: When determining how to code the question component for inquiry, coders considered the origin of the question, the selection of the question, and the nature of the question. Usually, there was a combination of student and teacher choice and/or direction in the question component. In some treatments, for example, teachers gave students a broad topic and then asked them to generate and/or select a more focused question within that topic. In other cases, students participated in the generation of questions and then the teacher selected a question(s) from those generated or the class made selections as a whole. In still other cases, teachers provided students with a list of possible investigation questions and allowed them to choose from the list. Coders considered these and other situations in light of the examples offered in the codebook and other contextual and descriptive information of the study to make their rating selections.

Design: Responsibility in design refers to who makes the decision about what the design will be. Active thinking refers to the process of considering the merits of the design in light of the investigation question and the process of finding the answer to the question.

Data Gathering: Responsibility in this component refers not to who *does* the data gathering, but rather, to who *decides* who does the data gathering. In studies where teachers directed the data collection process, even when the students collected the data, there was little student responsibility in this component. In contrast, student responsibility was present, for example, when students worked in small groups and had to assign roles and tasks within their groups. Active thinking in this component refers to students considering the merits of the data collection strategy, discussing when and how the data collection process is and is not working, and correcting that process along with way.

Conclusions: When coding the conclusions portion of the instruction, coders recognized that teachers who engage students in the act of drawing conclusions are, to some extent, demonstrating the presence of active thinking in the intervention. Conclusions have been broadly defined in the codebook, and coders took the nature of the conclusions and the process by which they were arrived at into consideration when deciding on the final rating.

Communication: For the purposes of coding for the presence of inquiry, communication is not synonymous with all student talk or discussion. The important considerations for inquiry are who decides what and when there will be opportunities to communicate, and the substance and nature of that communication. When inquiry is present in the communication component, it demonstrates the expectation that students will, among other things, decide when and how communication takes place; include presentation, debate, and defense of points of view in the communication; and focus on the science content of the instruction.

The presence of an assessment in which students convey data, results, or other aspects of an investigation is not, by itself, evidence of communication. There must be evidence that the conveyance was part of the instruction, not exclusively a source of student outcome data for the purposes of the research.

CODING RULES
To further assist coders in making more general decisions about what to consider in the coding process, several "coding rules" were developed. They are described here.

Defining the Student Instructional Intervention
The team devised a definition of "student instructional intervention" close to the beginning of the project as part of the inclusion/exclusion process (see *Technical Report 3*). In this second stage of coding, the definition was refined to make clear that an intervention could vary in length and duration, and from one to many lessons. Further, in studies where interventions took place over time, authors often included only one or two descriptions of individual lessons as illustrations of the intervention. These author-selected descriptions of instruction were considered as representative of the intervention as a whole and used for the coding process.

Inferences About the Student Instructional Intervention
Some studies contained incomplete descriptions of the intervention, making it difficult to identify the enacted curriculum and, therefore, to code for the presence of the inquiry elements. For example, a study might provide a full description of the instructional materials but provide little or no information about the actual instruction that used those materials. In cases where there was sufficient contextual information, it was permissible to infer about the presence or absence of a *component of instruction*, but a decision was made to not make inferences about the *presence or absence of inquiry elements* and, in these instances, a rating of "not reported" was made.

In other instances, descriptions of the intervention were intertwined with the findings. While it was permissible to use information about the intervention found in a findings discussion to inform the inquiry coding, it was not permissible to infer that merely because something was assessed and reported on, that it was part of the expectations for that instruction and, thus, could be evidence for the presence of inquiry elements. For example, an intervention description might provide a specific actual account of the questioning component of the *actual intervention* but provide no description of the design component. If the report provides a *general description of the curriculum* that includes a description of "design," the coder may infer the presence of design but may not make any inferences about the extent to which that design had elements of the inquiry domain.

Accommodating the Developmental Levels of the Students
While elements of the inquiry domain can be present for instruction provided for any student, the way those elements are emphasized in classrooms of different ages and abilities of students necessarily vary greatly. The examples in the codebook did not consistently include evidence for young children in early grade levels or students of varying abilities. Thus, a rule was created to reinforce the need to consider additional evidence of inquiry in instruction designed for young children or children of different abilities, and add those examples to the codebook lists.

Addressing Conflicting Information About the Instructional Treatment
The information available to coders was highly variable and sometimes conflicting, leaving coders to make their best informed decisions about whether to select "some emphasis" or "a lot of emphasis" for the presence of an inquiry element in a component. When several examples of instruction within a single intervention were presented in a study, and there was the *same amount of evidence* for both "some" emphasis *and* "a lot" of emphasis on a particular inquiry element, coders would base their rating on the example with the highest amount of inquiry. When there was conflicting evidence but there was more evidence for one rating than for another, coders would base their rating on the preponderance of evidence to make the rating for either "some" emphasis or "a lot" of emphasis.

Making Exceptions to the Rules
As stated above, there were many occasions when coders did not see concrete, specific examples of evidence for the presence or absence of inquiry elements in a study. On some occasions, however, the descriptive and contextual information was compelling in its contribution to illustrating the circumstances of the instruction and its intent. For example, a study may not provide explicit information about the design component of the instruction, but at the same time, it may provide a history of the lessons that came before, describe student work that demonstrated the nature of the design component in other lessons, and provide data from

interviews with teacher and/or students that strongly suggests the presence (or absence) of an inquiry element in the design component of that instruction. These cases were brought to the whole team or to a senior researcher, and together a decision was made whether or not to make exceptions to the rule requiring explicit evidence to determine an inquiry rating.

ONGOING QUESTIONS AND CHALLENGES

Throughout the coding process, new questions emerged. Those questions that could be addressed while the coding process was underway were, and small codebook adjustments were made. Other issues were identified as challenges or "lessons learned" for future efforts to conduct syntheses and/or to codify inquiry.

Distinguishing Elements of Inquiry and Their Relationships

While the three elements of the Inquiry Domain were developed as distinct, exclusive constructs, operationally, the distinctions were not always so clear. Some studies provided examples of instruction that demonstrated close relationships between the elements of inquiry that were difficult to extract from one another while others demonstrated relationships between the elements of inquiry that, on occasion, challenged the coding process.

Distinguishing the Reporting of the Intervention from the Intervention Itself

The amount of specificity in the descriptions of the interventions in the studies was highly variable. In some cases, this might be attributable to the author of the study. In other cases, it might be attributable to the constraints placed on the author by the publisher or sponsor of the research. Regardless of the reason, the variability of the description created limitations for coders. One limitation resided in the level of specificity with which the instruction could be described. Early on in the codebook development process, the team had identified highly specific components of instruction. Only after seeing that the studies did not provide information at that level of detail did the team decide to consolidate some of the categories into five.

Coding Comparison Treatments

When studies included a comparison treatment, these treatments were also coded for inquiry. It was common, however, for authors to limit their descriptions of comparison treatments and, thus, our ability to code them was impaired. Our ability to draw conclusions about the impact of inquiry science instruction compared with other forms of instruction will rely on authors' clarity and completeness of their descriptions of all the instruction they studied. Additionally, some studies included comparison treatments that differed in ways that were irrelevant to the inquiry components of instruction on our coding protocol, or the differences were so subtle that they were not discernable.

ADDITIONAL INFORMATION

For more information on this or other CSE research projects or to view additional technical reports, visit http://cse.edc.org/work/research/

Inquiry Project Staff
Daphne D. Minner, Ph.D., Principal Investigator (dminner@edc.org)
Abigail Jurist Levy, Ph.D., co-Principal Investigator
Jeanne Rose Century, Ed.D., co-Principal Investigator (August 2001-July 2005)
Erica S. Jablonski and Erica T. Fields, Research Associates

