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

Technical Report 3:
Operationalizing the Inclusion/Exclusion 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 3: Operationalizing the Inclusion/Exclusion Coding Process*. Retrieved [date retrieved], from http://cse.edc.org/products/inquirysynth/pdfs/technicalReport3.pdf.

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

INTRODUCTION

This technical report is the third 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 that 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 for describing inquiry science instruction and the rationale for and process by which the research team derived a systematic way to code for the treatment—inquiry science instruction—that accommodates the great variability in definitions of "inquiry science" commonly used by practitioners, academics, and the public. This report describes Stage one of Phase II—inclusion/exclusion coding. It explains the development of the codebook for report inclusion/exclusion and the operationalization of that codebook to determine which research reports had the potential to contribute to answering the research question of the synthesis project and, thus, were eligible to continue on to Stage 2 of coding. It also offers some perspective on the complexity of the inclusion/exclusion task.

REFINING THE INCLUSION CRITERIA

To identify research that reported on findings that would address our research question—*What is the impact of inquiry science instruction on student outcomes?*—we began with four broad parameters to assist with the search-and-retrieval process. Those four parameters were created to identify potentially appropriate research documents that would be reviewed in Phase I (Report Collection). These parameters established that studies had to

- examine one or more aspects of inquiry science teaching;
- assess student outcomes (e.g., science-subject-matter content knowledge, inquiry content knowledge, science attitudes, attitudes about instruction, school attendance, science course enrollment, science career choice, historical/social perspectives in science, participation in class, motivation);
- be completed in a written form between January 1, 1984, and December 31, 2002; and
- focus on students within the K–12 population.

These parameters were developed with the intent of "casting a broad net" to ensure that studies that could potentially be included would not be missed. We were aware that this broad net would capture a number of studies that ultimately would not be appropriate for inclusion. Thus, in the effort to operationalize these parameters in a more formal way for study identification and coding purposes, further refinements were necessary.

Early on in the coding process a decision was made to divide Phase II coding into stages in order to gather the key aspects of data—inclusion/exclusion, instructional treatment characteristics, and findings—in a manageable way. Each stage of coding would represent a filtering of studies that, in the end, would result in the final sample for analysis. For example, the refined inclusion criteria of Stage 1 focused on *science* instruction in general, recognizing that the specific kind of science instruction would be captured in the second stage of coding. Further, since the specific aspects of instruction (i.e., the characteristics that determine whether or not it is inquiry-based) would be coded in Stage 2, the search-and-retrieval process used working definitions of "inquiry instructional treatment" to locate potentially relevant documents. These definitions included a range of those identified by the authors, such as "inquiry," "hands-on," or some other closely related term. Then, as studies moved into Stage 1 of Phase II we applied more refined criteria to filter out those that, while meeting the original parameters, would not ultimately be appropriate for the synthesis. So, for Stage 1, in order to be included and move on to the next stage of coding, a research study had to meet the following criteria:

- focus on at least one student instructional intervention in science;
- report student outcomes;
- have at least one research question that was about the effect of a student instructional intervention in science on student outcomes;
- describe the student instructional intervention with sufficient specificity;
- be published or presented between January 1, 1984 through December 31, 2002; and
- be conducted with K–12 students.

OPERATIONALIZING THE INCLUSION CRITERIA

Confirming the Obvious: *Defining an Empirical Study*

The broad strategy employed in inclusion/exclusion coding was to progress from the most broad to the least broad of the criteria. Therefore, the first step in coding was to confirm that a document was describing an empirical research endeavor, which we defined as "a research endeavor that systematically collects data under a single research plan from a designated sample of respondents to answer one or more research questions." At this point, we began excluding documents that had been collected as part of our "broad net" search but did not fit this criterion. These included meta-analyses, literature reviews, commentaries, descriptions of instructional programs, and opinion/editorial pieces. Although efforts were made to screen these out during the collection process, in some instances the type of document was not easily discernible until a first careful reading of the piece, which took place in the inclusion/exclusion coding.

Then, we began to apply our inclusion criteria. The easiest to focus on were the last two listed above. Only empirical studies that had a publication or presentation date from January 1, 1984 through December 31, 2002 were included. Although this was part of the search criteria for identifying eligible reports (see *Technical Report 1*), this criterion was double checked and confirmed. Similarly, during this stage of coding, we confirmed that the research focused on K–12 students. Studies that included preschool or university students *in addition* to K–12 students, or studies that included the general public (such as those conducted in museums), were included only when the information about the K–12 students could be disaggregated and examined separately. Even though the focus of this synthesis is on K–12 students, the studies were not constrained to traditional K–12 classroom settings. We also included research conducted in experimental classroom settings (i.e., classroom settings created for the purpose of the study) and informal settings.

Next, we further examined the studies to confirm that each

- focused on at least one student instructional intervention in science;
- had at least one research question that was about the effect of a student instructional intervention in science on student outcomes;
- reported student outcomes; and
- described the student instructional intervention with sufficient specificity.

Defining the Student Instructional Intervention

To fit our criterion of having at least one student instructional intervention, we first had to define what an "instructional intervention" was for the purposes of our study. First, we determined that a student instructional intervention had to be within the domain of science and conducted, administered, or initiated by a teacher (or researcher or other instructor) for, to, or with students in a K–12 setting. We recognized that the studies had a variety of types of instructional interventions, ranging from regular classroom instruction to one-on-one interactions with students. "Regular classroom instruction," for example, had no specific prescribed intervention program or strategy that was added to or that replaced the "typical" instruction. In these studies, researchers documented and described practices in the classroom and their related outcomes.

Other interventions entailed a prescribed intervention program/curriculum/pedagogical innovation that was introduced into the classroom for the purposes of the research and then studied to determine the effect on student outcomes. Then, there were some interventions specifically designed to focus on eliciting student understanding of key science concepts or processes resulting from developmental phenomenon, such as the emergence of abstract reasoning, and not from an instructional/experimental treatment. More specifically, while these studies used experimental protocols that *could* be used in regular classroom instruction, their intent was to understand students' cognitive development of scientific concepts and processes, unrelated to any single instructional intervention students may have experienced. Basic developmental research studies were excluded when the intent of the research was an improved understanding of cognition, of how students learned, and/or of how they processed information, because our research question is about the effect of *instruction* (i.e., inquiry-based pedagogy) on student outcomes.

The Research Question
Another criterion for inclusion was that there had to be at least one research question on the impact of instruction on student outcomes. The research question(s) could be of any type, including exploratory, predictive, and hypothesis-testing. If no research question for a study was explicitly stated, but it could be inferred from the description of the research or from information provided about the intervention and related student outcomes, the study was still eligible for inclusion. If at least one research question was explicitly stated, no additional research questions were inferred.

Reported Student Outcomes
A next step was ensuring that the instructional intervention under study was related explicitly to reported student outcomes that could be considered dependent variables. In quantitative studies, dependent variables were typically stated at the outset and often assessed by a post-test. In qualitative studies, potential student outcomes also were sometimes stated at the outset, but often they emerged later in the study during the description of the intervention and analysis. These variations sometimes made it difficult to determine whether the outcomes were linked to the intervention. In cases where the connection between the two was hard to discern, project staff conferred as a group to reach a consensus decision.

Describing the Instructional Intervention with Sufficient Specificity
In anticipation of the second stage of coding (inquiry-science-instruction coding), a decision was made to exclude those studies that did not describe the enacted instructional intervention with a level of specificity that would allow us to code it. As part of this process, we decided that even the most meager description of what actually happened in an instructional setting would be sufficient to allow a study to pass through to the next phase of coding. This information could take many forms, as long as it met the basic requirement of describing what actually happened in the instruction setting during the time the data was collected. Acceptable descriptions could range from narrative about what students and teachers actually did to class transcripts. Even when an intervention had multiple sessions, specific description of only one was sufficient for inclusion.

The intent of the requirement of sufficient description while obvious on the surface is still worth articulating here. We recognized early on in this work that we needed to find a way to describe inquiry such that it would lead to a meaningful analysis. That process is described in *Technical Report 2*. Thus, to operationalize that framework, we needed to have descriptions of instruction that moved beyond general description. For our analysis to generate useful and practical findings for our audiences—policymakers, administrators, curriculum developers, and others—to make decisions about science curriculum and instruction, it was crucial for us to identify and specifically describe the nature of the instructional interventions that led to particular student outcomes, and this could only be done if the instruction was described specifically in each study.

Project staff encountered many studies that included no information about what actually happened in the classroom, or the descriptions were extremely limited. For example, studies were excluded if the description of all components of the instructional intervention or activity consisted only of a reference to or general

overview of a curriculum, set of instructional materials, or kind of instructional strategy rather than specific teacher and student activities. Examples of this include, "The class used *FOSS* kits," or "The science curriculum included lab activities and computer simulations." Other general descriptions of instruction that were insufficient included "inquiry," "hands-on," "experiential learning," and "constructivist."

We considered these descriptions to be insufficient because implementation of curricula, materials, and instructional strategies can be so varied that it would be misleading to make any assumptions about the nature of the instructional intervention based on these descriptions alone. However, if the author described a curriculum or instructional intervention and then indicated that through classroom observations it was known that the intervention was implemented as intended, the author's word was taken and the study was included.

Studies that evaluated the implementation of a particular instructional program were excluded if the evaluation findings focused on *program* implementation or on outcomes other than student outcomes. However, if the evaluation addressed research questions relating student outcomes specifically to the intervention, these studies were included.

This focus on specific descriptions of the instructional intervention resulted in the exclusion of many studies. However, we recognized that there would be value in keeping track of studies that met all other inclusion criteria but were excluded because of insufficient descriptions of the intervention for the reason that they offered potentially useful data. One specific example lies in those studies that used large-scale national samples on standardized instruments such as NAEP to look at the effect of regular classroom instruction on student learning. It would be impossible to draw any conclusions about the instructional components that these students received and what impact these components may have had on the NAEP scores. This is not the intended use of these kinds of assessment devices, but it is the intention of our synthesis project. Therefore, these studies were identified and set aside for possible later examination under a different research question.

Our initial search-and-retrieval process also identified studies that focused on professional development interventions. These studies were only included if (a) they stated a research question about the impact of the delivered instruction (presumably resulting from the professional development) on student outcomes; (b) there was a sufficient description of the student instruction; and (c) student outcomes were reported. Those that described only the professional development and the student outcomes were excluded; those that described only the professional development and the resulting classroom instruction were excluded; leaving only those studies of professional development reporting on classroom instruction and student outcomes eligible for the next stage of coding.

THE INCLUSION/EXCLUSION CODING PROCESS

The team's coding criteria were compiled in detail in a coding manual. First, coders determined whether the document met the definition of an empirical study of interest, meaning it was "a research endeavor that systematically collects data under a single research plan from a designated sample of respondents to answer one or more research questions." Then, coders sought to confirm that the study (a) focused on at least one student instructional intervention in science, (b) reported student outcomes that were associated with the instructional intervention; and (c) asked at least one research question about the effect of a student instructional intervention in science on student outcomes. If the document passed this threshold, the coders went on to confirm that the study pertained to K–12 students and was published between 1984 and 2002. Finally coders determined whether or not the document contained a specific description of the instructional intervention. If the study met these additional criteria, coders then noted the student outcomes that were related to the instructional intervention and the data sources from which the student outcomes were generated.

To begin, all studies were coded independently by three coders and then consensually coded for inter-rater agreement. During these rounds, the coding manual was iteratively revised until all coding items were clarified and the language of the codebook was specific enough to ensure consistent understanding and interpretation across the team of researchers.

Thereafter, each study was coded by two researchers, and then the pair met to consensually code so that disagreements could be reconciled. Evidence that supported each coding choice was both highlighted in the study itself and noted on the coding sheet with page numbers. Differences in coding were discussed based on researchers' interpretations of the evidence in the study. If disagreements persisted, a third senior researcher was called in to help reconcile coding. When questions could not be resolved in this manner, or when they raised a larger issue that required broader discussion, they were brought to the entire project staff for resolution. A record of the changes made in order to reconcile coding within the pair of coders was documented on each coding sheet. Once agreement was reached, reconciled data were entered into the computer archive.

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

