Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The Working Paper Series was created in order to preserve the valuable information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Suellen Mauchamer at (202) 219-1828 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N. W., Room 400, Washington, D.C. 20208-5652.

Susan Ahmed
Acting Associate Commissioner
Statistical Standards and Methodology Division

Samuel S. Peng
Statistical Service and Methodological Research

Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work

Prepared by

Mary S. Leighton and John E. Mull ens, Brenda Turnbull,
Lisa K. Weiner, and Angela S. Williams

Policy Studies Associates, Inc.

Table of Contents

Foreword ...111

Table of Contents ... iv

Preface ... vii

Part 1: MEASURING; INSTRUCTION: THE STATUS OF RECENT WORK
by Mary S. Leighton

The Role of Instruction Opportunity to Lear n ... 1

Recent Attempts to Assess Instruction as Part of OTL ... 17

The Implication of Recent Efforts to Assess Instruction for NCES Data Collection ... 31

References ... 34

Part 2: MEASURING CURRICULUM CONTENT: THE STATUS OF RECENT WORK
by Mary S. Leighton and John E. Mullens

1. Validating National Curriculum Indicators--RAND/UCLA/CRESST Research Teams ... 1

References ... 13

2. Reform Up Close (RUC)--University of Wisonsin Center for Education Research/Consortium for Policy Research in Education Team ... 1

References ... 16

Appendix A: Mathematics Content Codes ... A-l

3. Third International Mathematics and Science Study ... 1

References ... 10

Appendix A: Topics for In-Depth Topic Mapping ... 11

Appendix B: Aspects and Major Categories of TMSS Curiculum Frameworks ... 13

Part 3: MEASURING INSTRUCTIONAL RESOURCES: THE STATUS OF RECENT WORK
by John E. Mullens, Lisa K. Weiner, Angela S. Williams, and Mary S. Leighton

instructional Materials and Equipment 2

Facilities 4

Calculators, Computers, and Related Equipment 5

Implications for NOES 8

References 10

Appendix A: Selected Instruments Used to Assess Opportunity to Learn and Instructional Resources 11

Appendix B: Instructional Resources as an Element of Opportunity to Learn instructional Materials and Facilities 13

Appendix C: Instructional Resources as an Element of Opportunity to Learn Calculators, Computers, and Related Equipment Items 15

Part 4: MEASURING OPPORTUNITY TO LEARN: ADVANCING THE STATE OF THE ART
by Mary S. Leighton with the assistance of Brenda J. Turnbull and John E. Mullens

Purposes of NCES Studies of OTL 1

Key Dimensions of Context 2

Focus: K-12 Instruction 5

Focus: Eighth-Grade Math 6

Focus: Fourth-Grade Language Arts 8

Focus: U. S. History (Grades 4-12)... 9

Final Thoughts: Opportunities to Collaborate 10

Appendix A: QUESTIONNAIRE ITEMS FOR THE TEACHER FOLLOW-UP SURVEY EXPLORING OPPORTUNITY TO LEARN A-1

Preface

The four papers contained in this volume are four parts of series on the status of recent work on measuring instruction, curriculum content, and instructional resources. These papers are intended to identify aspects of instruction, to analyze the approaches used by several leading studies, and to describe the implications of recent work for NCES data collection efforts related to the opportunity to learn. The work documented in these reports was conducted by Policy Studies Associates, Inc. under Contract No. RN 9306100.

The authors of the papers presented in this volume are Mary S. Leighton and John E. Mullens, Brenda Turnbull, Lisa K. Weiner, and Angela S. Williams of Policy Studies Associates, Inc.

MEASURING INSTRUCTION: THE STATUS OF RECENT WORK

Mary S. Leighton

October 1994

Prepared for the U.S. Department of Education under a subcontract of Contract RN 9306100 between The National Center for Education Statistics and MPR Associates, Inc.

MEASURING INSTRUCTION: THE STATUS OF RECENT WORK

This paper has three purposes:

- To identify and list the aspects of instruction that shape students' opportunities to learn
- To analyze the approaches used by several leading studies to assess instruction in order to characterize opportunities to learn
- To describe the implications of recent work for NCES data collection efforts related to opportunity to learn.

The first section of the paper summarizes research findings about features of instruction that contribute to learning. It begins with process/product research results and then explains alternate approaches that add validity to the insights gained from correlational and experimental studies. The section concludes with an excerpt from a description of a second grade math class and a discussion of how analytic frames capture different aspects of the event. The second section of the paper describes how contemporary studies have approached assessment of opportunity to learn. It categorizes study questions and data collection strategies, and reports on the effectiveness of different strategies (to the extent that information is available). It shows how the information produced by these strategies maps onto the framework described in the first section, to demonstrate the extent to which they capture what other studies suggest are the determining aspects of instruction. The final section of the paper explores how the frameworks for understanding the contribution of instruction to opportunity to learn and recent experiences in assessing instruction can inform NCES efforts to assess opportunity to learn as a part of its national data collection programs.

The Role of Instruction in Opportunity to Learn

For the purposes of this series of papers, we have characterized opportunity to learn as a construct with three components: curriculum content, instruction, and instructional resources (such as materials, computers, texts, and labs). Certainly what students learn in any course depends in part on what was covered; if a topic never arose in school, student mastery could not be assumed as a consequence of schooling. The quantity and quality of instruction also influences learning. If a topic is covered too hastily or with erroneous explanations, for instance. it may leave no more trace in student achievement than if it were not covered at all. Likewise, the availability of resources to support learning-such as library books, online information retrieval systems, desks, and supplies—influences productivity.

These components have overlapping elements; in practice, aspects of one may also be aspects of another. "Curriculum content" includes not only major and minor topics covered, but also teachers' emphases and role orientation and student performance expectations, because, although they may be considered instructional rather than curricular variables, the latter features create operational definitions of each topic. For instance, a lesson may cover the concept of least common denominator, but how students experience the meaning of that concept will be influenced by whether it is a small digression in a bigger discussion of subtracting fractions, whether the teacher acts according to a conviction that her role is to stimulate discovery (rather than to explain), and whether as a result of the lesson students are expected to solve problems or only to name the common factors of two numbers. All of these cues communicate what the lesson is about, which is likely what the students will engage in learning. The curriculum content papers also touched on resource availability, to the extent of reporting the use of supplies, equipment, and facilities as instances of lesson content. In this paper, we examine the dimensions of instruction that sound research shows to have a clear bearing on student achievement and thus should be construed as influencing opportunities to learn.

The Results of Process/Product Research[1]

Process/product research examines the relationship between certain of teachers' overt and quantifiable instructional behaviors and the achievement of typical students. Extensive correlational studies, often supported by experimental studies and replicated in many settings, have identified four areas of instruction that significantly influence students' opportunities to learn across grade levels and subject matter:

- Amount and pace of instruction
- Presentation of information
- Questioning strategies

---

[1] The material in this section comes mainly from Brophy and Good's (1986) comprehensive research summary, which appeared in the Third Handbook of Research on Teaching.

- Reactions to student *responses*

Some evidence also suggests that particular features of seatwork and homework assignments make small but significant contributions to learning. The constitution of the instructional group-whether whole class. small group. or individual-has also been studied for its relationship to learning,

Until recently, many comprehensive studies of instructional effectiveness focused on models of direct teaching, and their findings apply most neatly to traditional teacher-centered, whole-class lessons. However, some findings have implications for other models of teaching that may involve more student-teacher or student-student interaction. In the following sections. such extrapolations have been suggested whenever they seemed supportable.

Amount and *pace of instruction*. According to Brophy and Good, "The most consistently replicated findings link achievement to the quantity and pacing of instruction" (1986, p. 360). Conventional measures such as the number of text pages covered or test items taught show strong, positive correlations with student achievement. Teachers who give top priority to academic activities and who allocate the most time to academics nurture the most learning among students. Maintaining an environment that supports students' sustained engagement in learning also contributes to achievement. The rewards of sustained engagement depend in part on students' success with the assigned tasks-if work is too easy, given the circumstances, it occupies students without advancing their knowledge or skill. Alternatively, if it is too difficult, it may hold their attention without generating understanding. The level of difficulty appropriate for different conditions varies. Very hard work may challenge and motivate a group of students who are deeply interested in a topic, while periodic assignments of very easy work-perhaps after a stretch of challenging exercises-may consolidate learning and promote facility. In general, active teaching strategies-those that (1) include stimulating exchanges between the teacher and students and among the students. (2) involve the teacher deeply as a mediator of learning, and (3) ensure that students experience high-quality representations of content–promote more learning. (This is in contrast to relatively unproductive situations in which students work unassisted trying to make sense of textbook explanations or other content exposure without the teacher's oversight to catch misconceptions or misinterpretations and monitor comprehension.)

Presentation of information. In lessons where the teacher serves as the primary source of information. usually by lecturing, demonstrating, and eliciting student recitation, several features boost student learning. Structuring material with advance organizers, periodic summaries of what has been done and what is ahead. and explicit reference to patterns as they emerge facilitate retention and give students better access to the big picture. Planned redundancy and careful sequencing of lesson elements and subtopics also stimulate learning.ry is an essential lesson element. Using appropriate analogies and examples makes it easier for students to master content. Showing enthusiasm and maintaining an appropriate level of momentum also contribute to learning.

Questioning strategies. Overall, the ways teachers frame questions and behave with respect to students' responses influence achievement. Both difficulty level and cognitive level are important considerations in questioning. Difficulty level reflects how hard it is for students to answer questions (regardless of complexity).' Most of the time, students learn more from lessons where most of the questions are readily answerable. In quick review sessions of basic facts, virtually all of the questions should be relatively easy to sustain pace and engagement, whereas in lessons intended to stretch students' cognitive limits, a higher proportion of questions may be difficult without compromising momentum. The issue of balance between lower and higher cognitive questions[3] is murkier, although the evidence suggests that overreliance on either produces little learning. For example, a lesson may begin with a rush of simple recall questions that serve to remind students about a body of information learned previously about the topic, but conclude with extended discussion of a single higher-level question that requires students to analyze and synthesize the facts gathered earlier. Practical competence depends on having easy access to some common knowledge and the ability to perform some complex cognitive operations. Research to date indicates that there is some relationship between the balance of higher and lower cognitive questions and student achievement, although the nature of the relationship is not well documented. Whether a question is difficult or easy, cognitively demanding or not, its clarity also influences students' capacity to learn from it.

Reactions to student responses. Once questions are asked, how the teacher manages students' responses also supports or restricts learning. Ordinarily, waiting a few seconds after asking a question before calling on a respondent positively affects the extent of student engagement. If the timing is right, more students will think through a question and prepare an answer while they wait to see who will be called upon to answer publicly; such mental rehearsal stimulates learning. Under some circumstances, selection of respondents can influence learning-sometimes, for example, patterned rum-raking may support engagement and learning, although other times, randomly calling

---

[2] Questions such as "What is the capital of .....?" involve only recall, a simple cognitive process. However, they are easy only when they apply to familiar material-a recently-studied or local state in this instance. When they apply to very new or foreign material-perhaps to a newly-formed country mentioned on the previous evening's news programs--they may be quite difficult.

[3] Cognitive level refers to the complexity of thought required to answer a question. Explaining how a certain city got to be the capital requires remembering details and then generating syntheses and analyses. It may or may not be hard for a student, but it is in any case complex.

on nonvolunteers who probably know the answer is a better strategy. Once the selection is made, how the teacher responds to correct, partially correct, and incorrect answers affects lesson productivity. Inappropriately extravagant praise for giving the correct answer to an easy question or failure to correct a wrong answer may communicate low expectations or leave a misconception uncorrected.

Seatwork and homework. Properly constructed seatwork or homework-if not used as substitutes for active instruction--promote learning when they provide opportunities to practice and apply new skills and knowledge, usually with a high rate of success. When teachers monitor students' performance on these tasks and provide feedback, students learn more. Teachers' treatment of students' products signals the degree of importance of each task and students' effort adjusts correspondingly.

Whole class, small group, and individualized instruction. Studies of the effects of teaching the whole class, small groups of different types, or individuals have not produced a simple directive. The variable that predicts learning is not group size, but rather student engagement in appropriate work, which can be orchestrated in a number of ways effectively. Student learning is supported if the whole-class lesson successfully engages every student in working toward the goal. Student learning is supported if the structure of the lesson activities for the small groups that are not being supervised by the teacher in a given period is sufficiently engaging. Student learning is supported if other individuals are productively engaged with their work while the teacher offers one-to-one coaching. In general, threats to learning come from the extent to which any strategy leaves students disengaged or occupied with unproductive busywork.

To summarize, this body of research construes opportunity to learn as a phenomenon characterized by certain enabling instructional conditions. First, content coverage is extensive and brisk. Second, information is presented in well-ordered fashion, with appropriate organization, clarity, and redundancy. Third, questioning strategies engage students in a productive balance of easy and hard, cognitivel y demanding and undemanding questions-a balance determined by the lesson content. Fourth, teachers' reactions to students' recitation and responses encourages participation without supporting low effort or misconceptions. Fifth, students' independent work--seatwork and homework-elicits successful and productive practice and teacher feedback. Finally, the instructional group ing system sustains all students' productive engagement and makes the most of teacher expertise. To the extent that any of these dimensions of a lesson is inadequate, the quality of the students' opportunity to learn from it may be compromised. No doubt as new methods permit subtler analyses. other important dimensions will be added to this list. but these form the heart of current. well-substantiated characterizations of good instruction.

Other Approaches to Analyzing Instruction

Each of the findings above draws support from competent research that has been replicated often enough to generate confidence in the results. However, the studies have examined features of a lesson in isolation from the lesson context--a strategy made necessary by the limits of method and the demands of social science, but also one that obscures in critical ways the image of opportunity to learn as teachers and students experience it. As Good (1983, p. 45) reminds us, "The desirability of a particular teacher behavior depends in part upon the teacher's total instructional plan, the content being taught, and the characteristics of individual students."

Other recent studies have approached the task of lesson analysis with frameworks that capture complex clusters of interactions, adding validity to portrayals of the lesson experience, if not yet suggesting an easily quantifiable metric. In the field of research on teaching, a consensus is emerging that some qualities defining opportunity to learn vary by subject matter-the features that create an opportunity for learning math may not have recognizable counterparts in social studies or literature. In the field of curriculum, researchers look at task structure to determine what a lesson is about and how it may stimulate learning. In the field of socio-linguistics, researchers examine discourse patterns to identify the communicative structures that shape learning by framing and labeling "opportunities."

Although topic coverage is one pretty good predictor of student achievement, recent international comparisons suggest that it doesn't explain as much variation in achievement as one might expect. That is, in countries that teach a given topic included on the international test, students usually (but not always) perform better on items covering that topic. However, students in some countries seem able to extrapolate from what they are taught with sufficient skill to succeed on items they have never encountered before at rates even higher than students in countries that cover the topic. This (among other things) indicates that other factors make strong contributions to opportunity to learn and subsequent achievement. Understanding what students have an opportunity to learn requires knowing more than what topics were covered; it requires knowing something about the quality of their learning experience. Among the dimensions of quality that appear to be important are the adequacy of subject matter presentations, the nature of the academic tasks, and the match between teachers' and students' socio-linguistic expectations.

Substantive adequacy. In the past two or three decades, researchers seem to have focused on "generic" features of teaching and learning, searching for evidence of effectiveness across disciplinary boundaries and, often, grade levels. Solid findings in math would spin off replications in science, language arts, and social studies to look for similar results--which were usually found. Until recently, the absence of a solid theoretical model connecting teachers' knowledge and representations of disciplinary content to student achievement constrained research on the features of good teaching that might be unique to a discipline.[4] A series of studies in the late 1980s undertook the task of developing a model. Their records show differences in portrayals of subject matter that arise from differences in teachers' backgrounds and that seem certain to have serious implications for student learning.[5]

The ways teachers present content shape the opportunities students have to learn it; math and science offer especially telling examples. In math, teachers with limited math backgrounds often teach procedures rather than concepts; their students are likely to experience mathematical computation but not mathematical reasoning. They may learn to add or even to find square roots without calculators, but they do not learn why or when to do so. Furthermore, the absence of conceptual learning leads students to use what some call "buggy algorithms"--procedures that happen to work in specific practice exercises with items of very similar types but that are not exactly correct and do not illustrate the application of a mathematical principle in a way that permits transference. Students acquire buggy algorithms from lessons that portray content in flawed ways. Instructors teaching out of field--for instance, biology teachers assigned to physics classes-often restrict the range of students' discussion and explorations, sticking to a relatively simplistic version of the text material. Without a foundation of knowledge on which to draw, they are unable to support and guide students' discoveries and limited to emphasizing rote memory and procedure. If math lessons provide opportunities to learn only computation, then mathematical thinking may be forever beyond reach.

Good instruction in subject areas builds on teachers' substantively well-developed knowledge and skill, and it provides students with opportunities to learn content that meets the challenges of real-world applications-content that transfers. Important aspects of the quality of an opportunity to learn subject matter are the extent to which instruction (Grossman, 1991):

- Presents "the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations-in a word, the ways of formulating the subject that make it comprehensible to others" (from Shulman, 1986, pp. 9-10)

---

[4] Much of this section draws on a brief but comprehensive summary of this early work on teaching in the disciplines written by Pamela Grossman (1991). Grossman was part of the research team led by Lee Shulman at Stanford that conducted an extensive, coordinated series of studies laying the groundwork for understanding dimensions of teaching that are rooted in subject matter.

[5] For example, Grossman, P., (1988); Gudmundsdottir, S (1989); Hashweh, M. (1987); McDiarmid, W., Ball, D., and Anderson, C. (1989).

- Addresses directly the pores of particular difficulty to the learners, taking into account that, at each developmental stage generally and at various points in learning specific subjects, learners are susceptible to predictable misconceptions

- Connects to students' prior knowledge of related topics as well as to upcoming lessons on the same subject

- Emphasizes the key ideas and defining principles of the subject, rather than easy but marginal concepts

Although the body of evidence connecting subject matter portrayal to students' achievement is not yet as compelling or extensive as the best of the process/product research, intuition and experience suggest that such evidence will eventually appear. Two perceptions support the hypothesis about the connection: First, teachers often report having covered material that students show no signs of having learned, and, second, ethnographic studies of teachers" presentations of content show wide variation in the adequacy of their portrayals in dimensions that have direct relevance to students' opportunities to learn. The difference between whether a student learns how to determine the theme of an unfamiliar work of literature instead of only the theme of a particular work, whether a student learns how to analyze the contribution of forests to an ecosystem instead of only the chemical processes of photosynthesis must inhere at least partly in the nature of the portrayal of subject matter in the relevant opportunity to learn. Characterizing that aspect of lesson quality is more difficult than other aspects successfully captured by the methods of process/product research.

Academic tasks. Academic tasks are where the rubber meets the road in teaching and learning

> The tasks on which students work structure to a large extent what information students select from the environment and how they process it. To understand classroom learning thus requires an understanding of children's progressive performance on assigned tasks; to understand the effects of teaching on learning it is necessary to ascertain the extent to which the intellectual demand of assigned work is appropriately matched to children's attainments (Bennett & Desforges, 1988, p. 222).

An "academic task" includes a product, the operations to produce the product, the resources available to support the operations, and the importance of the product in the class accountability system, that is, whether it will be on the test or has significant weight in the final grade (Doyle, 1986).[6] Academic tasks are like molecules of instruction-their subparts have names and distinguishing features, but only

[6] Recent studies on academic tasks usually begin with Doyle's definition and build on the foundation provided by his work.

when assembled do the parts constitute an opportunity to learn. Tasks occur in systems of work that are influenced by environmental features. Like the constructs of process/p roduct research, the construct of task is "generic" although obviously its specific features vary by subject.

The rationale for considering the task as a unit of analysis for understanding opportunity lies partly in failure of simpler and more traditional student activities such as "on-task" reading and writing to correlate strongly and consistently with achievement, according to Doyle (1992). In some situations, such behaviors as students' movement and apparent fidgetiness--conventionally thought to be evidence of distraction--have been found to be reasonably good predictors of learning, because, in the context of the lesson's academic task, they were symptoms of active engagement in problem solving. Students experience the curriculum as a series of tasks that embody "content"--a blend of subject matter and other things to be learned by performing an activity.

Consider, for example, the differences in opportunity to learn presented in two hypothetical (but typical) beginning reading lessons that emphasize strongly and with little digression the sound/letter correspondence for the letter "m" and last about 15 minutes. In both lessons, teachers spend about five minutes explaining the correspondence, using picture cards and objects around the room for cues and calling on many children to discriminate between words (including object names) that begin with the sound "m" and others. In one lesson, students spend the last 10 minutes circling on a page of pictures all the objects whose names begin with m and then coloring the page, which the teacher collects for later correcting. In the second lesson, students pair off, working with the same page of pictures. First, they determine together the names of the objects pictured.[7] Then they take turns deciding whether each name is an "m-word. " The listener either confirms or corrects the speaker's decision and then takes a turn as a speaker. If they finish before time is up, they play a game in which one says two or three words and the other has to determine whether any begin with the sound of the day. During the final minute, the teacher pulls the class together, points to the pictures one at a time and randomly calls on one person from each pair to name the picture and tell whether the name given is an m-word.

The intended product and the resources are the same for both lessons-demonstration of mastery of sound/letter correspondence, a worksheet, and commonplace school supplies. The

---

[7] A common student problem in this kind of exercise is choosing a label other than the one the teacher has in mind when correcting the paper (" cupcake" rather than "muffin," for example). then determining an answer on the basis of the "wrong" label, and being marked "wrong" regardless of whether the beginning sound identified was correct for the label chosen. In this case, both the student and the teacher may end up with misinformation about the student's comprehension of sound/letter correspondence.

operations needed to produce the object were different--in one case, the operations were solo naming and sound identification, followed by a probably longer period of choosing colors and coloring within the lines; and in the other case, operations included discussions about the objects' names, individual determination of whether the name begins with the m sound, giving and receiving critical feedback, and extended practice. In addition, the weight or the product in each accountability system is different. Students in the first group may know from past experience that it will take days to get the colored paper back and that it is one of dozens of little importance completed in a week, while students in the second group know they will have to demonstrate mastery in public immediately following practice. Doyle's observations and analyses of lessons indicate that these differences in task structure create very different opportunities to learn.

The nature of the academic task also explains the effectiveness (or ineffectiveness) of whole-group, small-group, or individual instruction under different conditions. In whole-group instruction, some tasks during a lecture/recitation episode elicit very little operational effort while others elicit quite extensive effort. Consider, for example, the implications of alternative approaches to the same objective:

- Either the teacher calls on one student for the answer to "2 + 6" or asks the whole class, when cued, to show the number of fingers that represents the answer

- Either the ***small*** group has to fill out and turn in, for a group grade, a sheet that asks for the names and demographic features of the capital cities of 6 states in a region or, given the same single worksheet, all members of the group will be expected to be able to name and describe the capitol city of a given state when called on randomly and publicly, to earn game points for their group at the end of the lesson

In each pair, the first task requires the attention of only a single person, and a small group of likely suspects in each class will vie for the chance to be that person while the other students watch--performing no operations, producing no products. They are experiencing, for the most part, an opportunity LO learn classroom "survival skills" rather than math or social studies content. The second task demands that each student perform the operations-at least most of them-and the "product" is viewed publicly. Students may be experiencing an opportunity to learn survival skills. but content remains a central focus, because demonstration of mastery weighs heavy in the accountability formula, at least within the episode.

Doyle (1986) points out that familiar tasks usually run smoothly and with at least the appearance of productivity (more pages get filled in, although more learning may not happen). Conversely, when attempting novel tasks, students ordinarily work slowly and with stops and starts.

Doyle observes that in circumstances where classroom order is under constant assault from the environment, teachers lean on familiar tasks to maintain order. They may break up complex tasks into subparts that lend themselves to routinization, purchasing order at the expense of subject matter rigor. Completing a whole page of 50 computation problems may not be intellectually productive, but it offers more easily sustained engagement than whole-group or small-group discussions about or individual attempts to solve complex applications of algorithms. The tasks designed to elicit difficult or multi-step operations representing useful levels of content mastery may be sufficiently challenging that they render students vulnerable to distraction.

The work of Doyle and others investigating the relationships between academic tasks and student achievement presents compelling support for construing opportunities to learn in terms of academic tasks, because they represent meaningful clusters of the activities and substance that or-her research shows are related to learning.

Socio-linguistic expectations. An opportunity to learn is an instructional event recognized as such by students. However, the event's defining features are culturally marked, and some students come to school without the cultural coding system that marks opportunities in lessons. For example, within the American English language community, subgroups pattern discourse and signal participation options in different ways. In her study of teaching and learning in Appalachia, Heath (1983, p. 280) noted that local children found incomprehensible their standard English-speaking teachers' indirect orders. "Is this where the scissors belong'?" seemed to them a rhetorical question requiring no attention, not a polite way of saying "Put the scissors away. " Others have observed that in some cultural subgroups children are to be seen and not heard, so that even somewhat direct requests for their participation might be ignored on the grounds that "The teacher could not possibly be speaking to me. " In other subgroups, children are given conversational equity as soon as they can command adults' attention on equal grounds, so they jump into discussions as full members rather than aspirants-which other groups find disconcerting.

These orientations toward discourse shape children's perceptions of tasks and opportunities in the classroom. Patterns of interaction that characterize "good" lessons may not be comprehensible as opportunities to learn for some students under some conditions. The teacher may pause expectantly or invite comment with a raised eyebrow or inflection, and misread students' lack of recognition of the opportunity for lack of intelligence or interest.

Furthermore, differences in students' accuracy in "reading" a lesson's demands arise not only from culture but from context variables that indicate the likelihood of some interpretations rather than others. "Put your heads on your desk" may mean "You have been too rowdy and I am angry" in some situations, but "Prepare for guided imagery as a creative writing warmup" in others. In some lessons, teachers will reduce the opportunity for error by framing questions to elicit a rapid series of one-word answers, whereas in other lessons, questions may require longer thought and permit more speculative responses. Recognizing and interpreting the cues that define academic tasks and producing the desired behavior often call for subtle perception and analysis on the part of students. Whether a given set of teacher behaviors, substantive objectives, and material resources provides a student with an opportunity to learn and what it provides an opportunity to learn depend on the match between the communication system and the students' ability to make sense of its parts.

Most members of the broader community, even children who spend some part of their days in ethnic or linguistic enclaves, are able to figure out the discourse rules and expectations of mainstream institutions in the long run with some degree of skill. However, as they encounter new schools, teachers, or subjects, some students experience a degree of disorientation that influences the extent to which the learning opportunities presented make sense to them. Accommodating the differences in their socio-linguistic assumptions may result in opportunities to learn with idiosyncratic characteristics that produce the desired types and extent of student engagement but do not look like conventional and productive opportunities to learn.

The Image of Opportunity: Analyzing a Lesson

Each analytic lens described above captures a different image of the phenomenon called "opportunity to learn" in a lesson. Each posits a different construction of the event, which implies collecting data on different features to generate a characterization of opportunity to learn. The following description of a real classroom event, documented by Gaea Leinhardt[8] during her studies of mathematics teaching and learning, provides "a fairly accurate description of the first 10 minutes of a very good lesson on subtraction with regrouping, as seen through the eyes of a hypothetical student." The excerpt serves to anchor brief analyses of instruction and its assessment that follow, illustrating how some of the analytic frameworks described above produce different images of an instructional event.

> Pretend for a moment that you are 7 years old and you are sitting in a room with 28 other children who are also 7, although one or two are 8. The room is in a school, and it is your homeroom. You spend about 5 hours a day in the room and about 1 1/2 hours a day in other parts of the building. You will be in this room for a total of 180 days during the school year

---

[8] Excerpted from The skill of learning from classroom lessons, by Gaea Leinhardt and Ralph T. Putnam, in American Educational Research Journal, Winter 1987, Vol. 24, No. 4, pp. 557-559.

unless you are sick. Lots of things happen to you in the room, and you are asked to do a lot of things, although most of the time you sit at a small desk and watch and listen.

Today the teacher is walking around to each of the desks and giving children Popsickle sticks, or rather telling them that they may take Popsickle sticks from the tin cans that she is holding. As the teacher comes to your desk, you see that there are loose sticks in the orange can and bundles of sticks with rubber bands around them in the blue and purple can. You like blue and purple better than orange, so you take out four of the Popsickle stick bundles. The teacher then pushes the orange can toward you and says, "Take out some ones, too. " You take out the loose sticks without counting them and simultaneously you glance at the bundled sticks on your desk. As the teacher continues to move around the room distributing Popsickle sticks, you start to think of ice cream and of how many ice cream Popsickles you might have had to eat to get so many sticks. You also start to think about the art project where you used two Popsickle sucks and wrapped bright yarn around them, and then about the social studies project last week when you and two other students built "log" cabins.[9] Your musing is interrupted by the teacher's comment, "You remember last month when we made these bundles of 10?" There is a flash of recognition and you do remember making tens bundles and, furthermore, using them a few weeks ago for adding. You remember that these sticks can be used for counting in math, and you realize these sticks will not be used for art or social studies. At this point, you might also realize that the cans holding the Popsickle sticks were painted colors to be pretty, or you may still be thinking that color is an important feature, as it often is in other classroom activities.

Having finished the distribution of sticks, the teacher walks around the room, asking individual students to give her back some sticks: "Susan, may I have 12 sticks? How many are left?" As this process is repeated several times, it becomes clear that the request for sticks is only half of what the teacher really wants. When it is your turn, you will have to give her the sticks and then quickly count the sticks left on your desk top and count quietly-- you have 49 sticks. That is a big number, and you start to count the sticks in your bundles: 1, 2, 3, and so forth. The teacher interrupts you and tells you to count by tens. You count the bundles as 10, 20, and so forth, and complete the transaction with her, reporting that there are 15 sticks left. At the next desk the teacher asks for 8 sticks. Your neighbor, Baron, looks puzzled. He has two bundles of 10 sticks and 6 loose sticks. There is a long pause; the teacher stands and smiles. The smile gives you the clue that she did not just make a silly mistake that should be politely ignored. This "mistake" was on purpose.

Finally, she says, "Can anyone help Baron out?" One of the bolder girls says, "Take the gum band off. " You think that is kind of a stupid answer because why would you put all those gum bands on if you were supposed to take them off? You think that lending Baron a couple of your sticks would be a better solution. But to your surprise, the teacher seems pleased and, indeed, the gum band is taken off and Baron completes his exchange.

After the interaction with Baron, the teacher quickly collects all the sticks and goes to the blackboard, where she writes the number 42. She picks up a long strip of pink felt and puts

[9] In elementary school, the homeroom teacher usually teaches social studies and often teaches some art.

...on the blue felt board propped up on the chalk shelf, and then she puts on another and another and another. The she picks up two little green felt squares and puts them on the board, too. You realize that there is a connection between the 42 written on the board and the felt strips and squares that have been arranged on the felt board, but you are not sure what it is. At this point, the teacher says, "Now I want more ones. " Instead of just reaching in the basket and taking out more squares (which you have decided must be what she means by "ones"), she takes down one pink strip and says, "Now how many ones should I put up'?" No one answers. You are confused. Why is she taking felt strips *away* if she wants ***more*** ones? You thought the ***squares*** were ones. After a pause, the teacher says, "Ten." She repeats the entire process a couple of times, and then you notice something about the pink strips: They have lines on them-nine lines. Maybe you think of this as nine somethings or maybe they look like ten little squares all attached together..

Process/product analysis. Coding for the amount and pace of instruction, one would note that of the ten minutes in the segment, about nine were invested in a review of previous work, and the main idea arose only after this review. The teacher talks on task while distributing the sticks, but it is not a content-packed segment. Students seem modestly challenged--Baron doesn't know what to do, but Susan does. The teacher is in charge of presenting information, but students are occupied only one at a time in following her lines of reasoning and modeling. An observer could document the extent of student engagement according to overt evidence (such as students' watching the teacher as she moves about the room). Her presentation is direct-it is the classic teacher-centered direct instruction model, in fact, with obvious sequencing from prerequisite skill review to introduction of new concept. Her lesson objective is not immediately clear to the student from whose point of view the story is told. The teacher appears to move the lesson right along. She opts for a kind of patterned questioning strategy, apparently calling on students randomly but starting with those who can answer correctly and then both including other students and making the questions a little harder. She waits patiently and with a smile when a student cannot answer, giving him time to think but evidently trying to keep pressure low. Rather than backing up herself, cognitively-speaking, to develop scaffolding to help Baron answer the hard question, she asks for someone else to "help" Baron; however, she does not have the other person answer for him, but limits the assistance to a suggestion about how to proceed. He then proceeds successfully. She collects the popsickle sticks (thereby eliminating one source of later distraction).

Mathematical analysis. ***The*** teacher is proceeding on a longterm approach that moves from simple to more complex mathematical concepts with the support of manipulatives of gradually increasing abstraction. That is, she is connecting the lesson in subtraction with regrouping to earlier lessons about addition with regrouping and subtraction without regrouping. Furthermore, she begins with objects that the students themselves have grouped into "tens" or left as "ones" (and reminds them of their previous activities) and chat are actual, separable objects. Then she moves on to illustrate with real objects that cannot be taken apart and that have other properties (different colors and shapes, essentially two-dimensional) that shift the perceptual supports toward more symbolic and less concrete metaphors for the mathematics. She emphasizes big ideas, but appears unaware of the extent to which the little participant is unable to distinguish figure from ground at first.

Academic task analysis. Much of this academic task remains unrevealed, because the narrative segment gives us only ten minutes of the two-day, two-period lesson that frames the task. However, at this point, students understand that some elements of the task are taking popsickle sticks out of the cans, performing subtraction with them, and observing the teacher's demonstrations on the board. The task product is unknown at this point, although the teacher calls on students randomly to perform lesson subtasks publicly, so these subtasks have some immediate weight to focus students' attention. The task operations currently involve listening, watching, determining which properties of the manipulatives and which word choices in the teacher's exposition are significant, and occasionally answering questions about the process or the solution. (The answers may function as interim products as well.) So far the resources used include popsickle sticks and felt objects, although it seems likely that others will come into play. It is not yet clear what value the task's final product will have in the course economy. However, inasmuch as this lesson appears to build on some familiar lesson protocols, students may well know already how important their participation in this task is, at least in an immediate way.

Socio-linguistic analysis. The narrative of the student at the center of the lesson provides some information about options for decoding the combination of the teacher's verbal and physical behavior. The student recognizes that this event requires sitting and listening. The teacher's implicit desire that students take bundles and individual sticks has to be made explicit to elicit student compliance. Her naming the objects "tens" and "ones" signals that this is a math lesson, which also implies that attributes such as color of the cans are not important and that sticks will be used for counting rather than building or designing. The teacher's questioning format, repeated several times, finally communicates that the students' work has two parts-counting out the sticks to give away, and counting out the sticks left. In the event, the notion of counting by tens has to be communicated directly, inasmuch as it has not been inferred by the respondent from the several previous examples. The teacher's smile conveys to students that a hard question was asked on purpose-not by mistake as originally suspected-and she permits her silence to stretch as a sign that she really does hope for an answer. The student narrator views the girl who answers the teacher's eventual request for help as "bold" even though no obvious cues suggest the question is rhetorical. (Using the term "gum band" establishes the children as members of a distinctive dialect group, which probably has a host of other implications.) The teacher shifts to a new form of manipulatives when she moves to the front of the room, and signals the change by referring to her need for "more ones. " From her repeated demonstrations of substituting small squares for long rectangles, student infer that she is using a new system to represent the concepts of tens and ones. For most of the ten minutes of this lesson segment, the student at the center has been piecing together information from fairly indirect verbal cues and seems about to arrive at the determination that the lesson is about subtracting with regrouping.

Choosing a framework. Characterizing an opportunity to learn is a matter of phenomenological decisionmaking: What constructions of the event can be supported by valid, reliable, and useful evidence? Three factors influence the decision. First, common sense and the results of social science research suggest that certain constructions of the event are more revealing than others when the goal of analysis is understanding the contribution of formal schooling to achievement. Further, we know that some dimensions actually seem to define "opportunity" within the each construction. For instance, in the process/product view, active teaching capitalizes on the expertise of the teacher and the focused attention Or the students to generate learning. In the view of those who concentrate on the substantive adequacy of a lesson, the way mathematics is construed enables or constrains learning. Both approaches assess the quality of school work, rather than the respective influences of, say, the home, the cultural community, or the neighborhood environment. Second, methodological and technical advances also favor some approaches over others, on the simple grounds of practicality. Researchers have sophisticated strategies for counting some kinds of teaching and learning activities and analyzing their relationships. Other aspects of instruction seem likely to be powerful influences on opportunity, but current research methods do not yet provide reliable quantification. Third, present educational policies focus more attention on some aspects of teaching and learning than on others, arguably of equal importance, but of less social or political interest. Given the limits of capacity to characterize fully any educational event, choosing the dimensions of greatest policy interest is one responsible option.

The three factors create a kind of operational definition of adequacy with respect to portrayals of opportunity to learn. An adequate portrayal includes dimensions that have proved important to learning, that can be assessed with some confidence, and that bear on policy-making. The goal of assessing opportunities to learn at the national level should be to generate as useful and adequate a portrayal as possible, to reduce the possibility that subsequent policy-making might target aspects of marginal significance. The next section of this paper analyzes the approaches taken by several major studies that included assessment of the instructional component of opportunity to learn.

Recent Attempts to Assess Instruction as Part of OTL

Overview of the Targets of Inquiry in Recent Major Studies

A broad view of opportunity to learn. ***This*** series or papers has defined "opportunity to learn" as a combination of curriculum content, instruction, and resources. The nine studies that provided the foundation for work in assessing opportunity also included context variables, such as teacher and student demographics and professional development activities. (Table I in the appendix summarizes the approaches used by the studies.) The studies used several data collection strategies to learn about the dimensions of interest. The RAND/UCLA and RUC teams tested the broadest array of strategies, but, like most of the others, they made extensive use of teacher surveys. The RAND/UCLA group collected artifacts. including teacher logs, text pages, and copies of tests, homework and class assignments. RUC collected fewer types or artifacts, but conducted classroom observations once or twice during the period of data collection to ascertain the accuracy of teacher log entries. RAND/UCLA, RUC, and to a lesser degree TIMSS interviewed school staff members, and the High Schools That Work (HSTW) teams interviewed students.

Using these strategies in their quest to assess opportunity to learn, the studies examined ten aspects of curriculum content:

- Topics covered (select from a list provided)
- Time spent on each topic
- Expected student mastery level (i.e., recall, apply, etc.)
- Students' prior experience with topic (new, review, extend)
- Text/materials coverage of topic (e.g., indicate pages taught)
- Emphasis on topics
- Nature of test items on topics
- Influences on topic coverage
- Teachers" content-related knowledge and skill
- Integration of topics with other subjects (interdisciplinary lessons)

Earlier papers have described studies' findings related to curriculum content at some length. However, curriculum content is in some ways integral to instruction and discussions that follow will sometimes refer back to this area.

Surveys, in some cases backed by interviews, of students and teachers, were the primary strategy for learning about resources (texts, electronic equipment, lab equipment and supplies, and facilities) and context features (e.g., demographics, school organization, and professional development). This paper touches only lightly on resources, which are discussed elsewhere. (Context variables are not the subject of this review of assessing opportunity to learn.)

Researchers used a similarly broad spectrum of approaches to collect data on instruction. The nine dimensions included in their focus were:

- Teaching practices
- Student activities
- Time allocated for content coverage
- Classroom management
- Grading formula
- Homework
- Non-academic time (e.g.. discipline, administration)
- Attitudes of teachers and students
- Planning and preparation time for teachers

In combination with curriculum content, instructional variables have a strong influence on students' opportunity to learn. This section will describe in some detail how studies attempted to capture dimensions of instruction.

Assessing instruction. Studies ask about teaching practices primarily in teacher surveys, with a few corroborating items in student surveys and observation reports. The projects reviewed for this paper addressed teaching practices in three ways: (1) a list of possible practices to be rated according to the degree of emphasis or amount of time spent on each; (2) items about questioning strategies in particular and teacher reactions to student responses in general; and (3) lesson structure items.

requiring the selection and ordering of lesson components from a brief but comprehensive list to create a sketch of a recent typical lesson.

Together, the teaching practices lists from RAND/UCLA, TIMSS,[10] and RUC provide a fairly complete set of general strategies (summarized in Table 1). In the RAND/UCLA studies, the teacher log lists major activities and asks teachers to check those that apply. This checklist is compared with the same teachers' responses to a survey administered to a larger sample. In RUC, the observation reports and teacher logs for the same days provide mutual confirmation. Some items attempt to capture aspects of the lesson structure as a collection of practices; for example, the RUC observation form asks separately about lesson elements that fit together into a model that resembles active teaching when analyzed together in an observation.

The teacher survey in one RAND/UCLA study listed 12 approaches to math instruction, apparently representing reform-oriented practices, such as "I routinely justify the mathematical principles and procedures I use" and "Students are provided frequent opportunities to discover mathematical ideas for themselves." The set also includes "Students are required to memorize and apply rules," possibly to capture the incidence of conventional practices. The response format asks for indications of emphasis (i.e., none, minor, moderate, major).

Several instruments have items that focus on questioning strategies and purposes and teachers' reactions to students' comments. Items developed for TIMSS ask teachers how often (i. e., never, rarely, sometimes, often) they ask questions to cultivate students' mastery of procedures, develop concepts, assess students' understanding, and explore students' potential misconceptions. Other items ask for the frequency with which teachers react in certain ways when a student gives a wrong answer; the reaction options are: correct students in front of the class, follow up with another question that is easier, call on another student for the correct answer, call on several other students to create a set of possible responses for further discussion, give the correct answer, or not correct the student. The RUC observation instrument asks for ratings of teachers' accessibility during seatwork, teachers' unsolicited feedback, and teachers' use of wait time after asking questions.

[10] Items attributed to TIMSS have been drawn from several documents produced in the development of the final instruments; they may not appear on the final instruments.

Table 1

Teaching Practices Items and Responses: RAND/UCLA, RUC, and TIMSS

Teaching Practice	Studies	Data Source	Response Format
Lecture to class	RUC RAND/UCLA • ·	Teacher Survey • • Teacher Log	0, 30, 60, 120, 180+ mins/week? Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
Lead oral recitation/drill	RUC RAND/UCLA	Teacher Survey Teacher Log	0, 30, 60, 120, 180+ mins/week? Check if it applies
Lead whole class discussion	RUC RAND/UCLA	Teacher Survey • •	0, 30, 60, 120, 180+ mins/week? Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Have students work in pairs, teams, or small groups	RUC RAND/UCLA • · TIMSS	Teacher Survey Teacher Log Teacher Survey Teacher Survey	0, 30, 60, 120, 180+ mins/week? Check if it applies Daily, 1-2 wk, 1-2 mo, 1-2 term, never? None, all, some of time in specific lesson
Have students work independency	RUC RAND/UCLA	Teacher Survey • •	0, 30, 60, 120, 180+ mins/week? Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Demonstrate	RUC RAND/UCLA TIMSS	Teacher Survey Teacher Log Student Survey	0, 30, 60, 120, 180+ mins/week? Check if it applies Scale: 1/never - 3 daily
Use manipulatives or audiovisuals to explain a concept	RAND/UCLA • ·	Teacher Survey Teacher Log	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
Have student-led whole group discussions	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Have students work in cooperative groups	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Have students give oral reports	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Discuss career opportunities in subject	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Review or discuss homework	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Have students use: computers/calculators/manipulatives texts other than course textbook math lab activities	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Discuss everyday applications of subject	TIMSS	Student Survey	Scale: 1 never to 3 daily
Have small groups find joint answer	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
Have whole class discuss small group answers	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
During lesson, teacher provides: overview connection to prior learning rationale summary at end	RUC	Observation Form	Scale for each: 1 not at all to 5 comprehensively 1 no link to 5 explicit link 1 no rationale to 5 lofty rationale 1 no summary to 5 good summary

Finally, as a result of analyses of several earlier studies, TIMSS researchers developed an item intended to capture overall lesson structure. The purpose of this item is to characterize the learning experience in less fragmented, more holistic terms than those used in other studies. RAND/UCLA teams had explored strategies for data collection on this point, and t-heir work led to adoption of the present structure. Teachers are asked first to recall the last typical lesson they taught in math. With this lesson in mind, they review the following list of lesson components:

- Review of previous lesson(s)
- Short quiz or test to review previous lesson
- Oral recitation or drill (students responding aloud)
- Review or correction of previous lesson's homework
- Introduction of a topic (class discussion, teacher explanation/demonstration, film, video, use of concrete materials, etc.)
- Development of a topic (class discussion, teacher explanation/demonstration, group problem solving, film, video, etc.)
- Small group activities (with or without teacher
- Students do paper-and-pencil exercises related to topic (not the same as homework)
- Assignment of homework
- Students work on homework in class
- Student laboratory or data collection activity (not a separate laboratory hour) or hands-on session

On blank lines to the right of each item, teachers indicate the order in which the component appeared in the target lesson (if at all) and the number of minutes spent on it. (The total length of the lesson is asked in a previous question.) If an element occurred more than once, it receives a ranking and duration number for each time it occurred.

In lessons, the counterman to teaching practices is student activities, and most studies reviewed for this project ask about student activities in the same kind of detail as they asked about teaching, using mainly teacher and student surveys. In general, the items fall into two categories: those that focus on specific, subject-related activities ("practice computational skills") and those that are more generic (" work in pairs or ***small*** groups"). Table 2 summarizes the content of items related to student activities, taken primarily from studies by RAND/UCLA teams and from work conducted in preparation for TIMSS. All of the TIMSS items did not appear in the final instruments, in part because time constraints ultimately required severe cutting, even of items that had proven to be well-constructed. (RUC has no student questionnaire.)

RUC observation instruments inquired about the nature and extent of student engagement using both narrative reports and rating formats. Items proposed for TIMSS probed for the details of students' interactions with each other in whole-group, small-group, paired, and individual settings.

Most of the studies ask teachers to indicate the length of an instruction period as an indicator of <u>allocated time</u>. Some also ask how many periods are in a week and/or how many instructional days are in a school year to calculate total instructional time available. These questions focus on the issue of time planned rather than on time spent on academic engagement. RAND/UCLA and RUC instruments attempt to determine the character of allocated time by asking about classroom management issues. Various RAND/UCLA items cover "maintaining order/disciplining students" as well as "performing routine administrative tasks." In the RUC observation form, observers are asked to rate teachers' "efficiency in classroom management," "effectiveness in handling discipline problems," and pacing. Teachers also report in daily logs the amount of time they spend on non-instructional activities. and both they and observers independently describe these activities.

RAND/UCLA and TIMSS instruments gather information on <u>grading practices</u>. RAND/UCLA items ask teachers to rate the following elements as "not, somewhat, or very important" in determining grades:

- Achievement relative to the rest of the class
- Absolute level of achievement
- Individual improvement or progress over past performance
- Effort
- Class participation
- Completing homework assignments
- Consistency attending class

Table 2: Student Activities Items and Responses: RAND/UCLA and TIMSS

	Student Activities	Studies	Data Source	Response Format
CONTENT RELATED	Explain the reasoning behind an idea in math	TIMSS	Teacher Survey	Never/almost never; some, most, all lessons
	Represent and analyze relationships using tables, charts, or graphs in math	TIMSS	Teacher Survey	Never/almost never; some, most, all lessons
	Work on problems for which there is no immediate solution in reads	TIMSS	Teacher Survey	Never/ almost never; some, most, all lessons
	Use computers/calculators to solve math exercises or problems	TIMSS RAND/UCLA	Teacher Survey Teacher Log	Never/almost never; some, most, all lessons Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
	Write equations to represent relationships	TIMSS	Teacher Survey	Never/almost never; some, most, all lessons
	Practice computational skills in math	TIMSS	Teacher Survey	Never/ almost never; some, most, all lessons
	Listen to/take notes on whole class instruction by teacher	RAND/UCLA	Teacher Survey Teacher Log	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
	Listen to teacher review homework problems	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Use books other than the textbook	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Work problems at the board	RAND/UCLA	Teacher Survey Teacher Log	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
	Do math/science activities/special projects	RAND/UCLA	Teacher Survey Teacher Log	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
	Work with manipulatives	RAND/UCLA	Teacher Log	Check if it applies
	Present solutions	TIMSS	Student Survey	Scale: 0= NA, 1 never to 3 daily
GENERIC	Copy notes from the board	TIMSS	Student Survey	Scale: O= NA, 1 never to 3 daily
	Complete worksheets individually for practice	TIMSS	Student Survey	Scale: 0= NA, 1 never to 3 daily
	Go on field trips	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Go outside to collect data or observe	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	View films, filmstrips, videotapes	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Review or discuss homework	RAND/UCLA	Teacher Survey Teacher Log	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Check if it applies
	Play instructional games	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Give oral reports	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Design and conduct extended projects	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Report on outside reading	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Take an examination or quiz	TIMSS	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Work individually	RAND/UCLA TIMSS	Tchr & St. Surveys Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Never/almost never; some, most, all lessons
	Participate in class discussions	RAND/UCLA	Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never?
	Work as whole class	TIMSS	Teacher Survey	Never/almost never; some, most, all lessons
	Work in small groups	RAND/UCLA TIMSS	Tchr & St Surveys Teacher Survey	Daily, 1-2 wk, 1-2 mo, 1-2 term, never? Never/ almost never; some, most, all lessons

The final version of the TIMSS teacher survey asks teachers to indicate whether they gave "none, little, quite a bit, or a great deal of weight" to these elements in assessing students' work:

- Standardized tests produced outside the school
- Teacher-made short answer or essay tests that require students to describe or explain their reasoning
- Teacher-made multiple-choice, true-false, and matching tests
- How well students do on homework assignments
- How well students do on projects or practical/laboratory exercises
- Observations of students
- Responses of students in class

It may be that differences in weight given to process variables (e.g., consistent attendance, improvement over past performance) compared with that given to outcome variables reflects differences between the contexts of the studies. Some evidence suggests that standardized tests carry greater weight in instructional decisions in some countries than they carry in the United States.

Homework questions cover three main areas: the length of time it takes an average student to complete assignments, the nature of the assignments, and the instructional uses of the assignments. Most surveys have an item that asks something like, "Approximately how much homework do you typically assign to this class?" and provides a blank line on which respondents are directed to report in minutes. A follow-up question asks how many days per week such assignments are given. RUC and RAND/UCLA instruments probe on the issue of instructional use, using almost identical questions: "How often (never, sometimes, most of the time, or always) do you do the following with homework: Keep a record of who completed it; return it with grade or correction; discuss it in class; and/or include it in computing course grade?"

TIMSS and RAND/UCLA studies gather information on the kinds of homework that are assigned. The composite list of options includes the following:

Completing worksheets/work books	Completing text problem/question sees
Reading text/supplementary material	Writing short assignments
Conducting small investigations	Finding uses of content covered

Preparing oral reports	Keeping a journal
Working on long-term projects	Writing definitions of concepts
Applying concepts to new situations	Solving problems with no obvious answers
Preparing written reports	Extending results found in class
Solving applied problems	Explaining journal articles

The major studies took two approaches to assessing teacher attitudes that might influence teacher behavior in the classroom: (1) presenting a list of position statements to which teachers indicated their levels of agreement and (2) presenting a brief scenario of a particularly distinctive strategy followed by a series of reactions or instructional responses, which teachers ordered into a series or chose the one most like their own. (Some of the TIMSS items were deleted from the final version because of respondent time constraints.) The position statements represented positions at different points on the traditional/reform continuum in some cases. RUC and RAND/UCLA used a similar set of polar statements, asking teachers to mark with an X the position representing their own thoughts on these three continuums:

Item **I:**

A	B	C	D
Teacher is a facilitator	Teacher should ask some leading questions	Teacher should guide and initiate student discovery	Teachers should structure, explain, give practice

Item **II:**

As a teacher, my primary aim is to help students..

A	B
Learn mathematical terms, master computational skills, solve word problems	Achieve a deeper conceptual understanding of mathematics

Item **III:**

In mathematics class, as a teacher I aim for..

A	B
In-depth study of selected topics and issues even if it means sacrificing coverage	Comprehensive coverage, even if it means sacrificing in-depth study

TIMSS asked whether respondents thought boys or girls more likely to (1) be good at science/math, (2) be interested in science/math, (3) like science/math, or (4) be able to do science/math experiments. Items asked further about teachers' views of science and math as disciplines and as areas with particular pedagogical features. RUC observers reported on apparent student attitudes about the math and science classes they visited.

Most studies asked directly-although not extensively--about planning, with a focus on the amount of time available and the nature and extent of collaboration with other teachers.

Effectiveness of Assessment Strategies

Of the three major study teams, only RUC has completed full analysis of the properties of its items and instruments and developed scoring systems that collect related information from different data sources. The RAND/UCLA team expects to complete its analyses within a few months (perhaps by late winter 1995); its focus in current work is specifically on validating data collection strategies. TIMSS has pilot test results from item and instrument development work, but data collection on the major project is still under way. In several cases, TIMSS items represent the most recent incarnation of RAND/UCLA items, adjusted according to analyses in progress; the lesson structure item is one example of this carry-over. However, the TIMSS item pool is limited by the team's focus on international, large-scale data collection; some items that have been validated have not been used because they do not fit into the TIMSS framework. Some approaches that could work in a many-layered national study will not be tested in TIMSS.

RUC developed scales reflecting the extent to which data indicate the nature of teachers' academic demands on students and their use of strategies that engage students in active learning and knowledge construction. (Several other scales were also developed.) To create scales, researchers converted item responses to standard score form to ensure that each item received equal weight. In addition, they calculated coefficients of internal consistency and item intercorrelations for each scale, and screened out some items on the basis of their findings. The scale for teacher demands on students was constructed of responses to six items: (1) agree/disagree that "teachers... push students pretty hard.. ."; (2) amount of homework per week; (3) frequency of recording homework completion; (4) frequency of returning assignments with grades or comments: (5) frequency of using homework in calculating grades; and (6) weight of homework in grades. Four items contributed to the active learning scale: (1) amount of time spent on lecturing, whole-class discussion, and pair/team/small-group work; (2) amount of time students spend listening/taking notes. discussing, writing reports, and doing lab or field work; (3) importance of observing, measuring, ordering, comparing, and classifying in class: and (4) importance of interpreting data, recognizing patterns. designing experiments in this class.

RUC also included some instructional dimensions in its definition of curriculum content, notably the mode of presentation, but although the item was sufficiently reliable, it did not provide information as revealing as had been hoped (according to Porter).

RUC found a high level of correlation between observer and teacher accounts of lessons with respect to items on log forms --primarily content covered and emphasis. This appeared to reflect the results of shared mining-both observers and participating reachers were taught how to apply the terms used in the brief log forms. The RAND/UCLA team, on the other hand, found that teachers and researchers disagreed within and across their groups on many aspects of interpreting artifacts and events. Their view was broader-not limited to the items on a log form.

Both RUC and RAND/UCLA teams reported expecting that social desirability would create a response bias in some of their items-both popular opinion supporting reform initiatives in science and math and important state and local policies seemed influential forces in the professional arena. However, neither team found bias to be a noticeable problem: that is, responses might have been more favorable than an objective observer would suggest, but they were still quite conservative, holding usually to traditional arrangements. RAND/UCLA's focus group discussions with participants at various points in their work indicated that teachers were unapologetic about their instructional decisions, however old-fashioned they seemed to be. Response options for items intended co measure reform orientations sometimes had to be expanded to include sufficient range at the traditional end of the continuum to discriminate usefully among practices. Furthermore, teachers used both reform and traditional practices in various circumstances, so assumptions that practices as a whole could be located at a single point on a continuum had to be reconsidered. Most researchers assume response bias would be a problem if self-report data collection strategies related to assessing opportunity to learn were part of an accountability system (rather than some type of indicator system).

Items on several instruments across studies attempted to get at the use of small groups, but we have no information on how well they succeeded. Assessing this dimension of instruction may be important, given the current popularity of "cooperative learning" strategies. The problem is that small group work in itself is not a predictor of learning. The extent to which the lesson structure engages students is the variable of interest--and small group work only achieves that goal under certain conditions, usually requiring the teacher's careful structuring. Judging from the evolution of items from early to later RAND/UCLA studies, researchers are attempting to find questions that get at this structuring.

RAND/UCLA collected artifacts to provide elaboration of instructional dimensions of opportunity to learn, but has not yet developed a technique for classifying and calibrating what the artifacts reveal and integrating it with the results of other data collection strategies. The team expected ultimately to find an analytic framework that would work on medium-scale data collection (i. e., smaller than national but bigger than case studies). However, at midpoint in analysis the team was recommending that observations and focus groups were needed to make clear the meanings (and precise limitations) of survey responses.

At this stage it appears that information about instruction can be gathered reliably from teacher logs that have a restricted range or topics with low-inference coding systems; observations using the same system do not add much to the data base. However, researchers have had limited success in gathering data that reflects the texture of classroom interactions in what some perceive to be more meaningful chunks using instruments with relatively low respondent demands. The lesson structure item in TIMSS is one attempt to address this challenge in a survey.

Some Comparisons of Actual Characterizations of Instruction with Theoretically Useful Models

What dimensions of the role of instruction in OTL do current approaches document? Overall, *the* most commonly used approaches to assessing instruction as a component of opportunity to learn target variables identified in process/product research. In combination with items assessing curriculum content. the items assessing instruction gather data on the amount and pace of instruction (e.g., pages/topics covered, allocated time, classroom management); the presentation of information (e.g., lecturing, demonstrating, eliciting recitation); questioning strategies (e.g., level of cognitive demand); and reactions to students' responses (e.g., wait time, follow up). They ask about the size of instructional groups (whole class, small group, individual), the amount and kind of homework, and the uses of homework and classwork in grading. Questions about student activities provide evidence about the use of active teaching strategies, the level of cognitive engagement, and the amount of time invested in learning. The most recent form of the item about lesson structure attempts to capture the sequence of events in a lesson, providing a selection of components that can establish the extent to which active teaching takes place.

The studies address substantive adequacy by assessing the content coverage as reported by teachers (and corroborated by texts and/or curriculum guides). They use topic lists and scales that allow rating of the extent to which students are exposed to the governing principles and central knowledge of a discipline. RUC calculates depth of coverage by combining information about topics and strategies used to cover them. TIMSS locates individual class coverage on a national coverage "map" developed on the basis of intensive analysis.

The items about student activities provide some information about the nature or academic tasks, the products themselves [e. g., conduct an experiment, report on outside reading) and/or the operations used to achieve them (e. g.. work problems on the board, analyze relationships using graphs). Several instruments ask about instructional resources, such as computers or lab equipment. Both RUC and TIMSS have several items that examine the relationship between homework or classwork and grading or instructional activities. The weight of task products in the classroom accountability system is an important aspect of the way academic tasks shape opportunities to learn.

Socio-linguistic dimensions of opportunities to learn may become evident in responses to items about questioning strategies and reactions to students' responses, pacing, and classroom management. None of the data collection strategies in the studies reviewed for this project used a socio-linguistic approach.

What dimensions of the role of instruction in OTL remain obscure? Several variables identified in the process/product approach as important influences on learning are not addressed in the major studies that included opportunity to learn. The match between what the student is prepared to learn and what is presented in the lesson is not examined, yet lesson productivity depends in part upon having an appropriate degree of challenge. Furthermore, assessing the quality of a lesson's structure on the basis of its having a good lecture may be problematic. Findings about the importance of teachers' adequate presentation of content have been seen as specifying that the lecture/demonstration component of direct instruction is essential. However, research on teaching and learning is expanding the repertoire of effective strategies-for example, adding whole-language methods, discovery activities, project-based lessons-which entails associated changes in the teacher's role as a mediator of high-quality opportunities. In these lesson formats, forms of teacher support outside of direct instruction may help students avoid misconceptions or misinterpretations. Current studies do not accommodate the new developments. The research on wait time (after questioning and after students' responses) balances the need for reflection and mental rehearsal against the requirements of cognitive engagement; engagement hinges in part on pacing. Knowing that a teacher waited three seconds after asking a question or followed up a nonresponse with a different (easier) question does not provide insight into the accessibility of an opportunity to learn in a particular instance. Documenting the extent of whole-class, small-group. and individualized instruction does not provide insight into levels of engagement and effort, because it does not address the context. In process/product research, the main effects are small (though notable) and the interaction effects are many. This suggests that, in assessing opportunities to learn, characterizing the configurations of various attributes may be as important as capturing evidence of their individual presence.

Well-understood topic lists and coding strategies that report instructional methods for each topic do illuminate to a useful degree two dimensions of substantive adequacy. However, they do not provide information about the quality of content presented. A teacher might have "covered" the topic of base-two number systems using pictures and popsickle sticks without communicating the principles of place value and number theory that make the concept useful. Coverage of the *Civil* War may include mention of several causes without ever raising the deep social and economic questions that made it so incendiary an episode of history. Given the extent of misassignment of secondary school teachers, especially in science and math, and the limits of substantive preparation of elementary school teachers in light of the breadth of their responsibilities, it seems likely that educational productivity is affected by ill-conceived explanations of content. Most schools operate with curriculum guidelines that somewhat constrain content coverage to ensure minimum coverage, and checklists of topics and subtopics would show something like the coverage expected–perhaps "good" coverage in terms of extent. What they do not show is whether the biology teacher covering physics or the bilingual teacher with an emergency credential (based on her BA in accounting) teaching a unit on habitats in children's home language provides examples and explanations that make the content known to students. The adequacy of representations of content seems likely to influence the nature and productivity of an opportunity to learn. but items and instruments in current use do not assess this aspect of adequacy.

In analyzing artifacts and accounts of class sessions from teacher logs, the RAND/UCLA study may get closest to characterizing academic tasks. RUC assembles proximal measures into scales and snapshots to generate images of the work setting in different classrooms. The scale for teacher demands on students explicitly includes the weight of class and homework assignments in grading-an attribute that communicates the value of a task to students and that may be instrumental in eliciting their serious effort and marking an opportunity to learn. On the whole, while some items do gather data on elements that define academic tasks, no analytic strategy specifically addressing that construct has been undertaken.

RUC observation reports and raring forms invite observers to comment on climate, pace, management, questioning, and student and teacher attitudes--all of which have socio-linguistic implications. However, data collection on this dimension of opportunity to learn requires special observation and coding strategies that no study to date has used except on a very small scale.

The contributions of different characterizations of instruction to understanding opportunity to learn. Despite the gain in popularity of more student-centered and interactive approaches to instruction. the strongest evidence about effectiveness comes from the process/product research. Whale developing knowledge of cognitive function indicates the likelihood that new strategies will prove powerful in promoting student achievement, such hard evidence as social science has been able to produce with current techniques comes from studies of discrete and quantifiable teaching methods, classroom conditions, and student behavior. The information may not acid up to a meal or even a side dish, but it provides reliable if incremental sustenance on the path to understanding the contribution of opportunity to achievement.

On the other hand, an increasing body of studies supports the hypothesis that the individual features of "effective instruction" influence learning not in isolation but in configurations that respond to given conditions. For instance, using more wait time and easier questions may promote learning in a geography class conducted in English for English language learners, while less wait time and harder questions may promote learning in a native language math class for the same group engaged in the final review for a test. In addition, the constructs of substantive adequacy and academic task structure seem to add great explanatory value to characterizations of opportunity to learn, without entailing very different approaches to data collection than one might use to identify configurations of the more commonly targeted features of an opportunity to learn.

The Implications of Recent Efforts to Assess Instruction for NCES Data Collection

Effective Large-Scale Strategies

Of the three recent major studies that have attempted large-scale data collection and that included technical evaluation, only RUC has actually finished its work and reported on the properties of its instruments. The RAND/UCLA study team has not yet completed analysis of its most recent study, Validating National Curriculum Indicators. One of its central concerns was the adequacy of instruments, so that will form an important part of its findings. The RAND/UCLA team worked with the TIMSS team on item development, so some items used by TIMSS represent the most up-to-date versions, often informed by yet-unpublished RAND/UCLA and TIMSS pilot test results; this is the case with the lesson structure item. However, some of the TIMSS items reflect an interest in reducing ambiguities for an international respondent group, as well as other compromises demanded by time and differences in academic focus. Such items may work reasonably well in an international study, but not as well as others at the national level, where certain ambiguities do not arise and some compromises need not be made.

Teacher surveys targeting practices in one subject area have been the instruments of choice for most of the recent studies. Where studies have assessed opportunities to learn in more than one area, they use parallel survey forms. adjusted to accommodate differences in subjects. For example, it seems that more science than math classes include labs, whereas more math than science classes require homework regularly. Long term projects appear to be more common in social studies, while research papers may be common in both English and social studies. Using relatively low-inference response formats appears to result in reliable (if not precise) response patterns. Distributing related questions throughout the instrument and clustering them later for analysis generates evidence that inspires greater confidence.

At least in the areas of science and math, the best collection of items and instruments would seem to emerge from comparing the results of the RAND/UCLA studies with RUC as soon as the former results are available. Limits imposed on the RUC approach by emergent technology may have been resolved in the later study.

Related Small-Scale Investigations

Finer-grained studies within the larger scope of RUC and RAND/UCLA work provided important information about the survey data and enhanced the validity of findings. RUC developed a log form that could be completed for one period in five or ten minutes--easily enough that most of the teachers who said they would do so did keep them for the great majority of days in the school year. Whether a year's worth of logs is better than five weeks' worth, as the RAND/UCLA team collected, remains to be seen. It also remains to be seen whether the RUC log form--the more complex of the two and seemingly more useful--includes the right questions. Some questions that might have mapped tidily onto the survey did not, and so a chance for validating measures of depth of coverage, for example, was lost. However, it is not clear what could be asked in a way that maps onto the survey and does not make the log form too long for sustained use.

RUC conducted one or two observations of teachers who were also keeping logs, but the observation data did not seem to figure largely in analysis, except to confirm reliability of logs. RAND/UCLA did not use observations, but did use teacher focus groups to debrief data collection events and learn more about teachers' understanding of items and instruments. Ultimately, dust team reported that observations would have answered a lot of questions (for instance, about discourse patterns) that surveys and focus group discussions did not, and recommended that observations be included. Part of the TIMSS data collection process will be to videotape 100 random hours of instruction in math in a particular research population in each participating country. The videotapes will be coded according to several dimensions and used to inform analysis of national trends. At this point, conducting case studies (modeled on those of RAND/UCLA and RUC) of randomly selected schools and classrooms nested within the survey population seem likely to provide important information about opportunities to learn that will make larger-scale data collection and analysis more productive.

References

Bernet. hi., & Desforges. C. (1988). Matching classroom tasks to students' attainments. The Elementary School Journal, 88 (3).

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching. New York: MacMillan.

Doyle, W. (1986). Content representation in teachers definitions of academic work. Curriculum Studies 18(4), 365-379.

Doyle., W. (1992). Curriculum and pedagogy. In P.W. Jackson (Ed.), Handbook of research on curriculum. Macmillan Publishing.

Good, T., L. (1983). Research on classroom teaching. In L. S. Shulman, & G. Sykes (Ed.), Handbook of teaching and policy. New York, New York: Longman Inc.

Grossman, P. (1991). Mapping the terrain: Knowledge growth in teaching. In H. C. Waxman & H. J. Walberg Effective teaching: Current research. Berkeley, California: McCutchan Publishing Corporation.

Grossman, P., & Richert, A. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. Teacher and Teacher Education, 4

Gudmundsdottir, S. (1989). Knowledge use among experienced teachers: Four case studies of high school teaching. Palo Alto, California: Stanford University.

Hashweh, M. (1978). Effects of subject matter knowledge in teaching biology and physics. Teaching and Teacher Education, 3

Heath, S., B. (1983). Ways with words: Language, life and work in communities and classrooms. Cambridge, Massachusetts: Cambridge University Press.

McDiarmid, W., Ball, D., & Anderson, C. (1989). Why staying one chapter ahead doesn't really work: Subject specific pedagogy. In M. Reynolds (Ed.), Knowledge base for the beginning teacher. Oxford, England: Pergamon Press.

Leinhardt, G., & Putnam, R., T. The Skill of learning from classroom lessons. American Educational Research Journal, 24(4), 557-587.

TABLE 1 - Summary

Recent Approaches to Measuring Curriculum Content, Instruction, and Classroom Resources as Elements of Opportunity to Learn

Subject or Focus of Inquiry	Where are the data found?											
	Artifacts					Text & Materials	Observation Reports	Survey Data		Interview Data		
	Teacher Logs	*Tests*	*Homework*	*Classwork*	*Other (e.g. ...)*			*School Staff*	*Students*	*School Staff*	*Students*	*State, district*
1. *Curriculum Content:** 1 list of topics covered	AI RAND RUC	RAND SRA	RAND	RAND	HSTW (suggested programs) SRA (syllabi) TIMSS	RAND RUC TIMSS	AI RUC	CLAS RAND RUC, SRA TIMSS	CLAS SRA	RUC		RUC
2 time on each topic	AI RAND RUC	RAND	RAND	RAND		RAND	AI RUC	RAND RUC, SRA TIMSS				
expected student mastery level	AI RUC#	RAND		TIMSS				RAND RUC, SRA		TIMSS		
4 st's prior experience with content		RAND	RAND	RAND				RAND, SRA TIMSS			HSTW	
5 text/materials content coverage	AI RAND RUC	RAND	RAND	RAND	SRA	CAM RAND	AI	RAND RUC, SRA TIMSS		RUC		RUC
6 emphasis on topics	AI RAND RUC	RAND	RAND	RAND		RAND	AI	CLAS RAND RUC, SRA TIMSS			HSTW	
7 test items on topic	RUC	RAND SRA						RAND, SRA TIMSS	CLAS			
8 influences on topic coverage								RUC, SRA TIMSS		RAND RUC		RUC
9 tchr's knowledge of content							AI	CLAS RAND RUC, SRA TIMSS	TIMSS	RUC		RUC
10. content integrated w/ other subjects	AI					TIMSS	AI	JHU TIMSS	TIMSS		HSTW	

Subject or Focus of Inquiry	Where are the data found?											
	Artifacts					Text & Materials	Observation Reports	Survey Data		Interview Data		
	Teacher Logs	*Tests*	*Homework*	*Classwork*	*Other (e.g. ...)*			*School Staff*	*Students*	*School Staff*	*Students*	*State, district*
B. Instruction * 1. teaching practices	RAND RUC		RAND	RAND			AI RUC	JHU, RAND RUC, SRA TIMSS		RUC		
2. student activities	AI RAND RUC	RAND	RAND	RAND			AI RUC	HSTW RUC, SRA TIMSS	SRA TIMSS		HSTW	
3. allocated time for content	RUC		RAND	RAND		RAND	AI	RAND RUC TIMSS				
4. classroom management	RAND						AI RUC	TIMSS				
5. grading formula		RAND	RAND	RAND				RAND, SRA TIMSS				
6. homework (nature, frequency)	AI RAND RUC	RAND	RAND				AI	RAND RUC, SRA TIMSS	SRA TIMSS			
7. nonacademic time	RUC						RUC					
8. attitudes of teachers, students							AI RUC	JHU RAND RUC, SRA TIMSS	TIMSS	RUC		
9. prep/planning time (including paperwork)								TIMSS				
*C. Materials/Equipment**: 1. texts	AI RUC							RAND RUC, SRA TIMSS		RUC		
2. electronic equipment	RUC							RAND RUC TIMSS	TIMSS	RUC		
3. lab equipment & supplies								RAND RUC TIMSS				
4. facilities (room, chalkboard)							RUC	TIMSS		RUC		

Subject or Focus of Inquiry	Where are the data found?											
	Artifacts					Text & Materials	Observation Reports	Survey Data		Interview Data		
	Teacher Logs	*Tests*	*Homework*	*Classwork*	*Other (e.g. ...)*			*School Staff*	*Students*	*School Staff*	*Students*	*State, district*
D. Setting/Context*: 1. student demographics							AI	JHU, RAND RUC, SRA TIMSS	SRA TIMSS	RAND RUC		RAND
2. teacher/counselor demographics							AI	JHU, RAND RUC, SRA TIMSS		RAND RUC		
3. administrator demographics								JHU		RUC		RAND RUC
4. student achievement level/tracking	AI	SRA					AI	JHU, RAND RUC, SRA TIMSS	TIMSS	RAND RUC		RAND
5. school organization								JHU, RUC TIMSS		RUC		RUC
6. external influences								JHU, SRA RUC	TIMSS	RAND RUC		RAND RUC
7. professional development								RUC TIMSS		HSTW RUC		
8. course organization/offerings								JHU, RAND SRA		RAND RUC		RUC
9. parent involvement								JHU, SRA TIMSS	TIMSS			RUC
10. school location (surroundings)								TIMSS				
11. report card information								JHU				
12. incentive								JHU	TIMSS			
13. extra curricular activities								JHU	TIMSS			

*KEY

A. *Curriculum content:* What topics are taught? How familiar or unfamiliar are they? What is the emphasis on each? How is each portrayed in tests and other assignments?
B. *Instruction:* What do teachers and students do during classtime? For homework? How does grading weight activities?
C. *Materials/Equipment:* What do students and teachers have available to use and do use in the classroom?
D. *Setting/Context:* What are the factors that affect classroom and institutional climate?
Intended Learner outcomes

AI Academic Instruction for Children in Poverty
CAM Curriculum and Achievement in Math
CLAS California Learning Assessment System
HSTW High Schools That Work
JHU Johns Hopkins University [NELS] Enhancement Survey
RAND RAND/UCLA Validating National Curriculum Indicators
RUC Reform Up Close (also RUC/CPRE, on some instruments)
SRA School Reform Assessment Project (a RAND/UCLA study)
TIMSS Third International Math and Science Study (including the Survey of Math and Science Opportunities--SMSO)

MEASURING CURRICULUM CONTENT: THE STATUS OF RECENT WORK

Mary S. Leighton
John E. Mullens

October 1994

Prepared for the U.S. Department of Education under a subcontract of Contract RN 9306100 between The National Center for Education Statistics and MPR Associates, Inc.

MEASURING CURRICULUM CONTENT: THE STATUS OF RECENT WORK

1. Validating National Curriculum Indicators--RAND/UCLA/CRESST Research Teams

What Is the Purpose of the Projects?

In the Validating National Curriculum Indicators study, a team of researchers from the RAND Corporation and the Center for Research on Evaluation, Standards, and Student Testing at UCLA was funded "to design and field-test a model for collecting benchmark data that can seine as.. .anchors against which the validity of routine data collected through national efforts" can be assessed (Burstein, April 1993, p. 6). The central purpose is to create a benchmarking process that provides detailed information about the context of teachers' and students' responses on national surveys in order to ensure appropriate interpretation. The surveys themselves provide insights about aspects of educational quality, including the extent to which the substance and the routine processes of schooling are changing in this era of reform. The team's work is the latest in an evolving series of investigations pursuant to recommendations in the 1987 RAND report on indicator systems (Shavelson et al.). That report called for adjunct studies of (1) substance to "go significantly beyond the findings of the indicator system in their depth of analysis and power of explanation" (p. 49) and (2) methods to enhance the technical capacity of an indicator system to measure such variables as teacher quality, student achievement, and enacted curriculum. Data collection for the current study, Validating Indicators, funded by NSF and NCES, was carried out in conjunction with the 1992 NELS Second Follow-up.

What Do the Projects Measure: Subjects? Grade Levels?

Both the earlier School Reform Assessment (SRA) (McDonnell et al., 1990) project and the indicators study conducted in conjunction with NELS 1992 focused on high school math. SRA also examined history and government, while the indicators study included science, dividing both math and science roughly into upper and lower division courses. Math I surveyed teachers of algebra I, geometry, and lower-level courses leading to them, and Math II surveyed teachers of algebra 2, trigonometry, and calculus. One science survey targeted teachers of biology and the other, teachers of physics or physical science. The final sample of the Validating Indicators study included about 70 math teachers responsible for 20 distinguishably different courses in math (general math, pre-algebra. Math A & B. Interactive Math Project, algebra 1 and 2, honors algebra 2, algebra 3/trigonometry, geometry, honors geometry, math analysis and pre-calculus, calculus, and AP calculus) and 18 science teachers, responsible for biology, advanced biology, AP biology, physical science, physics, and/or honors physics. Within these areas, data collection focused on three aspects of practice:

1. Content coverage, including emphasis (how important is a topic? how often is it taught?) and student performance expectations (for example, are students expected to recall, apply, and/or evaluate lesson content?)

2. Instructional practices and conditions, including teaching strategies (amount of classtime the teacher spends lecturing, monitoring groupwork, supervising labs, etc.); time allocations for regular class sessions and lab sessions; homework (how often does teacher record, grade, or discuss? what kinds are assigned? how much time do assignments take to complete?); and assessment/grading issues (how important are specific factors in grading? how are tests constructed?)

3. Goals, objectives, and teacher beliefs (what do teachers try to convey about processes associated with the discipline of study? how do they view their role--as informer, leader, or facilitator?)

How Do the Projects Measure Curriculum Content?

On what previous and concurrent work do the projects build? Members of this research team have been involved in developing, analyzing, and reporting on the Second International Math Study (SIMS), the Third International Math and Science Study (TIMSS), the School Reform Assessment project, several studies conducted in conjunction with staff of the Center for Research on Education Standards and Student Testing (CRESST) for use with the California Learning Assessment System, and the State Collaborative on Assessment of Student Standards (SCASS), among other projects. The items and instruments used in the indicators study were created and piloted collaboratively with researchers from these projects.

What items and other instruments have been developed? To generate information that linked survey responses as closely as possible to course experiences, researchers used a variety of collection strategies to gather data from one class section of a course. They interviewed principals, department heads, and counselors and examined curriculum guides and transcripts to sketch a broad picture of the context for the detailed, class-level data set. Formal data collection for the course section included:

- a brief initial teacher survey (ITS)
- collection of major assignments for the entire semester
- collection of all assignments for five selected weeks
- a daily teacher log for the same five selected weeks
- an enhanced end-of-semester teacher survey

In addition, researchers conducted focus group discussions with teacher participants before and after formal data collection to learn more about teachers' perceptions of instruments and their interpretation of emerging findings. Copies of the instruments are included in Appendix A.

The relationships among various types of data collection are displayed in Table 1 (adapted from Table 10, Burstein et al., 1993). Note that the table is organized according to categories established in the final survey, which represents the team's operating conceptual framework, one similar to but more elaborate than the NELS framework evident in the initial survey. The discussion that follows the table first presents the data collection strategies in the order teachers experienced them, from initial survey through artifact collection to final survey, and then summarizes preliminary findings about the usefulness of various question types and response formats.

Table I: Validating National Curriculum Indicators Study: Comparison of Artifacts and Selected Teacher Survey Data

Focus		Enhanced, End-of-Semester Teacher Survey Items Math Forms 1 & 2 [Science Forms 3 & 4]	Initial Teacher Survey (ITS) Items Taken from NELS: 1992	Artifacts (✓ = Direct or indirect confirmation of survey responses)					Tex
				Logs	Exams	Homework	Classwork	Other	
Content Coverage, Emphasis, and Performance Expectations	Coverage and Emphasis	7 [7]. Name text 8 [8]. List chapters covered during target semester 9 [9]. List other chapters that have been/will be covered this year. 10-11 [10-11]. For each of about 20 or more subject-specific topics [listed in related clusters], indicate (a) whether it was introduced or reviewed, and, (b) if taught, for how many periods, by circling appropriate responses. 22 [23]. Indicate how frequently T uses expendables and equipment (e.g., graph paper, calculators, computers, oscilloscope) by circling daily, weekly, etc. [24]. Indicate how frequently material or equipment is used in certain ways (e.g., a computer for data acquisition) by circling daily, weekly, etc. 23 [25]. Indicate course emphasis on each of 20 or more subject-specific process objectives (original 8 to 10 plus others) by circling none, minor, moderate, or major.	14. Indicate course emphasis on each of 8 to 10 subject-specific process objectives--e.g., understanding proofs, memorizing rules, solving equations--by circling none, minor, moderate, or major. 15. For 9 to 11 subject-specific topics taught this year, indicate the extent to which each was introduced or reviewed by circling previously taught, reviewed, introduced, will introduce, or beyond course scope.	✓	✓	✓	✓	✓	
	Performance Expectations	12. For each of about 10 to 12 subject-specific concepts, indicate the type of student understanding expected by circling recognition, recollection, application, explanation, or not relevant to course.			✓	✓	✓		
Teacher Goal and Role Orientation		25 [27]. Indicate T's role orientation on a 4-point scale, from 1 (facilitating) to 4 (didactic). 26 [28]. Indicate T's goal orientation on 2 continuums: (a) skill mastery vs. conceptual understanding (b) depth vs. breadth of coverage		✓	✓	✓	✓		

Focus		Enhanced Survey Items--Forms 1 & 2 [3 & 4]	Initial Survey (NELS)	Logs	Exams	Homework	Classwork	Other	Te
Instructional Practices and Conditions	Strategies	13 [12]. Indicate how often T uses each of 20 generic strategies (e.g., recitation, homework review, small and whole group discussion) by circling daily, weekly, once/twice a month, etc. 14 [15]. Same as 11, ITS [14]. Indicate how often T uses each of 12 strategies in lab sessions - e.g. T explains, Ss follow text instructions or do lab work in pairs or individually--by circling never, sometimes, mostly, or always. 24 [12 and 26]. Indicate emphasis on 12 or more subject-specific teaching strategies by circling none, minor, etc. [Response options on #12, Forms 3 and 4, are daily, weekly, etc.]	11. Indicate percent of classtime/week spent on each of 7 activities (whole group, small group, or individual work; classroom management; tests; administrative tasks; and labs) by circling none, <10%, 10-24%, 25-49%, 50-74%, 75-100%. 12. Indicate how often T uses specified lesson presentation or activity formats (lecture, use computers or AV materials, teacher- or student-led whole group discussion, recitation, cooperative groups, individual written work, oral reports) by circling daily, almost daily, 1-2 times/week, 1-2 times/month, never/rarely	✓	✓		✓		
	Allocated Class Time *Class*		9. Record the minutes/week class meets (not counting lab sessions)	✓					
	Lab	[13]. Indicate frequency of lab sessions by circling daily, weekly, etc	10. Record the minutes/week spent in lab sessions	✓					
	Homework *Time Demand*	19 [20]. Same as 7 in ITS	7. Record the number of minutes/day homework assignments are assumed to occupy						
	Instructional Use	20 [21]. Same as 8 in ITS	8. Indicate frequency with which T records, grades, or discusses homework by circling never, sometimes, mostly, or always	✓		✓			
	Frequency & Type	21 [22]. For each of 13 types of homework assignments, indicate how often they are given by circling daily, weekly, monthly, etc.							
	Assessment	15 [16]. Circle percent use of each item response format on quizzes & exams--multiple choice, short answer, essay, open-ended, or other. 16 [17]. Circle percent of use of each item cognitive demand type on tests--recall, apply, describe, transfer, analyze, etc. 17 [18]. Circle percent of use of each item type on tests--minor variations on familiar problems, more than one right answer or approach, multi-step, and table- or graph-based. 18 [19]. Describe grade distribution for students in target class.	13. Indicate importance of each of 7 factors in grading--class curve, absolute level, progress, effort, participation, homework, and attendance--by circling not, somewhat, or very		✓	✓	✓		

The Initial Teacher Survey (ITS)

The Initial Teacher Survey is composed of NELS items. The survey asks for information about the class for which the reacher is reporting; the goals, content, and instructional activities of the course; and the teacher's background. Teachers respond with reference to a single class section, anchoring their recollections to a specific context to improve reliability. One benefit of asking these questions briefly before the rest of the data collection takes place–for instance, as part of the regular NELS survey-is to provide a point of comparison that indicates whether subsequent responses are contaminated by teachers' other reporting activities, which may exaggerate the perception of certain features of instruction beyond the extent warranted by actual occurrence. This survey has about 25 short-answer questions, about half of which focus on aspects of curriculum and instruction.

Two questions address curriculum content:

> Question 14 asks respondents to indicate course emphasis on each of 8 to 10 subject-specific process objectives-e. g., understanding proofs, memorizing rules, solving equations-b y circling the appropriate answer: none, minor, moderate, or major.
>
> Question 15 asks respondents to indicate for 9 to 11 subject-specific topics taught this year the extent to which each was introduced or reviewed by circling the appropriate answer: previously taught, reviewed, introduced, will introduce, or beyond course scope.

Artifacts

To serve as the benchmarks to validate the survey responses and provide a richer data base for explaining the nature and content of instruction, teachers collected major assignments for the whole semester and, for five selected weeks, collected all assignments and kept a daily log.

Major assignments. For the entire semester of the study, participating teachers collected copies of examinations, papers more than three pages in length, and projects. Researchers also obtained copies of relevant textbooks for course content analysis. These materials were coded and blocked into daily lessons using a system described below.

All assignments. For five weeks altogether-one week during the first five weeks of class, then three consecutive weeks (usually semester weeks 9 through 11), and finally one week near the end of the semester--teachers collected all materials used regularly in instruction: homework, classwork, lab reports, other reports and papers. quizzes, exams, projects, and any other written work, in addition to making a complete record of textbook coverage. Each separate item was tagged with a short pre-printed label on which the teacher checked options indicating the assignment's purpose, its connection to ocher classwork, and the setting in which it was to be completed (i.e., individually or in a group, in or outside of class).

Teacher's daily log. During weeks when teachers collected all assignments, they also tilled in a one-page daily log form. First, teachers recorded "content covered in this class period" on three blank lines. Second, in a list of "modes of instruction" (e. g., lecture, demonstrate exercise, work with individuals or small groups), they checked all that were used during that period. Third, in a list of student activities (e. g., listen and take notes, use calculators, discuss, conduct lab, write report), they checked all that students used that period. Finally, they were provided space to add further comments, if they chose.

Enhanced End-of-Semester Survey

At the end of the semester, teachers completed a survey that included the original questions of the ITS plus others soliciting more detail. For example, as noted above, ITS item 14 asks for emphasis on just 8 to 10 subject-related process objectives and item 15 asks about coverage of 10 to 12 topics. In contrast, the enhanced surveys ask for this information in three or four much longer items, in addition to items asking about text coverage. Enhanced math and science survey items 10 and 11 ask about coverage of 25 to 30 topics, in terms of both students' prior knowledge and the number of periods taught. Enhanced math survey item 23 [25 on the science surveys] asks for indications of emphasis 20 or more subject-related process objectives (the original 8 to 10 plus others). Enhanced math survey item 22 [23 and 24 on the science surveys] asks about the role of specialized equipment in the lessons, a role that may bear on the substantive content. The broader array of prompts and responses on the enhanced surveys gives a far more detailed account of course goals with respect to discipline-related concepts and processes than did the generic prompts and responses on the initial survey.

In addition, enhanced math survey item 12 provides a separate list of 12 related topics for which teachers indicate the level of cognitive engagement and mastery expected of students with respect to each item; for example, "The Pythagorean Theorem-[Students are expected to] (1) Recognize; (2) When given, apply correctly; (3) Know when and how to apply; (4) Apply and explain; or (5) The topic is beyond scope of class." The purpose of this item is show the depth of coverage in art area included in most versions of the targeted math courses. The surveys also ask for information about the text in items 7, 8, and 9.

In both math and science, the original list of topics for each subject and level came from a careful analysis of those most commonly taught, and the subsequent list of subtopics was similarly developed with a view to capturing the broadest sample of responses, even though the subtopics themselves were quite narrow. For instance, under the major topic "environment/ecology," the subtopics were "population and environment, technology and societal issues..., biomes and ecosystems, heredity, and habitats and niches." The major topic is covered in most biology courses; the subtopics, while not usually constituting a major portion of such courses, very often appear in a biology curriculum. Developing a comprehensive but brief list of common topics supplemented by smaller, narrowly-focused list of subtopics balances the needs for minimizing participant response burden and optimizing the generalizability of the data obtained.

What Are the RAND/UCLA/CRESST Team's Preliminary Findings?

The research team is still analyzing data, untangling effects and associations, and developing interpretation schemes that are robust and illuminating. They have already gained some useful new insights about the productivity of various response options, topic list prompts, and artifact collections. but they anticipate that even more information will emerge in the final months of the project.

Response formats for survey items. Preliminary analyses suggest that asking teachers to indicate emphasis by choosing "none, minor, moderate, or major" for each topic or process objective in a list does not elicit useful, unambiguous information about either teachers' behavior or students' experiences. Some teachers expressed conviction that particular objectives were "very important" or received "major" emphasis even though other evidence showed that they spent little time on the objectives. Because time is one of the most precious resources in a course, time invested constitutes a highly reliable indicator of the value of a curriculum component. In addition, artifacts from classes reported on surveys and follow-up conversations with teachers revealed agreement only about the "none" and "major" options, in general. Responses in the mid-range were not reliably defined in practice.

Response options in the enhanced survey sometimes substituted time units for more subjective emphasis metrics; instead of "none, minor, moderate, or major," teachers circled the frequency with which a topic or strategy appeared: almost daily, once or twice a week, monthly, once or twice a semester, or never. An alternative time-bound response option was "percent of time spent in a typical week" on various topics or processes Results on the effectiveness of the period metric are not yet clear, but they appear to map more accurately onto less inferential indication of emphasis and importance. In addition, they permit differential interpretation: "Once or twice a week" may be a high frequency for using cooperative learning, but a low frequency for lecturing. "Once or twice a semester" may be a high frequency for assigning projects that last two weeks, a moderate frequency for using highly technical equipment, and a low frequency for testing.

In items asking whether topics were reviewed, newly introduced, or not covered in a given course, some suggest that a sixth option--"Not in the curriculum'' --would provide information usefully different from but now folded into the fifth option ("Beyond the scope... "), which implies that a topic is in the overall curriculum, but not in the course to which the survey responses apply. Unfortunately, experience so far indicates that many teachers know little about what is taught in other courses. and therefore their assertions about what is "not in the curriculum" are little better than guesses.

Topic list prompts on surveys. ***The*** short lists of topics in the initial surveys appear to have a few items that teachers interpret in various ways. For example, some teachers see "Patterns and Functions" as distinct topics and others see them as closely related, a difference in understanding that results in different responses for what evidence suggests are very similar sets of lessons. More agreement may be elicited by an elaborated topic list, such as the ones in the enhanced surveys. Topic and process objective list prompts must be sufficiently inclusive to capture reasonably well the behaviors of both traditional and reform-oriented courses without being too long and time-consuming or so plainly "correct" that they introduce social desirability as a contaminant. For example, the list of instructional practices in Form 1 of the enhanced math survey includes old favorites such as "lecture, " "teacher-led group discussions, " test administration, drill and practice in computation, and working exercises at the board and newly-recommended practices such as "student-led whole group discussions, " small group work on common problems, long-term projects, using computers, and writing about math discoveries.

Artifacts. Sorting, coding, and determiningg the weights of each element of artifact collections posed a challenge to researchers. Development of practical, valid, and reliable groundrules for translating artifacts into data relevant to the survey items is still under way. With respect to content coverage, researchers used coding categories drawn from the initial survey--the survey prompts became labels. Because other evidence suggested that secondary teachers find it easy to report coverage in terms of a class period, that was selected as the metric for content quantity. As often as possible, assignments, log entries, textbook exercises, and other indicators of content were blocked into daily lessons (Guiton et al., 1994, p.13). The artifacts furnish a valuable source of explanation for survey responses.

Coding for content of artifacts proved more problematic than expected, in part because of a lack of agreement among coders (who were expert secondary math teachers themselves) about the meaning of some terms associated with curriculum reform. For example, teachers had trouble agreeing on examples and nonexamples of concepts such as "mathematical modeling," "proportional reasoning," "estimation," and "tables and charts," and sometimes compensated by defining them too narrowly, other times too broadly, which resulted in under- or over-reporting their frequency in lessons. Coding for content emphasis was also a matter of dispute. By examining the relevant text pages and artifacts, researchers usually agreed on what was the major topic of a lesson, but seldom on the presence and rank order of the secondary emphases, which varied in nature and extent. Thus, they could determine how often a topic was the major emphasis in a lesson--one critical indicator of its importance in a course--but could not calculate the cumulative importance of topics occurring to different extents as secondary emphases, although the latter assessment was also necessary to generate an accurate portrayal of course content.

Such differences in the ways well-educated coders and teachers defined the terms characterizing "traditional" and "reform" approaches to content led to some of the same participants' being rated as "highly traditional" by coders, when survey responses and focus group comments provided evidence that the same individuals might also be seen as "highly innovative"--initially conceived as opposite endpoints on what was envisioned as a curricular continuum. Coders used narrower definitions than others. Focus group discussions with teacher participants revealed that even knowledgeable professionals are still evolving workable definitions of the terms of their arts. Furthermore, the same teachers often <u>are</u> both traditional and reform-minded, at different points in the curriculum or under different teaching conditions.

Distinctions teachers make about the nature of their curriculum and instruction that appear only in the enhanced surveys and artifact collections constitute a response subtext that significantly alters the apparent meaning of their ITS answers, at least in some cases. On one hand, using well-known conventional language enhances survey clarity but reduces its capacity to capture shifts in practice that might occur as a result of reform initiatives. On the other hand, using reform language enhances the survey's sensitivity to reform shifts but seems to increase ambiguity.[1]

---

[1] This ambiguity was resolved to some extent by the Reform Up Close team, which generated definitions for some terms and then taught the teacher participants how to apply those definitions to reporting their practice. However, that approach was not used by the RAND/UCLA group, which gave higher priority to non-obtrusiveness and manageable response burden. The RAND/UCLA team judged respondent training sessions to be impractical in the context of a national data collection effort, where keeping subject response burden to a minimum is very important.

Ongoing disagreements among teachers, coders, and other researchers are one illustration of the value of in-depth data collection. Teachers in focus groups confirmed that the differences among data analysts are the same as their own differences in interpretation: Some of the language of reform (and indeed some of the language of "traditional" practice) remains unclear in its translation to daily teaching. The research team anticipates that when analyses of artifacts and survey data are complete, a template will emerge that indicates how much of each kind of artifact is necessary for effective triangulation and useful elaboration and what coding and weighting strategies produce the most relevant, valid, and reliable data set. Current analyses are focused on determining what constitutes the critical mass of artifacts and what procedures yield the quickest and most reliable triangulation of other responses.

Despite the puzzles that remain to be solved in the final stages of analysis, some early findings are already evident (McDonnell, 1994). First, enhanced surveys and artifact collection in selected sites do contribute materially to understanding the responses of a broader sampling of teachers' on the NELS survey items. Second, while intrusive data collection could have made participants aware of the particular interests of the research project, so far there is little indication that data collection activities or social desirability contaminate responses. Changes between initial and final survey administrations seem to reflect only the influence of experience-describing what has been done is easier to do accurately than predicting what will be done. (This finding is also backed by the experience in Reform Up Close.) Third, response formats that offer "major, moderate, minor, no" or "very, somewhat, little" choices threaten validity and reliability. Fourth, lack of common understanding of the meaning of some terms related to curriculum and instruction generates inaccurate or misleading responses. Fifth, adding observation to the data collection is probably necessary to improve understanding of aspects of lessons (for example, student/teacher discourse patterns) that are difficult to capture in self-reports, but that have important implications for learning. A teacher log form that gets at lesson structure more accurately may provide better information about content emphasis, among other things. Furthermore, evidence developed so far suggests that "findings about the mathematics curriculum generally apply to physical science courses but not to those in the life sciences .... or other subjects such as social sciences" (McDonnell, 1994, p. 1). However, further analyses may reveal broader transfer. The science classes represented a small sample in this study and additional examination of the data is required to understand their implications.

How Might This Work Inform NCES Data Collection?

Within a year, perhaps as soon as six months, the team expects to provide information precisely aimed at informing NCES data collection on opportunities to learn in secondary science and math. At the moment, the team (McDonnell, 1994) offers three recommendations:

- Use the findings of studies such as RAND/UCLA's to revise items and otherwise improve national surveys
- Conduct small, focused, indepth studies that can inform interpretation of survey results
- Develop a plan for ongoing validation studies that target areas of interest, such as lower-level courses or new reform initiatives

This study involved almost 90 teachers in activities that occupied only about five hours over a semester (slightly less than 2 hours to complete both surveys, about 5 minutes a day to complete daily log on 25 school days, and perhaps an hour or two over the semester to collect a portfolio of major and other assignments and tag them appropriately). Each teacher received a stipend of $175 and each participating school received $1,000 to compensate for time spent gathering and copying transcripts and artifact data.

References

Burstein, L. (1993, April). Validating national curriculum indicators: A conceptual overview of the RAND/CRESST NSF Project. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Burstein, L., Guiton, G., Mirocha, J. and McDonnell, L., Ormseth, T., van Winkle, J. (1993, April). Validating national curriculum indicators: Student work samples as benchmarks of learning opportunities and instructional practices. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Guiton, G. & Burstein, L. (1993, April). Indicators of curriculum and instruction. Draft of a paper prepared for the mini-conference on educational indicators at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Guiton, G., Mirocha, J., van Winkle, J., Ormseth, T., and Burstein, L., McDonnell, L. (1994, April). Validating national curriculum indicators: Instrumentation, data collection, and coding. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

McDonnell, L. (1994, April). Validating national curriculum indicators: (Potential) implications for future curriculum indicator design. Draft of a paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

McDonnell, L., Burstein, L., Ormseth, T., Catterell, J., & Moody, D. (1990). Discovering what schools really teach: Designing improved coursework indicators. Santa Monica, CA: RAND Corporation.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N., and Picus, L. (1987). Indicator systems for monitoring mathematics and science education. Santa Monica, CA.: RAND.

Validating National Curriculum Indicators: Instruments and Other Documents

- The Collection of Instructional Materials (Teacher directions)
- Initial Teacher Survey
- Mathematics Teacher Surveys I & II (Survey Forms 1 & 2)
- Science Teacher Survey (Form 3: Biology and Form 4: Physics and Physical Sciences)

2. Reform Up Close (RUC)--University of Wisconsin Center for Education Research/Consortium for Policy Research in Education Team

What Was the Purpose of the Project?

The National Science Foundation funded this study to document the relationships between (1) reform-related state and local policies and practices regarding high school science and math teaching and (2) the enacted curriculum, that is, the content teachers attempted to cover and students experienced. The research team listed five purposes (Porter et al., 1993):

- To determine whether increased course participation stimulated by higher graduation requirements resulted in "watered down" course content
- To characterize state and local policy activities that apply to secondary science and math
- To create a baseline description of practice at the onset of reform activity in the late 1980s
- To learn how curriculum policy influenced classroom practice
- To improve measurement of opportunity to learn

To find answers to the study's questions, RUC researchers interviewed 18 representatives of state education agencies, 74 district-level administrators, and 76 school-level administrators. However, documenting effects at the classroom level required improving existing methods and developing new methods of capturing complex instructional events and comparing them across sites. To this end, researchers administered surveys to math and science teachers (obtaining useable, complete surveys from 312, about 75 percent of them), interviewed 81 teachers of target courses, conducted a total of 116 observations of 75 of those teachers, and collected daily logs from 62 teachers. Interpreting the data generated by different methods involved what amounted to validation studies: in a small subset of a larger sample (about 20 percent of all cases), researchers compared the results of fine-grained, multidimensional, labor-intensive data collection strategies with the results of less labor-intensive surveys targeting the same events. Using these comparisons, the research team was able to determine the extent to which survey item responses produced characterizations of classroom events similar to the characterizations produced by closer scrutiny, thereby making better sense of the survey responses of the whole sample. This effort to represent "opportunity to learn" in the classroom was an important part of the larger study but not the primary focus.

What Did the Project Measure: Subjects? Grade Levels?

RUC's focus was on secondary science and math. In each of 18 schools serving grades 9-12, located in 12 districts in six states, RUC studied four class sections, two in science and two in math courses that had experienced substantial enrollment gains or maintained high enrollment following adoption of more demanding state graduation standards. The team chose to work in Arizona, California, Florida, Missouri, Pennsylvania, and South Carolina, states that varied in interesting ways with respect to their approaches to the reform of math and science K-12 education. For instance, some states aimed to improve basic education and others to stimulate more advanced learning in the content areas. Some provided ground-breaking leadership in program development, such as state-of-the-art curriculum frameworks, with few mandates regarding courses, while others left program leadership to local forces but enacted more course mandates or high-stakes assessments as accountability measures.

Within each state, two schools in a large urban district and one school in a smaller, suburban or rural district were recruited. To get the clearest picture of the effect of reform on students most at risk of school failure, the team targeted schools with high proportions of minority and low-income students, below average achievement, and stable populations. Among the faculty for each targeted math and science course, the team recruited mid-career teacher volunteers who were typical rather than outstanding in their levels of professional competence and who taught the greatest number of sections of the target course. Teachers for 72 target courses were selected; in this sample, 62 teachers ultimately participated in the full range of data collection activities, including log-keeping and outside observations.

RUC used ERIC course classifications, based on curriculum content, to label courses in the study. (The actual courses had different titles in many cases.) The final sample of target courses included the following kinds and numbers of course sections:

Math Course	Sections	Science Course	Sections
Basic Math	8	General Science	2
Prealgebra	5	Physical Science	8
Algebra 1	11	Earth science	3
Algebra 2	4	Life Science	2
Geometry	3	Biology	12
Trig/Precalc	1	Chemistry	2
		Ecology	1
Total	32		30

How Did the Project Measure Curriculum Content?

On what previous work did the project build? Following the discussions of the adequacy of U.S. educational standards surrounding publication of A Nation at Risk in 1983, educational institutions raised requirements for coursetaking, especially in science and math. Some educators and policymakers expressed concern that higher requirements would drive out students, especially minority students who had achieved poorly under the former, lower standards. Instead, rates of graduation and enrollment in low and intermediate level courses remained stable or increased slightly Studies (cited in Porter *et al.,* 1993) confirmed the occurrence of increases in coursetaking in science and math (e.g., Clune and White, 1992) and positive correlations between such coursetaking and achievement (e. g., Schmidt, 1983a, 1983b; Sebring, 1987; and Walberg and Shanahan, 1983), controlling for initial differences in students' ability and resources. A second concern was that a big influx of low-achieving students would cause a watering-down of the curriculum.

The RUC research team members had conducted and/or participated on advisory panels for earlier studies in related areas. Most notably, they extended the methods of the content determinants research done at the Institute for Research on Teaching at Michigan State University, which had developed a three-dimensional taxonomy to characterize elementary mathematics instruction. The conceptual foundation of that taxonomy wedded subject matter (e. g., whole numbers, fractions) to operation (e.g., add, subtract) and student performance expectation (e.g., understand, apply). Within this taxonomy, for example, one topic might be: "applying procedures for adding fractions" and another, "understanding division of decimals. " The purpose of this amalgam approach is to account for the interplay among general instructional intent, the nature of materials and the operation to be performed, presentation, and purpose that shapes a lesson's substance. What (if anything) students learn in a lesson nominally about adding fractions is materially influenced by certain features of their instructional experience[1], as well as by how hard they try. Adapting this approach for topics at the secondary level and incorporating insights gained in other studies about how its usefulness might be improved, the RUC team developed a four-dimensional taxonomy (described below) with a two-part content code (topic, subtopic), a code for lesson presentation format, and a code for student performance expectation.

What items and instruments have been developed? The scope of data collection for the whole study included interviews, surveys, observations, teacher logs, and minimal artifact collection (for example, teacher-made tests). Researchers interviewed state, district, and school-based administrators to learn about the context of classroom instruction: the policy climate, regulations, supervisory practices, and accountability mechanisms (if any) connecting policy to practice. They surveyed all math and science teachers in participating schools (including teachers of target courses). Finally, they interviewed four teachers of target courses in each school and had them keep daily logs, supplemented by a pre-log survey about specific context issues and a weekly questionnaire. Observers visited the target classes once or twice during the study and filled out daily logs, completed summary forms covering various other aspects of instruction, and wrote narrative reports shaped by a set of common questions. Information most directly related to opportunity to learn came from responses to math and science teacher surveys and target teacher instruments, observations, and interviews.

The relationships among various types of data collection focusing on curriculum content are displayed in Table 1. The table is organized according to the variables constructed by the research team for analysis; it does not include all variables in the study, only those germane to discussions of curriculum content as an aspect of opportunity to learn. The section that follows the table also addresses only issues related to assessment of curriculum content. It begins with the teacher survey, continues with analysis of target teacher logs and log-related data, and ends with observation and interview items.[2]

---

[1] What students study-tie content of the lesson-is different from how well they learn it, to some degree. In this aspect of lesson analysis, researcher were trying to capture mainly the components of what students had an opportunity to learn, a phenomenon composed of (1) a bit of substance apprehended through (2) some medium that has its own influence on how the substance is conveyed and understock and influenced by the behavioral consequences of (3) what the teacher intends students to know or be able to do. The nature and effectiveness of any instructional strategy also have a bearing on what is learned; however, the focus on instructional practice includes a broader and somewhat different array of lesson elements, which will be discussed elsewhere.

[2] The administrator interview, prelog sur-ey, and observation form are included in this table. Although they do not touch directly on curriculum content, they do address instructional practices and resource availability, which will be discussed in other papers in this series. In those papers, Table 1 wilt be expanded to show the relative emphasis of each form of data collection.

Table 1:
Third International Math and Science Study:
Data Collection Strategies for Documenting Opportunity to Learn Curriculum Content[11]

	Focus	Teacher Questionnaire
CURRICULUM CONTENT	Major topics	22. (24.) For a list of major topics (22 in mathematics and 22 in science) indicate the number of class hours spent teaching each. (5 levels: none, 1-5 class hours, 6-10, 11+, will cover) Topics are from math and science curriculum frameworks. 21b. (23 b.) Indicate primary written source of information when deciding which topics to teach. Response options include textbook, teacher's edition of text, school, regional, and national curriculum guides.
	Topic and subtopic	23. (25.) For a list of sub-topics (18 in mathematics and 19 in science) indicate the number of class hours spent teaching each. (5 levels: none, 1-5 class hours, 6-10, 11+, will cover) Topics are from math and science topic trace mapping. 32b-36b. (35b-38b.) Given two examples each of student exercises illustrative of subtopics (5 in math and 4 in science), indicate if anything is done in class that would enable students to complete similar exercises that address subtopic. (If yes, choices are: earlier in year, currently, later; if no, choices are: covered in earlier grade, later grade, not in curriculum, don't know.) 8. (8.) Indicate amount of teacher influence on content to be covered. Response options include 4 levels (none, little, some, a lot).
	Time on topic	22. (24.) (see item above) 23. (25.) (see item above)
	Expected student mastery level	27a. (30a.) For a list of 4 possible purposes for questioning students (to develop a procedure or concept, to determine understanding or explore possible misunderstandings), indicate frequency of questioning to determine level of student conceptual understanding. (4 options: never, rarely, sometimes, often.) 31a. (34a.) Using a list of 7 assessment techniques (standardized, open-ended, and objective tests; homework, labs, teacher observation, student responses) indicate the weight given each in assessing students. (4 options: none, little, some, a lot)
	Text/materials content coverage	8. (8.) Indicate amount of influence on content to be covered and specific textbooks to be used. Response options include 4 levels (none, little, some, a lot). 18. (20.) Indicate primary student text. (Write in text information.) 19. (21.) Indicate what is used in place of or in addition to a textbook. (Open-ended response option.)
	Emphasis on topics	32c-36c. (35c-38c.) Given two examples each of student exercises illustrative of subtopics (5 in math such as such as units of measurement, whole numbers, rational numbers, and exponents; 4 in science: cells, organs, life processes, and life cycles), indicate if subtopic is emphasized in class this year. (Response options are yes/no.)

[11] Items are referenced to Mathematics (and Science) Teacher Questionnaires, Population 2, Review Versions, dated May 1994.

The Math and Science Teacher Survev

Math and science teachers in target schools completed a two-part survey. The first part, identical for both subjects, asked for information about student demographics, teacher preparation and experience, opinions related to the school's professional climate, and other details of working conditions. The second part, similar but different for math and science, included questions about student and course characteristics; responses were tied to a single section of a single course designated by the researchers.

Curriculum content items. In both the teacher surveys and the target section daily log form, respondents used topic lists to characterize the content of their instruction with respect to mathematics or science. The RUC team developed these lists by consulting professors of math and science, classroom teachers, the California and Wisconsin Mathematics Frameworks, the Wales (Great Britain) and California Science Frameworks, the National Science Teachers Association Scope and Sequence, the objectives of the National Assessment of Educational Progress, the NCTM Standards, and widely-adopted comprehensive textbooks, among other resources. The RUC team constructed the lists to provide as complete as possible a selection of easily intelligible, discreet, non-overlapping options. Their success was evident in the fact that using items on the list teachers were able to describe the content of instruction more than 98 percent of the time. Teachers used a special code to indicate topics not on the list. (The lists are attached as Appendix A.) Item 85 in the survey gave the list for science or math and asked teachers to indicate for each topic/subtopic area:

- The amount of time each was taught during the semester in the designated course, by circling the appropriate number: O for not taught, 1 for less than two hours, 2 for two to ten hours, or 3 for more than 10 hours.

- The depth of coverage, by circling the highest level of cognitive demand made during instruction: 1 for memorizing facts, 2 for solving routine problems, 3 for interpreting data or solving novel problems, and 4 for building theory or developing proofs

From responses to item 85, the researchers calculated variables related to curriculum content. The responses in Exhibit 1[3] show how one teacher described the content of a remedial math course with about 70 hours of meeting time in a semester; it provides a simple data set to illustrate how RUC researchers assessed curriculum content. Only the "circled" number is given for each to make it easier to read the display.

---

3 The responses come from a PSA summer staffer who teaches such a course during the school year.

Exhibit 1: **A** Sample Set of Responses to **RUC** Teacher Questionnaire Item #85

Topic / Subtopic	Not taught	Amount of Time Taught: < 2 hrs	2-10 hrs	> 10 hrs	Depth of Coverage: Memorize	Routine Problems	Novel Problems	Develop Proofs
	0	1	2	3	1	2	3	4
0 Number and Number Relations (Subtopics 0-9)	0							
1 Arithmetic								
O Whole numbers				3				4
1 Ratio/proportion		1				2		
2 Percent			2			2		
3 Fractions				3			3	
4 Integers			2			2		
5 Decimals				3			3	
6 Exponents			2		1			
7 Radicals		1			1			
8 Absolute value	0							
9 Relationships between operations		1			1			
2 Measurement								
O Time (not arithmetic, but units)		1					3	
1 Length			2				3	
2 Perimeter			2				3	
3 Area			2			2		
4 Volume (including capacity)			2		1			
5 Angle				3		2		
6 Weight		1				2		
7 Mass	0							
8 Rates (derived and indirect)		1			1			
9 Relationships between measures		1			1			
3 Algebra								
O Variable		1						
1 Expressions		1						4
2 Linear equations			2					4
3-8 Subtopics	0							
4 Geometry								
0 Points, lines, etc.				3			3	
1 Relationships, lines, angles		1				2		
2 Triangles, etc.		1				2		
3 Quadrilaterals			2			2		
4 Similarity		1				2		
5 Symmetry		1			1			
6 Circles		1				2		
7 Solid geom.		1				2		
8 Coordinate geom.			2				3	
9 Transformations		1			1			
Trigonometry								
0-7 Subtopics	0							
statistics								
O Collecting data			2				3	
1-8 Subtopics	0							
Probability								
O Events, outcomes					1			
1 Relative frequency					1			
2 Empirical probability			2			2		
Advanced Algebra								
0-9 Subtopics	0							
Finite/discrete mathematics								
0-1 Subtopics	0							
2 Business math			[illegible]			2		
3-8 Subtopics	0							

Time spent on each major topic. For each of the eight major topics in science and 10 in math, a value approximating the proportion of instructional time was calculated. Using the numbers circled by the respondents to indicated time taught, researchers calculated a value indicating the proportion of total class time allocated to each major topic area. According to responses in Exhibit 1, the teacher spent about one-third of the course time on arithmetic topics, about one-quarter each on measurement and geometry, and relatively small amounts of time on topics in algebra, statistics, and probability. He spent no time teaching number or number relations, trigonometry, advanced algebra, and, in the larger area of finite/discrete math, taught only business math.

Breadth of coverage. The breadth of coverage was calculated as the number of topic/subtopic areas that received scores higher than zero on "amount of time taught. " In this target math course section the breadth of coverage value is 35, because he spent some time on 35 different topic/subtopic areas.

Depth of coverage. On the survey, "depth" was calculated as average time rating per topic/subtopic area. In the sample, the value of the ratings for total time is 57 and the number of different topic/subtopic areas is 35, which yields a depth value of about 1.63 (57 divided by 35). This value can be compared to depth values for other courses and the other subject in the study, to generate a sketch of emphasis within a course and curriculum.

Higher order thinking ("HOTs" or "D marginals"). An estimate of the relative emphasis on student performance expectations at various points along a cognitive demand continuum was calculated by comparing the sum of the scores in each of the four categories with the sum for all categories. In the sample, adding the values circled under the "Depth of Coverage" (related to the "D" dimension elsewhere) heading gives a total score of 74. The total value in the "memorize" column is 10; according to the formula, less than one-sixth of the course emphasis was on this low level of student performance. Similar calculations indicate that more than one-third of the course's emphasis was on routine applications of concepts and operations; nearly one-third on extending applications; and slightly more than one-sixth for developing proofs, the highest level of cognitive demand. Researchers note that this scale system tends to underestimate the incidence of the two lower-level tasks and overestimate the higher level tasks.

Other indicators of content. In addition to item 85, the teacher survey included four other items that asked for information related to the cognitive demands and presentation modes of the course. Item 57 provided a list of live performance expectations on a continuum similar to that of the four in item 85 and asked respondents to rank the expectations according to the degree of emphasis given in the target course and to indicate approximately what percentage of time in a semester was devoted to learning of the types listed. In item 74, teachers indicated their positions with respect to two continuums reelecting different approaches to instruction; on one continuum, the extremes were "basic skills emphasis" and "conceptual understanding," and on the second, they were "basic skills learning first" and "basic skill learning in the context of analysis." In item 75, teachers indicated the relative emphasis given (none, minor, moderate, or major) to each of nine kinds of learning objectives sketching an array of cognitive demand levels. Item 56 connects loosely to the notion of presentation mode, asking teachers to indicate how often (e.g., daily, weekly, monthly, etc.) they use field trips, audiovisual resources, and print materials for lesson development.

Teacher's Dailv Log and Related Activities

Log forms. Every day teachers completed an instructional log form based on the activities of the target course section. (Log entries used in the final analysis covered a range from 109 to 177 days--with a median of 165--in the 180-day school year, collected for two complete semesters.) The form had 7 items on two sides of a single page. Three items specifically addressed the issue of curriculum content. Item 3 asked teachers to list the five most important topics of the lesson first using a brief descriptive phrase and then using a 4-digit code indicating topic, subtopic, presentation mode, and student performance expectation. This taxonomy associated with this coding system permitted labeling of 5922 different 4-digit content characterizations of math lessons and 4284 of science lessons. To promote accurate application of the coding system to lesson content, the research team developed a coding manual for participating teachers and provided them a training session on how to code.

For each of the five 4-digit codes in a lesson, the teacher indicated emphasis ranging from 1--occupying less than 20 percent of the lesson to 3--occupying more than half the lesson. Item 4 asked for the primary modes of instruction used (lecture, demonstration, recitation, whole class discussion, small group or independent work) and the emphasis on each (using the same emphasis scale as in 3). To provide some indication of the meaning of the time allocation indicators for content, Item 2 asked the amount of time spent that period on non-instructional activities. Logs were collected and reviewed for completeness weekly, and resulting data were used to create variables of several types. Those related to curriculum content are described below.

Time spent on each content tvpe. The researchers calculated a percent time value for each emphasis code in the day's lesson. An emphasis code of 1, for example, converted to a value between 8 and 20 percent of instructional time, depending on the number of other topics covered in the same period and their emphasis code. An emphasis code of 2 converted to a value between 15 and 50 percent of instructional time, and a code of 3 converted to a value between 50 and 100 percent of instructional time. The total amount of instructional time was calculated by subtracting time spent on noninstructional activities (log question 2) from the period's allocated time. Table 2 illustrates a set of codes for a hypothetical week of science instruction about population biology. (In the study itself, all combinations of content codes and emphasis codes were assigned logical values, according to the same formula that generated these.)

Exhibit 2: A Sample Set of Content Codes for a Week in a Biology Course

4-digit codes for lesson content	Emphasis pattern	Assignment of % Time	Value in a 50-minute Period with 10 Non-instructional and 40 Instructional minutes	Value in a W-minute Period with 50 Instructional minutes
3401	3	100%	40 minutes	50 minutes
3401 3301	3 1	80% 20%	32 minutes 8 minutes	40 minutes 10 minutes
3401 3301 3113	3 2 1	60% 25% 15%	24 minutes 10 minutes 6 minutes	30 minutes 13 minutes 7 minutes
3401 3301 3213 3406	3 1 1 1	55% 15% 15% 15%	22 minutes 6 minutes 6 minutes 6 minutes	28 minutes 7 minutes 7 minutes 7 minutes
3401 3301 3213 3406 0021	2 2 1 1 1	35% 35% 10% 10% 10%	14 minutes 14 minutes 4 minutes 4 minutes 4 minutes	18 minutes 17 minutes 5 minutes 5 minutes 5 minutes

Breadth of coverage. As with the teacher survey, researchers calculated a breadth score based on the total number of areas depicted by different 4-digit codes covered over the year. For example, the breadth score for the set of lessons above is six; the codes 3401, 3301, 3113, 3213, 3406, and 0021[4] define the content of the five lessons logged.

[4] In the science codes, these translate as follows: 3xxx - areas within the domain of the biology of populations; 3401 - lecturing for student understanding of population genetics; 3301 - lecturing for student understanding of natural groups; 3113 - using pictures to enable students to compare elements in a lesson on cycles in nature; 3213 - using pictures Lo enable students to compare elements in a lesson on producers, consumers, and decomposers; 3406 - using data presented in a lecture or text about population genetics to promote students' ability to recognize patterns; 0021 - in the domain of biology of the cell, using manipulatives to help students understand cell structure.

Depth of coverage. The depth score was calculated on the basis of the number of different ways each topic/subtopic combination was taught, that is, the number of different combinations of presentation mode and student performance expectation per topic, on average. The greater the number of presentation modes and performance expectation types, the greater the depth of coverage. Topic/subtopic "34xx" was approached in two ways: lecturing for student understanding and exposition of data so students can recognize patterns. The other four topic/subtopic areas were all approached in only one way: by using pictures to enable students to compare elements occurs in three codes or by using models to promote understanding. The depth score for this small set of lessons is 6/5, or 1.2. (In the study itself, all depth scores were calculated over a year of log data.)

Prelog survey. The prelog survey asked for demographic information about the students and some general information about the course, the text, and the school.

Weekly questionnaire. The weekly questionnaire was simpler still, asking only about course enrollment changes, special lesson activities that might not fit on the log form, and professional development events.

Observations and Related Reports

Log ***forms***. Teachers of target course sections were observed once or twice over the two semesters of their participation. Observers completed a log form identical to the teachers' log form.

Observation forms. Observers completed an additional form that focused primarily on instructional activities, including discourse patterns, levels of student engagement, classroom management, and overall teaching effectiveness. No questions specifically addressed curriculum content.

Observation reports. For each observation, researchers wrote a narrative report. The outline called for descriptions of lesson activities, extent to which students all studied the same material, constructivist practices, student interactions over content, proportion of time spent on instruction, student and teacher attitudes about the class and subject matter, and the physical features of the classroom.

Interview

The interview protocol covered a few topics related to curriculum content in a very general way ("What is your emphasis in this course?" "What has changed over the past few years?"). However, for the most part it focused on policy-related topics, such as the importance of various influences on course offerings and content, the ways students were assigned to classes, the professional climate, and instructional resources.

What Were the RUC Team's Findings?

Overall Findings. In general, the study determined that state and local standard-setting did have a positive influence on the number of students enrolled in science and math courses of better quality than in the past. However, the detailed descriptions of course content revealed that new conceptions of the disciplines and new information about effective instructional practices were having a limited impact in most schools. The science and math courses that were examined emphasized basic knowledge and skills presented to a passive student audience expected to meet a low level of cognitive challenge. Furthermore, testimony at various levels of the education infrastructure made it clear that few states employed "systemic" reform. In most, change efforts were fragmentary; the exception was South Carolina, which boasted a comprehensive and far-reaching approach, although its modest goal was minimum competency in basic skills.

Methodological Findings Related to Curriculum Content. The research team found that the taxonomy developed to characterize lesson content offered a clear and convenient common language for teachers. Comparisons of log data recorded separately by teachers and observers for the same lessons indicated that some dimensions of instruction could be described with a high degree of inter-rater reliability in an activity that takes only a few minutes a day. Comparisons of log data and survey data showed a degree of agreement, although differences in data collection procedures and in items made some response sets difficult to compare. For instance, in some cases the periods covered by log and survey data collection by the same teachers for the same class sections were not identical, and the log item for measuring student performance expectations did not perfectly map onto the survey item.

Taxonomy. In order to make clear the extent to which taxonomic codes were able to characterize lessons accurately over a year, teachers of target courses were directed to use a dash in their log entries to indicate when none of the topic/subtopic, presentation mode, or student performance expectation types accurately described the day's lesson. In only 1.7 percent of the science log entries and .8 percent of the math log entries did teachers use a dash for topic/subtopic codes. In less than .1 percent of the entries did they use a dash for presentation modes or performance expectations. In addition, weekly questionnaires included an item inquiring about problems using the logs and taxonomies; respondents seldom indicated difficulty with either. More than 98 percent of the time, teachers found the taxonomy adequate. This suggests that the taxonomies have considerable practical value in coding lesson content, at least in circumstances that permit offering participants some kind of written directions as well as training in how to use them properly.

Observation vs log. Thirty-nine teachers were observed twice and 36 were observed once during the study. Within this set of activities. 48 teachers keeping logs were observed 62 times by researchers who also completed entries on identical log forms for the same lessons. Researchers used these 62 pairs of logs to work out a sound method for determining the rates of agreement between teacher and observer taxonomic codes listed for each lesson (log item 3). On each pair of logs for the same lesson, they calculated the percent of agreement and averaged it over all 62 pairs. Agreement was high overall:

Topic	.78
Topic/subtopic	.68
Presentation mode	.67
Performance expectation	.59

Questionnaire vs log. Comparisons of questionnaire and log data for the same teachers and courses were confounded by problems in data collection that resulted sometimes in poor matches of timeframe. One set of correlations is based on daily log data compiled for a full year and survey data for the first half of the same year from the same teachers. A second set of correlations is based on fall log and questionnaire data; however, of these, some questionnaires were completed with reference to the previous year's fall class and others with respect to the fall class then in progress. Thus, some survey responses are affected by the error introduced by recollections of work completed and others by the error of projections of work not yet undertaken. RUC Tables 2.9 and 2.10 (excerpted on pages 17 and 18) show the correlations between log and survey responses on items about the same elements in the taxonomy. The first number in each pair shows agreement between aggregated log data for a full year and survey responses for the first half of the year. The second number shows agreement between log data for the first half of a year and questionnaire data for the first half of a year, but some survey data refer to the first half of the previous year and some to the first half of the current year (when log data were being recorded). On the basis of the overall evidence, RUC researchers assume that these discrepancies result in depressed correlation figures; in their view, more precise matches between timeframes of different data collection strategies would result in even more compelling reliability coefficients than those obtained.

In the first set of correlations (log year/survey fall) for math (N = minimum of 24), six of the 10 values are positive, significant[5], and high, ranging from .50 to .93, with values of .76 or higher for topics in algebra, geometry, trigonometry, and precalculus. In the second set for math (fall log/prior or current fall survey), five of the 10 values are positive, significant, and high, ranging from .58 to .89, with values of .77 or higher for same four topics. Only one of the six correlation figures for the lowest math levels (number, arithmetic, measurement) is significant, a finding that researchers hypothesize may be attributable to variations in the start dates of data collection at the beginning of the year, when such topics might be covered.

In science (N = minimum of 27), seven of the eight correlations in the first set (log year/survey fall) are positive, significant, and high. Only the correlation for general science is not significant. In the second set of correlations (fall log/prior or current fall survey), six are positive, significant, and high. Only general science and human biology are not significant.

With respect to student performance expectations, the survey and the log divided the continuum from the lowest level-memorizing facts-to the highest level-developing proofs-in different ways. The survey provided only four categories, while the log provided nine. (See RUC Table 2.11, excerpted on page 19.) Focusing only on correlations between categories identified the same way in each document (memorizing facts, solving routine problems, solving novel problems, and developing proofs), researchers found only three significant correlations in the 8 figures. In the first set of four, comparing full-year logs with fall surveys, only the correlation on "memorizing facts" was significant, with a value of .48. In the second set of four, comparing fall logs with fall surveys, the correlations for "memorizing facts" and "solving novel problems" were significant, although both had values less than .50.

No correlations were calculated for presentation mode as part of the taxonomic coding, because survey data did not include an item corresponding directly to that item on the log. Porter[6] reported

---

5 Significance is at the .01 or .001 levels.

6 At the School Quality Expert Panel Meeting, January 19, 1994

that, in general, the categories for coding presentation modes did not appear to be as valid or reliable as the other parts of the taxonomy. He felt that this dimension needed more work. It may be that RUC's typing scheme for modes is too unlike teachers' usual ways of considering a lesson. The identifiers seem to be accurate enough with respect to the construct, but they may not be familiar to teachers. Those particular terms may not be in common use among practitioners in the same ways that the terms for topic/subtopic areas are within disciplines and student performance expectations are across disciplines. Whether a teacher sketches a picture or idea in words by lecturing, or presents a still or moving representation on film, or illustrates with a formula is one substantively significant aspect of what a lesson may convey to students; it is a proper part of a taxonomy such as RUC proposes. However, if teachers cannot apply the terms used to characterize this aspect of a lesson accurately even with the help of a coding manual and training, then their reports on it are not helpful in documenting what students had an opportunity to learn-the object of this data collection exercise.

How Might This Work Inform NCES Data Collection With Respect to Curriculum Content?

On the whole, RUC's findings support the assertion that a taxonomy such as theirs could make a notable contribution to capturing aspects of curriculum content in the context of a relatively low-burden data collection effort specifically targeting math and science courses. In disciplines where experts have reached similar levels of consensus about what topics and subtopics constitute the core curriculum, a process similar to RUC's could be used to generate topic lists appropriate for other grade levels and other subjects. (The orignal lists developed by the content determinants group were for elementary math classes.) As national groups continue to develop content standards such as those now widely accepted in math and science, topic/subtopic lists should not be very difficult to derive.

The nature of presentation modes and student performance expectations does not differ by grade levels (although the emphasis on each may vary). This suggests that the third and fourth coding categories could be used from K-12, after the third has been adjusted to improve reliability.

RUC's approaches to characterizing lesson content show substantial consistency across methods. Teacher log items focus on aspects of curriculum content that are sufficiently clear to participants that their representation of a lesson's five most important areas of instruction matches the representation made independently on the same form by an outside observer. Where the same aspects are also covered in surveys, teachers' log entries match their survey responses for the same periods.

Recent expert testimony at meetings held to shape OERI's new agenda under the institute structure summarized emerging evidence that learning is most appropriately cha- :terized within a disciplinary framework. This conception has been developed in a series of studies (undertaken by Shulman and colleagues at Stanford and others at Michigan State) that elaborate on the intimate connections between pedagogy and content. One implication of this work is that "opportunity to learn" is not a generic phenomenon, but rather one shaped in important ways by the subject under study. RUC's approach to documenting opportunity to learn is framed within disciplines. This framing enabled researchers to draw conclusions about trends and to learn more about the influence of policy and other forces intended to enhance student achievement. If NCES chooses to document opportunity to learn at the national level, RUC's work provides one successful model of how to conduct studies based on the premise that learning is best characterized in terms of subject matter.

****

References

Clune, W. & White, P. (1992). Education reform in the trenches: Increased academic course taking in high schools with lower achieving students in states with higher graduation requirements. Education Evaluation and Policy Analysis, 14,(1), 2-20.

Porter, A., Kirst, M., Osthoff, E., Smithson, J., & Schneider, S. (1993). Reform up close: An analysis of high school mathematics and science classrooms. Madison, WI: Wisconsin Center for Education Research.

Schmidt, W. (1983a). High school course-taking: A study of variation. Journal of Curriculum Studies, 15(2), 167-182.

Schmidt, W. (1983 b). High school course-taking: Its relationship to achievement. Journal of Curriculum Studies, 15(3), 311-332.

Sebring, A. (1987). Consequences of differential amounts of high school coursework: Will the new graduation requirements help? Education Evaluation and Policy Analysis, 9(3), 257-273.

Walberg, H. & Shanahan, T. (1983). High school effects on individual students. Education Researcher, 12(7), 4-9.

Excerpt from RUC Table 2.9: Correlation Between Questionnaire and Log Data on Major Math Topics

	Number A0	Arithmetic A1	Measurement A2	Algebra A3	Geometry A4	Trigonometry A5	Statistics A6	Probability AI	Precalculus A8	Discrete Math A9
QA0	.42[7], 35[8]									
QA1		.29, .38								
QA2			.25, .58*							
QA3				.76**, .77**						
QA4					.93**, .89**					
QA5						.92**, .86**				
QA6							.50*, -.08			
QA7								-.05, -.09		
QA8									.80**, .80**	
QA9										.59*, .22

significant at .01
**significant at .001

[7] First entries in each cell represent correlations between a full year of log data and a fall semester of questionnaire data.

[8] Second entries in each cell represent correlations between fall semester log and questionnaire data, but Phase I questionnaire data describe the frill semester of the year preceding the fall semester described by logs and Phase II questionnaire data are prospective.

Excerpt from RUC Table 2.10: Correlations Between Questionnaire and Log Data on Major Science Topics

	Cell Biology	Human Biology	Biology of Organisms	Population Biology	Chemistry	Physics	Earth Science	General Science
QA0	.71**[9], .73**[10]							
QA1		.61**, .32						
QA2			.78**, .68**					
QA3				.62**, .71**				
QA4					.66**, .70**			
QA5						.81**, .66**		
QA6							.88**, .96**	
QA7								.32, .40

*significant at .01
**significant at .001

[9] First entries in each cell represent correlations between a full year of log data and a fall semester of questionnaire data

[10] Second entries in each cell represent correlations between fall semester log and questionnaire data, but Phase I questionnaire data describes the fall semester of the year preceding the fall semester described by logs and Phase II questionnaire data are prospective.

RUC Table 2.11: Correlations Between Questionnaire and Log Data on Content Dimension D (Performance Expectation)

	D0 Memorize Facts	D1 Understand	D2 Collect Data	D3 Order/Estimate	D4 Routine Procedure	D5 Routine Problems	D6 Interpret Data	D7 Novel Problems	D8 Develop Theory/ Proof
QD 1 Memorize Facts	.48**[11], .45**[12]	.24, .18	.07, .05	-.09, -.13	-.36*, -.32	-.36*, -.33	-.17, -.17	-.26, -.22	-.07.-.05
QD 2 Routine Problems	-.03, -.06	-.22, -.20	-.15, -.11	.07, .02	.18, .19	.11, .17	.15, .07	-.17, -.24	04, -.01
QD 3 Novel Problems	-.36*, -.35	.07, -.12	.18, .18	.05, .13	.06, .00	.17, .10	-.01, .11	.34, 39*	-.07, -.07
QD 4 Develop Proofs	.07, .13	.00, -.06	-.12, -.18	-.10, -.09	.01, .04	-.05, -.09	-.05, -.11	.07, .08	.14, .22

*significant at .01 ● *significant at .001

[11] First entries in each cell represent correlations between a full year of log data and a fall semester of questionnaire data.

[12] Second entries in each cell represent correlations between fall semester log and questionnaire data, but Phase I questionnaire data describe the fall semester of the year preceding the fall semester described by logs and Phase II questionnaire data are prospective.

APPENDIX A

In Reform Up Close, each lesson topic was described with a 4-digit code--ABCD--one digit for each coding dimension: A (major subject), B (subarea within major subject), C (presentation mode), and D (student performance expectation) below. For example, lesson topic 1523 in math is "Using manipulatives (C2) to demonstrate to students how to compare (D3) decimal values in arithmetic (B5 within A1)" and lesson topic 0010 in science is "Using pictorial models (C1) to help students learn the names (D0) of parts of cell structures in the general area of cell biology (B0 within A0)."

Mathematics Content Codes

Dimension A

0: Number and number relations

Dimension B

- 0: Sets/classification
- 1: Whole number
- 2: Ratio/proportion
- 3: Percent
- 4: Fractions
- 5: Integers
- 6: Exponents
- 7: Decimals (including scientific notation)
- 8: Real numbers (rational/irrational)
- 9: Relations between numbers (order, magnitude)

1: Arithmetic

Dimension B

- 0: Whole numbers
- 1: Ratio, proportion
- 2: Percent
- 3: Fractions
- 4: Integers
- 5: Decimals
- 6: Exponents
- 7: Radicals
- 8: Absolute value
- 9: Relationships between operations

2: Measurement

Dimension B

- 0: Time (not arithmetic - but units)
- 1: Length
- 2: Perimeter
- 3: Area
- 4: Volume (including capacity)

2: Measurement (Dimension B, cont.)

- 5: Angle
- 6: Weight
- 7: Mass
- 8: Rates (including derived and indirect)
- 9: Relationships between measures

3: Algebra

Dimension B

- 0: Variable
- 1: Expressions
- 2: Linear equations or inequalities
- 3: Nonlinear equations or inequalities
- 4: Systems of equations or inequalities
- 5: Exponents or radicals
- 6: Sequences or series
- 7: Functions (polynomial)
- 8: Matrices

4: Geometry

Dimension B

- 0: Points, lines, segments, rays, angles
- 1: Relationship of lines; relationship of angles
- 2: Triangles and properties (including congruence)
- 3: Quadrilaterals (and polygons) and properties (including congruence)
- 4: Similarity
- 5: Symmetry
- 6: Circles
- 7: Solid geometry
- 8: Coordinate geometry (including distance)
- 9: Transformations (informal or formal)

5: Trigonometry

Dimension B

- 0: Trigonometric ratios
- 1: Basic identities
- 2: Pythagorean identities
- 3: Solution of right triangles
- 4: Solution of other triangles
- 5: Trigonometric functions
- 6: Periodicity, amplitude, ...
- 7: Polar coordinates

6: Statistics

Dimension B

- 0: Collecting data
- 1: Distributional shapes (e.g., skew, symmetry)
- 2: Central tendency (e.g.. mean, median, mode)

6: Statistics (Dimension B, cont.)

- 3: Variability (e. g., range, standard deviation)
- 4: Correlation or regression
- 5: Sampling
- 6: Estimating parameters - (point estimate)
- 7: Estimating parameters - (confidence intervals)
- 8: Hypothesis testing

7: Probability

Dimension B

- 0: Events, possible outcomes, trees
- 1: Equally likely - relative frequency probability
- 2: Empirical probability (e.g., simulations)
- 3: Simple counting schemes (e.g., combinations and permutations)
- 4: Conditional probability
- 5: Discrete distributions - binomial
- 6: Discrete distributions - other
- 7: Continuous distributions - normal
- 8: Continuous distributions - other

8: Advanced Algebra/Precalculus/Calculus

Dimension B

- 0: Functional notation and properties
- 1: Operations with functions
- 2: Polynomial functions
- 3: Exponential functions
- 4: Logarithmic functions
- 5: Relations between types of functions
- 6: Matrix algebra
- 7: Limits and continuity
- 8: Differentiation
- 9: integration

9: Finite/Discrete Mathematics

Dimension B

- 0: Sets (e.g., union, intersection, venn diagrams)
- 1: Logic (truth values, logical argument forms, sentence logic, · ...)
- 2: Business math (interest, insurance, · ...)
- 3: Linear programming
- 4: Networks
- 5: Iteration and recursion
- 6: Markov chains
- 7: Development of computer algorithms
- 8: Mathematical modeling

Science Content Codes

Dimension A

0: Biology of the ceil

Dimension B

- 0: Cell structure
- 1: Cell function
- 2: Transport of cellular material
- 3: Cell metabolism
- 4: Photosynthesis
- 5: Cell response
- 6: Genes

1: Human biology

Dimension B

- 0: Nutrition
- 1: Digestive system
- 2: Circulatory system
- 3: Blood
- 4: Respiratory and urinary systems
- 5: Skeletal and muscular system
- 6: Nervous and endocrinic system
- 7: Reproduction
- 8: Human development/behavior
- 9: Health and disease

2: Biology of Other Organisms

Dimension B

- 0: Diversity of life
- 1: Metabolism of the organism
- 2: Regulation of the organism
- 3: Coordination and behavior of the organism
- 4: Reproduction and development of plants
- 5: Reproduction and development of animals
- 6: Heredity
- 7: Biotechnology

3: Biology of populations

Dimension B

- 0: Natural environment
- 1: Cycles in nature
- 2: Producers, consumers, decomposers: N_2, O_2, CO_2 cycles
- 3: Natural groups and their segregation
- 4: Population genetics
- 5: Evolution
- 6: Adaptation and variation in plants
- 7: Adaptation and variation in animals
- 8: Ecology

4: Chemistry

Dimension B

0: Periodic system
1: Bonding
2: Chemical properties and processes
3: Atomic and molecular structure
4: Energy relationships and equilibrium in chemical systems
5: Chemical reactions
6: Equilibrium
7: Organic chemistry
8: Nuclear chemistry
9: Environmental chemistry

5: Physics

Dimension B

0: Energy: sources and conservation
1: Heat (content and transfer)
2: Static and current electricity
3: Magnetism and electromagnetism
4: Sound
5: Light and spectra
6: Machines and mechanics
7: Properties and structures of matter
8: Molecular and nuclear physics

6: Earth and space science

Dimension B

0: Physical geography
1: Soil science
2: oceanography
3: Meteorology
4: **Geology**
5: Earth's history
6: Solar system
7: Stellar system
8: Space explorations

7: General

Dimension B

0: Nature and structure of science
1: Nature of scientific inquiry
2: History of science
3: Ethical issues in science
4: S I system of measurement
5: Science/technology and society

Codes Used for Mathematics and Science

Dimension C: Presentation Modes

0: Exposition - verbal and written

1: Pictorial models

2: Concrete models (e. g., manipulatives)

3: Equations/formulas (e. g., symbolic)

4: Graphical

5: Laboratory work

6: Field work

Dimension D: Student Performance Expectations

0: Memorize facts/definitions/equations

1: Understand concepts

2: Collect data (e.g., observe, measure)

3: Order, compare, estimate, approximate

4: Perform procedures: execute algorithms/routine procedures (including factoring), classify

5: Solve routine problems, replicate experiments/replicate proofs

6: Interpret data, recognize patterns

7: Recognize, formulate, and solve novel problems/design experiments

8: Build and revise theory/develop proofs

3. Third International Mathematics and Science Study

What Is the Purpose of the Study?

The Third International Mathematics and Science Study (TIMSS) is a research project conducted by the International Association for the Evaluation of Educational Achievement (IEA). Its purpose is to learn more about effective teaching and learning in math and science, through investigations of student achievement, instructional practices and student opportunity to learn, and curriculum content and text materials in about 50 countries. Because of the relationship between TIMSS student achievement testing and its investigation or classroom practices, findings will help contextualized variation in student achievement within and between countries. Four questions are at the heart of data collection activities:

- What are students expected to learn in mathematics and the sciences?
- Who delivers the instruction and how is it organized?
- What have students learned in math and science?
- What are the relationships among those three: the intended curriculum, the implemented curriculum, and the attained curriculum?

This study builds on prior efforts [First and Second International Mathematics Study (FIMS and SIMS), First and Second International Science Study (FISS and SISS)] of the IEA sharing the premise that to be meaningful, cross-national comparisons of student achievement must a) account for variations among countries in curriculum content and in teaching practices, and b) determine the effects of those variables on teaching and learning. Pilot administration of two questionnaires was conducted in late 1993 and early 1994 to refine instruments especially with respect to cross-national terminology and appropriateness; results contributed to major revisions in both questionnaires. TIMSS is now underway in the southern hemisphere, and data collection will follow in the northern hemisphere at the appropriate time in the school year.

What Does the Study Measure: Subjects? Grade Levels?

TIMSS measures student achievement, teachers' attitudes and qualifications, curriculum and instruction, and resources related to science and math instruction for 9- and 13-year-olds in 50 countries. Three student populations will be studied:

- Population 1 includes students in the two adjacent grades enrolling most of the 9-year-olds.
- Population 2 includes students in the two adjacent grades enrolling most of the 13-year-olds.
- Population 3 includes two groups of students:

  Students in their last year of formal secondary study in any component (e.g. vocational or academic) of a comprehensive system. The primary purpose of this sampling is to identify major characteristics to further refine the definition of this population for later use and testing for cross-national comparisons.

  A subset of these students who are enrolled in advanced programs in math or science. Separate samples will be selected for mathematics and physics.

TIMSS will collect data at three levels. School-level information will be provided by principals or department heads; class-level data on student opportunity to learn and the teacher's professional background and instructional practices will be collected from teachers; and student-level information on achievement, opinions, and attitudes will be collected from individual students.

How Does the Study Measure Curriculum Content?

On what previous and concurrent work does the study build? This project is the third in an evolving series of IEA international studies. The first set of IEA studies focused on overall school achievement results separately in math (FIMS) and science (FISS); the second round (SIMS and SISS) looked at international variation in curriculum content, teaching practices, and broad student outcomes. From the earlier results, TIMSS researchers have learned about the technical properties of items and instruments: from the activities associated with the Survey of Math and Science Opportunity (SMSO), funded to extend knowledge of effective data collection, they have further refined effective approaches to documenting teaching and learning.

What items and other instruments have been developed? Formal data collection instruments and analysis for each of the four populations include:

- In-depth Topic Trace Mapping (TTM)
- Textbook and curriculum guides document analysis
- Administrator questionnaire
- Teacher questionnaire
- Student achievement test and questionnaire

Topic trace mapping. An indepth curriculum analysis is intended to answer the first research question of student learning expectations or goals at the national/regional, school, and teacher levels. This analysis provides a description of the context within which student achievement results can be interpreted. In-depth TTM identifies: a) specific curriculum topics to which students in each country are exposed during their school career; and b) the order of instruction, depth of coverage, duration, and degree of classroom emphasis given each topic. Building on results from those trial administrations, TIMSS researchers have created six detailed topic groupings in mathematics and five in science, each with up to 14 subtopics. Subtopics were selectively chosen from the more inclusive curriculum frameworks as representing those specific subtopics most commonly included in intended school curriculums in a majority of participating countries. Two general topics, measurement and data analysis, are included in both disciplines. (See Appendix A.)

Textbook and curriculum guides document analysis. Curriculum guides and textbooks will be analyzed to ascertain the intended math and science curriculum for each participating country. The intended curriculum defines what the country's education officials expect to happen in their schools, the boundary within which classroom teachers plan and actualize their curriculum. Math and science frameworks, developed consensually by representatives from several countries and revised on the basis of reviews by every participating country, will function as international benchmarks of actual curriculum content. They will be used as a map to represent what the participating countries currently teach in mathematics and science, and to facilitate understanding of student achievement in the context of country-specific pedagogical approaches and classroom and school materials. Both frameworks have three major categories: content, performance expectations. and perspectives. (See Appendix B.)

The TIMSS analysis will occur around these sets of frameworks, using information collected at two levels. Document analysis, expert questionnaires, and interviews with education officials provide information on a national level; separate school-level questionnaires for school administrators and classroom teachers provide a finer grain of detail.

Teacher questionnaire. ***The*** teacher questionnaire is the primary method of obtaining information about classroom teacher qualifications. The instrument collects data on a teacher's professional background, pedagogical beliefs, subject matter orientation, instructional practices, and student opportunity to learn. Information in this domain will help define the available teacher incentives (such as formal and informal rewards and the quality of schools as workplaces) as well as teachers' responses to those incentives, and connect the implemented curriculum with the intended curriculum. The same questionnaire will also obtain information on classroom practices and the classroom organizational structure. The questionnaire obtains information on use of texts and curriculum guides in instruction, time allocated for and emphasis on topics, and student's prior experience with topics.

Administrator questionnaire. A separate questionnaire for school principals will obtain information on school organizational structure, teachers' professional environment, student academic programs, courses of study, and course sequencing.

Student achievement test and student questionnaire. To measure student achievement and to understand the attained curriculum, the study uses math and science achievement tests with content spanning national boundaries. Subject-matter content of each test is based on a careful analysis of science and math subjects covered by students in each population in each country. The student questionnaire is intended to shed light on the social, cultural, and economic characteristics of students that are related to student achievement in a systematic way.

Table 1 provides an overview of the foci of data collection strategies. The table is organized using six of the original ten curricular elements included on the original summary chart for this project. (See Attachment B to the memo dated June 6, 1994.)

Table 1:
Third International Math and Science Study:
Data Collection Strategies for Documenting Opportunity to Learn Curriculum Content[a]

	Focus	Teacher Questionnaire
CURRICULUM CONTENT	Major topics	22. (24.) For a list of major topics (22 in mathematics and 22 in science) indicate the number of class hours spent teaching each. (5 levels: none, 1-5 class hours, 6-10, 11+, will cover) Topics are from math and science curriculum frameworks. 21b. (23b.) Indicate primary written source of information when deciding which topics to teach. Response options include textbook, teacher's edition of text, school, regional, and national curriculum guides.
	Topic and subtopic	23. (25.) For a list of sub-topics (18 in mathematics and 19 in science) indicate the number of class hours spent teaching each. (5 levels: none, 1-5 class hours, 6-10, 11+, will cover) Topics are from math and science topic trace mapping. 32b-36b. (35b-38b.) Given two examples each of student exercises illustrative of subtopics (5 in math and 4 in science), indicate if anything is done in class that would enable students to complete similar exercises that address subtopic. (If yes, choices are: earlier in year, currently, later; if no, choices are: covered in earlier grade, later grade, not in curriculum, don't know.) 8. (8.) Indicate amount of teacher influence on content to be covered. Response options include 4 levels (none, little, some, a lot).
	Time on topic	22. (24.) (see item above) 23. (25.) (see item above)
	Expected student mastery level	27a. (30a.) For a list of 4 possible purposes for questioning students (to develop a procedure or concept, to determine understanding or explore possible misunderstandings), indicate frequency of questioning to determine level of student conceptual understanding. (4 options: never, rarely, sometimes, often.) 31a. (34a.) Using a list of 7 assessment techniques (standardized, open-ended, and objective tests; homework, labs, teacher observation, student responses) indicate the weight given each in assessing students. (4 options: none, little, some, a lot)
	Text/materials content coverage	8. (8.) Indicate amount of influence on content to be covered and specific textbooks to be used. Response options include 4 levels (none, little, some, a lot). 18. (20.) Indicate primary student text. (Write in text information.) 19. (21.) Indicate what is used in place of or in addition to a textbook. (Open-ended response option.)
	Emphasis on topics	32c-36c. (35c-38c.) Given two examples each of student exercises illustrative of subtopics (5 in math such as such as units of measurement, whole numbers, rational numbers, and exponents; 4 in science: cells, organs, life processes, and life cycles), indicate if subtopic is emphasized in class this year. (Response options are yes/no.)

[a] Items are referenced to Mathematics (and Science) Teacher Questionnaires, Population 2, Review Versions, dated May 1994.

What Are the Team's Preliminary Findings?

At this point, the only reportable findings from TIMSS come from the results of pilot teacher questionnaires and school questionnaires administered within 22 countries during September and October, 1993 (Schmidt, 1994). A total of 1439 responses to six questionnaires were analyzed to evaluate the school and teacher questionnaire items, directions, and response options and to guide their revision. Except for evaluating the appropriateness of variation in item responses, cross-national comparisons have not been made from this administration, and results have been analyzed for their methodological implications only. Technical reports are not yet available on details of validity and reliability. However, inclusion in the final set of instruments is evidence that items and instruments met TIMSS requirements with respect to these issues.

In addition to teacher and school administrator responses to individual items, data from this pilot included written advisory comments keyed to individual survey items, written-in responses to "other" options, and teacher responses to "ideal student response" items (asking teachers to present an ideal solution to a problem on the student test). Preliminary findings and changes made to the teacher questionnaire include:

- Pilot questionnaires were too long, requiring between 90 and 105 minutes for most non-U. S. teachers to complete. In combination with other revisions, the final teacher questionnaire was substantially shortened and time necessary for completion effectively reduced. All questionnaires are now estimated to require 45-60 minutes to complete, a reduction of 30 minutes from the original instruments.

- Language in some of the items was too complicated, especially in some areas and for teachers of population 1 students (9 year olds). Some items with troublesome terminology have been eliminated; wording in others has been changed.

- Some response options on some items were sufficiently out of range so as to be unused or underused. These were eliminated or revised.

- Skip patterns were difficult to follow and caused some confusion among respondents resulting in some invalid responses. All skip patterns were eliminated.

The section covering ideal student response was eliminated, and the one on pedagogical beliefs was substantially reduced. Single items regarding the teacher's role, characteristics for success with math, and computer use in the classroom were eliminated. Furthermore, TIMSS researchers added an introduction and general directions to the beginning. incorporated directions for each item into the body or the text, and restructured and reordered individual items.

Findings and changes made to the administrator questionnaire include eliminating items about the school's emphasis on specific content goals and coordination between subjects. Separate versions of the population 3 (students in their final year of secondary schooling) school questionnaire were prepared, differentiating between schools with a comprehensive program and schools with multiple student tracks.

Overall, revision of the questionnaire resulted in a shorter document with power-packed questions. Items remaining in the revised document hold great promise for highlighting important cross-national variation in educational practice and opportunity. "Every one of the items and options retained in the revised editions demonstrated significant variance in the pilot across the participating countries. In fact the cross country variations in the data suggested that the items measured concepts which seemed particularly important for cross national comparisons (Schmidt, 1994). "

Over the course of the TIMSS project development, researchers have made substantial revisions in construction of some items measuring student opportunity to learn. Initial teacher survey forms presented an item from the student achievement test and asked teachers whether classroom instruction had covered the material necessary to allow students to respond appropriately to the item. Among the problems that emerged from this approach were that teachers had differing opinions on the concept tested by the item and on the kinds of lessons that would bear on the item's topic. What initially seemed a direct and simple approach did not produce valid and reliable information. Experiments with format led to several changes culminating in the current version. First, the content is named ("units of measurement") to focus respondents' attention on the point of interest to the researchers. Second, one or more student achievement test items illustrating the content are listed, for example:

i. 2 meters + 3 millimeters equals how many meters?

ii. What is the number of 750 ml bottles that can be filled from 600 L of paraffin? "

Having thus framed the content to reduce ambiguity, the survey asks, "Is anything done in your mathematics class that would enable your students to complete similar exercises that address this academic topic'?"

Instrument developers have likewise adapted response formats to improve reliability of answers to questions about topic emphasis or time spent on a topic. Early studies asked for the percent of course time allocated or the degree of emphasis ("none, minor, moderate, major") afforded a topic. However, asking for a. percent of course time introduced measurement error, because teachers cannot give reliable estimates with such a fine metric. Similarly, asking for relative emphasis provided too little comparability among respondents-one person's "minor" was another person's "moderate." Subsequent studies, some done under the aegis of SMSO, led to the adoption of response options using an hour-cluster metric: "have covered this year in 1-5 periods, 6-10 periods, 11-15 periods, > 15 periods, will cover later this year, not covered this year, covered in a previous year." While perhaps not a fine-grained metric, researchers expect it to be more reliable than earlier measures and discriminate to a useful degree among responses.

How Might This Work Inform NCES Data Collection?

Items from the TIMSS questionnaires have undergone substantial trial testing and piloting with a wide range of education professionals in 22 countries. All items remaining in the final version have been revised, rewritten, and clarified to obtain maximum mileage for minimum respondent time. Should NCES desire to collect information in the same realms as those examined by TIMSS, the items included here could prove to be a logical and economical starting point for item development.

TIMSS ties information on school and classroom characteristics theoretically associated with student opportunity to learn to measures of student achievement, using a test that matches the universe of common curriculum across participating countries. In this way TIMSS researchers will be able to go beyond descriptive summaries of current educational practice and connect the student achievement body of data with the classroom and school body of data, identifying statistically significant structural characteristics and classroom pedagogical practices that relate substantively to high levels of student learning. Their hypothesis is that relating student achievement data to data on school, classroom, and teacher characteristics is likely to be more helpful in understanding the contributions of policy-relevant variables to educational productivity than separate measures of either OTL or achievement would be.

The international scope of the TIMSS may limit its relevance to NCES surveys. Because the TIMSS questionnaires were designed to be used in 62 countries to document international variation in learning opportunities among participating nations, the items and instruments were intentionally stripped of international ambiguities to make the concepts examined and the language used clearer to international audiences. Concurrently it may have eliminated language that functioned to clarify distinguishing effects within this country. According to Schmidt (May 12, 1994), "There was a tremendous amount of international variation demonstrated on the items. However, while this variation is what makes international comparisons interesting, it is also what makes it difficult to develop items that are always meaningful and relevant within any one country."

References

Burstein, L., Guiton, G., Bayley, L., & Isaacson, A. (1991). Compilation of items measuring mathematics and science learning opportunities and classroom processes from large-scale educational surveys (Survey of Mathematics and Science Opportunities Research Report Series No. 1). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, Graduate School of Education.

McKnight, C., Britton, E., Valverde, G., & Schmidt, W. (1992). In-depth topic trace mapping, Draft 4 (Survey of Mathematics and Science Opportunities Research Report Series No. 43). East Lansing: Michigan State University.

McKnight, C., Britton, E., Valverde, G., & Schmidt, W. (1992). Document analysis manual (Survey of Mathematics and Science Opportunities Research Report Series No. 42). East Lansing: Michigan State University.

Schmidt, W. (no date). Opportunity to learn and the SMSO study.

Schmidt, W. (1994). Teacher and school questionnaire pilot-report. East Lansing: Michigan State University.

Third International Mathematics and Science Study. (1991). Project overview. Author.

Third International Mathematics and Science Study. (1992). Mathematics curriculum framework, Draft 4 (Survey of Mathematics and Science Opportunities Research Report Series No. 38). East Lansing: Author.

Third International Mathematics and Science Study. (1992). Science curriculum framework, Draft 4 (Survey of Mathematics and Science Opportunities Research Report Series No. 37). East Lansing: Author.

Third International Mathematics and Science Study. (1993). TIMSS: Concepts, measurements and analyses (abbreviated version) (Survey of Mathematics and Science Opportunities Research Report Series No. 56). East Lansing: Author.

Third International Mathematics and Science Study. (1993). TIMSS: Concerns, measurements and analyses: Appendices (Survey of Mathematics and Science Opportunities Research Report Series No. 56). East Lansing: Author.

APPENDIX A

Topics for In-Depth Topic Trace Mapping

Mathematics

A. Place Value and Decimals

- 1.1.1.1 Meaning of whole numbers, as applied to place value and numeration
- 1.1.1.2 Decimal fractions
- 1.1.2.3 Relationship between common and decimal fractions
- 1.1.2.5 Properties of common and decimal fractions (as applied to decimals)

B. Fractions and Proportionality

- 1.1.2.1 Common fractions
- 1.1.2.3 Conversion of equivalent forms
- 1.1.2.4 Ordering of fractions and decimals fractions (as applied to common fractions)
- 1.1.2.5 Properties of common and decimal fractions (as applied to decimals)
- 1.5.1 Proportionality concepts
- 1.5.2 Proportionality problems
- 1.5.3 Slope and trigonometry
- 1.5.4 Linear interpolation and extrapolation

C. Geometry

- 1.4.2 Congruence and similarity

D. Linear Equations

- 1.6.2 (Linear) Equations and formulas

E. Measurement

- 1.2.1 units
- 1.2.2 Perimeter, area, and volume
- 1.2.3 Estimation and errors

F. Data Analysis

- 1.7.1 Data representation and analysis

Science

A. Human Biology

1.2.5.1	Nutrition
1.2.5.2	Disease
1.2.5 /1.2.1.4	Organs, tissues
1.2.5 /1.2.1.5	Cells
1.2.5/1 .2.2.1	Energy handling
1.2.511 .2.2.2	Sensing and responding
1.2.5/1.2.3.1	Life cycle
1.2.5/1.2.3.2	Reproduction
1.2.5/1 .2.3.3	Genetics
1.2.5/1 .2.3.4	Evolution
1.2.5/1.2.3.5	Biochemistry of genetics
1.2.5/1.2.4.3	Interdependence of life
1.2.5/1.2.4.4	Human behavior
1.2.5/1.6	Man's impact on environment

B. Earth Features

1.1.1.1	Composition
1.1.1.2	Land forms
1.1.1.3	Bodies of water
1.1.1.4	Atmosphere
1.1.1.5	Rocks, soil
1.1.1.6	Ice forms

C. Energy

1.3.3.1	Energy types, sources, conversions
1.3.3.2	Heat and temperature
1.3.3.3	Wave phenomena
1.3.3.4	Sound and vibration
1.3.3.5	Light
1.3.3.6	Electricity
1.3.3.7	Magnetism

D. Measurement

2.3.1	Using apparatus
2.3.2	Conducting routine experimental operations
2.3.3	Gathering data

E. Data Analysis

2.3.4	Organizing
2.3.5	Interpreting data
2.4.4	Interpreting investigational data
2.4.5	Formulating conclusions from investigational data

APPENDIX B

Aspects and Major Categories of TIMSS Curriculum Frameworks

Mathematics Framework

Content
Numbers
Measurement
Geometry: position...
Geometry: symmetry ...
Proportionality
Functions, relations, equations
Data, probability, statistics
Elementary analysis
Validation and structure
Other content

Performance Expectations
Knowing
Using routine procedures
Investigating and problem solving
Mathematical reasoning
Communicating

Perspectives
Attitudes
Careers
Participation
Increase interest
Habits of mind

Science Framework

Content
Earth sciences
Life sciences
Physical sciences
Science, technology, mathematics
History of science and technology
Environmental issues
Nature of science
Science and other disciplines

Performance Expectations
Understanding
Theorizing, analyzing, and
solving routine problems
Using tools, routine procedures
Investigating the natural world
Communicating

Perspectives
Attitudes
Careers
Participation
Increase interest
Safety
Habits of mind

MEASURING INSTRUCTIONAL RESOURCES: THE STATUS OF RECENT WORK

John E. Mullens
Lisa K. Weiner
Angela S. Williams
Mary S. Leighton

October 1994

Prepared for the U.S. Department of Education under a subcontract of Contract RN 9306100 between The National Center for Education Statistics and MPR Associates, Inc.

MEASURING INSTRUCTIONAL RESOURCES: THE STATUS OF RECENT WORK

The third component of opportunity to learn (OTL) as defined for this study is instructional resources. Instructional resources are an important link in the relationship between curriculum content and classroom teaching practices (addressed in earlier papers), because like them, classroom resources are likely to have implications for student achievement. Documenting measures of such resources and tracking them over time may be instructive in understanding variation in student achievement or learning.

Recent studies analyzing student opportunity to learn have identified three important levels of information about instructional resources. These levels distinguish among the resources available, the resources used, and the nature of their use. Simply assessing the number or amount of resources available provides information different in type and usefulness than does assessing resources actually used in the process of instruction. The former requires counting objects (e.g., pencils, books, computers); the latter necessitates a more complex assessment procedure, but provides a more intimate picture of the relationship between those materials and student achievement.

The third level of information distinguishes how those resources are used in the process of instruction: the nature, extent, and timing of their use. This differentiation among the ways teachers use instructionally -embedded resources recognizes that it is not the mere presence of the resources that makes a difference in instruction, but the details surrounding their use within the lesson. Knowing how materials are used in a lesson may provide for a more detailed understanding of the relationship between instructional materials and student learning. Woronov (1994) suggests that, "...computers in themselves do not automatically change the nature of teaching and learning; rather it is the way teachers integrate computers into classrooms, the content of technology-aided lessons, and the quality of the software programs selected that determine whether and how computers in schools really benefit students." Thus information about the circumstances in which teachers use computers (or any other resource) and the method they choose to employ can provide a level of educational insight richer and more directly connected to understanding student opportunity to learn than afforded by the simple enumerating of available resources.

information included here about instructional resources as a measure of student opportunity to learn was taken from three primary sources of national or international survey data: Reform Up Close

RUC), the Third International Mathematics and Science Study (TIMSS), and the Computers in Education Study (CIE). Principal research questions and methods for each study are summarized in Appendix A. RUC and TIMSS data include information on math and science materials only, while CIE includes information on computer usage in all curriculum areas. General information on both RUC and TIMSS surveys has been provided in previous papers in this series: similar background information about CIE is included here.

The International Association for the Evaluation of Educational Achievement (IEA) sponsored a two-stage international study of computers in education. (IEA also sponsors TIMSS.) The 1989 and 1992 Computers in Education studies were designed on the premise that, in concentrating on school-based installation of computers and technology, proponents may be neglecting the quality of computer and technological instruction (Anderson, 1993). During 1989, questionnaires were completed by principals, computer coordinators, and teachers in elementary, middle, and high schools in 23 countries. In 1992 the process was repeated in 2500 schools in 13 countries. In addition, 69,000 students enrolled in a required grade 5, 8. or 11 language arts class, selected at random in each school, completed a cognitive test and questionnaire on attitudes. published results allow cross-national comparisons as well as a more indepth look at computers in U.S. education, specifically teacher and student use of computers in the classroom.

With special attention to instructionally -embedded resources, three important elements of instructional resources will be treated below in separate sections: the types and amounts of instructional materials and equipment in use, and the types and adequacies of school facilities that promote or discourage appropriate use of those instructional materials. Because the availability and use of computers is a current and important topic and the object of substantial scrutiny, a separate section is included detailing data collection efforts in that area.

Instructional Materials and Equipment

Instructional materials and equipment provide the tangible resources with which teachers work. While no doubt some education may occur with minimal material resources, having adequate resources appropriately integrated into instruction improves opportunities to learn. Resources make the difference between a totally theoretical discussion and a hands-on experience actually using the item, perhaps providing a child's first experience with an important tool of learning and living. A basic teaching aid such as a globe can provide a crucial difference in a student's understanding of earth's continents. moving from a strictly theoretical acceptance of the teacher's word to an understanding based on that sensory experience.

Questionnaire items in this area probe the classroom use of instructional equipment, texts, and other printed materials; the frequency and purpose of use; teacher influence on material selection; and effects of material shortages. Many questions are phrased to determine the level of teacher satisfaction with the selection, amount, and use of these instructional materials. Items are listed in Appendix B and discussed below.

Texts and Coverage

Research interest in textbooks and other printed materials generally concerns the type of materials used and the percent of their content that is covered during the week or year. For example, the RUC questionnaire and the TIMSS instruments for both math and science teachers include an open-response item querying primary text use, in which respondents provide specific details about title, author, publisher, date, and edition; and the amount of the textbook covered during the week or the course. Response options for both instruments are expressed in percentages, with the TIMSS' options in four equal percentage categories (e. g., 0-25 percent), and the RUC options in roughly equal categories although with smaller high-end options (75-90 percent/ greater than 90 percent).

Frequency and Nature of Use

With the exception of the above items identifying specific classroom texts, virtually every instrument seeks information not just on specific materials used but on the method and context in which those materials are used. Recognizing that the mere presence of materials available in the school or classroom neither guarantees their use nor contributes to student opportunity to learn, researchers seek to understand how and when the materials and equipment are incorporated into the teaching/learning process. TIMSS, for example, has two related questions. The first asks teachers to estimate the percent of their weekly teaching that is based on the text, while the second seeks to determine teachers' reliance on texts and other specific materials when planning lessons.

Selection Process

Items ask about textbook selection process and the teachers' influence over purchase of supplies and the amount of money allocated for them. For example, TIMSS questionnaires seek information on the locus of responsibility for purchasing supplies and determining required texts. More specific questions are included on the RUC math and science teacher questionnaire. In asking teachers to "indicate the persons or groups who helped determine that you would use this particular textr in this... class." the RUC survey provides five choices--teacher, principal, group of teachers, distr...wide textbook commit-tee, state-wide textbook committee--in addition to an open-ended option. RUC also queries teachers about the amount of control they have over selection of texts and other instructional materials, offering six response options ranging from "none" to "complete control."

Material Shortages

School staff are asked about the effects material shortages have over their classroom program. For example, TIMSS asks teachers and school administrators to indicate the extent to which shortages of equipment limit student and teacher math and science activities.

Facilities

Facilities provide the overall setting in which instruction takes place. Well-designed and constructed facilities, properly maintained with good lighting, ventilation, and other health considerations may be taken for granted when they work properly, but a poor facility can pollute the climate, dampen enthusiasm, and impede the flow of learning by blocking or diluting efforts to provide a high-quality learning experience. However, surveys examined devote little space to cataloging information on this topic, and the items included are limited in scope and emphasis. Although recent books (Kozol, 1991) and media articles (school fire-code violations in the District of Columbia) have focused public attention on the deteriorating condition of school facilities, survey items seek information only on the condition of specific-use facilities such as laboratories, not on the condition of general-use facilities such as lavatories,

Questions about the type, adequacy, and use of facilities are included in RUC, TIMSS, and the CIE studies, and are outlined here in Appendix B. In a followup question to a query about the allocation process for instructional materials in the teacher interview protocol, for example, RUC researchers inquire about the adequacy of laboratory facilities for teaching math and science. Classroom observers note the availability and quality of a variety of classroom features, such as bulletin boards and supplies, as part of their description of the classroom. Both instances allow for openended responses. Similarly, the CIE study asks the computer coordinator the extent to which "not enough space to locate computers appropriately" is a problem, providing three response options from "not a problem" to "serious problem." The TIMSS teacher questionnaire includes a query about the extent to which "inadequate physical facilities" limit the teacher's instructional program, but the

TIMSS school administrator questionnaire is the most inclusive in its questions about school facilities. Administrators evaluate the impact on the school's instructional program of the adequacy of instructional space (e. g., classrooms), science and math laboratories, school buildings and grounds, heating/cooling systems, and lighting. Both TIMSS surveys use four response options ranging from "not at all" limiting to "to a great extent".

Each question above is embedded in the context of classroom instruction. The intent of the questions is not to generate counting of objects (such as lab tables or chalk boards), but to solicit an opinion on the adequacy of those facilities. Thus, whatever the facilities may be, however new or old, large or small, the respondent's opinion is placed within the classroom context, where the teacher's instructional goals seine as the frame of reference, rather than comparisons external to the school, classroom, or course.

CIE, the survey for which the most detailed results are published, reports that space is not a problem in 29 to 36 percent of the schools surveyed, is a minor problem in 33 to 42 percent, and constitutes a serious problem in 29 to 31 percent of schools.[12]

Calculators, Computers, and Related Equipment

Reelecting American society in general, school-based student interaction with computers is already of fact of life in the United States. Estimates from one study suggest that 99 percent of the elementary and secondary schools in the country have computers and 92 percent of students use them in their educational program (Anderson, 1993). The prevalence is easy to understand: computers are omni present in American commercial and business interactions; in schools they have educational applicability across grade levels and subject areas. Their classroom use varies from enhanced drill and practice to advanced science applications hitherto impossible. Advocates promote the improvement in quality and relevance of instruction related to regular classroom computer use. The promise of computers and the allocation of substantial resources required is questioned by many, however, at a time when some of the teaching cadre are themselves computer-shy, and severe budget constraints in some locations preclude basic expenses such as scheduled maintenance of facilities. Whether this is still the beginning or already the impending finale of a temporary phenomenon, information collected now and in the coming years may document changing attitudes and classroom uses of computers and other educational technology.

---

[12] Item results are reported separately for grades 5, 8, and 11. Percentages included here are the low and high values of those grade levels.

Perhaps with calculators and computers more than any other material resource, the presence of the machines in the school or classroom guarantees neither their use by teachers or students nor any change in the type and quality of a student's learning experience. Accordingly, no questionnaire asks solely about number of machines present. Rather, after briefly surveying the quantity of materials available (including information on the effect of shortages on educational programs), all researchers then ask how resources are incorporated into the teaching/learning process. Questions about computer use in the classroom are more extensive than similar items about calculators. After first detailing calculator questions, this section presents information on surveys of computers and related equipment. All items are included in Appendix C.

Calculators

Calculator questions are less complex than inquiries about computers, primarily seeking to differentiate among possible classroom uses. TIMSS offers five possibilities, while RUC lists 11, including drill and practice, problem solving, learning math content, homework, testing and evaluation. Both TIMSS and RUC also include questions about the frequency of use and the amount of classroom time during which calculators are actually used. TIMSS researchers also distinguish between types of calculators, querying availability and use of four-function and specialty calculators including graphing calculators, which some research has shown to have a positive effect on student achievement (Dunham & Dick, 1994).

Computers

Questions about how computers in schools affect student opportunity to learn are in four general categories: (1) number, type, and location; (2) frequency and nature of use; (3) peripherals and software; (4) the effects of equipment shortages.

Number, type, and location. TIMSS *asks* school administrators the number of computers "available in your school for use by teachers or students" providing an open response option. Not surprisingly, since each of four questionnaires is exclusively about computers, CIE items about numbers and availability are much more extensive, for example asking for availability counts by specific brand and model type, listing 11 options, each option incorporating a family of machines such as "IBM or compatible with 386 processor (includes IBM PS/2-70, PS/2-80, PS/2[y386SX], and PS/1[386SX]". A second question asks for a subset of those, distinguishing among laptop, notebook, and portable computers. Teachers are also asked to identify computer models and types used by their students "for the work they do during this class. " CIE researchers also find out how many computers have been available in the past, how many the school plans to purchase, and what model of computers they are likely to be. As the first (or in some cases second) generation of school computers begins to need replacing or updating, the model and type of replacement computer becomes of interest. As Anderson (1993) notes, while older computers may serve adequately for games and drills, educational software is no longer being developed for them, and students who are not exposed to the more modem equipment and multimedia software used in most workplaces may be at a competitive disadvantage upon graduation. CIE also collected information on the number of computers not currently being used and the reasons for their disuse, offering options such as "... broken, not working correctly/ no software for it/ no one trained to use it/ no space". Finally, CIE researchers also sought information on the location, availability, and access of computers within the school, using items such as "How many different rooms in the school are equipped with computers for teacher or student use at any one time'?" and "How many computers remain in one room and how many move from room to room?"

Frequency and nature of use. Every instrument sought information on the classroom use of computers. RUC offers teachers 11 possible response options ("teacher demonstrating computer use/ writing programs/ learning math content/ laboratory tool/ drill and practice/ using simulations/ problem solving/ using computer graphics/ games/ testing and evaluation/ homework"), while TIMSS presented nine possible use options such as demonstration, instruction, or administration. RUC also seeks data on the amount of classroom time during which students actually use computers, offering six response options in 15 minute segments. CIE researchers ask questions about school use of computers from a variety of perspectives. For example, computer coordinators are asked to assess teacher usage with an estimate of the number of "computer-using teachers" by subject area (ten listed) by grade level (three listed in high school version) and an estimate of the frequency with which teachers use computers for each of nine activities (such as exchanging messages or files between computers in your school, preparing/developing educational software, etc.) using a four level response option from "not used" to "every week. " Coordinators also assess student use of computers by type of activity (seven listed, such as "writing copy for a school newspaper or yearbook, using a computer for a computer club, " etc.) and by the subject of that activity (13 listed, such as how to type on a computer keyboard, computers for learning science, recreational use, etc) using an 11 point scale from "none" to "all". The teacher questionnaire also examines student use of machines exploring frequency of use (options range from "every day" to "less often than once per week"), the percent of instructional time students use computers (never to 75 percent), the number of days a typical student spends using a word processing program for writing essays and reports (none to 21 days or more), and the frequency with which students use specific types of computer programs (such as spell checkers, electronic thesaurus, graphics printing, etc). Similarly, the student questionnaire includes six questions on specific types of computer use in the classroom, examining frequency, subject matter, and type of use.

Peripherals and software. CIE researchers also included inquiries on peripherals and software. Similar to the level of detail incorporated in their questions about types of machines, CIE researchers also pose detailed questions about the presence and quantity of disk drives, video screens, printers, and other peripherals (such as a joystick, mouse, drawing tablet, modem, CD-ROM, etc.) as well as use of internal networks, external computer links, and network access. RUC seeks specific information on computer software used during instruction, asking for title, author, publisher, and copyright date, while CIE asks for the availability of different types of software (listing 26 options, such as drill-and-practice programs, tutorials, word-processing programs, spelling checkers, etc.) and software subject areas (listing eight options, such as mathematics, science, English, etc.).

Equipment shortages. Both TIMSS and CIE instruments ask about the effect on instruction resulting from computer and electronic equipment shortages. TIMSS asks school administrators the extent to which the instructional program of the school is negatively affected by shortages of computers and software for instruction. CIE researchers survey principals, computer coordinators, and students alike to discover the depth of computer usage problems in their schools.

Implications for NCES

In this paper, we have surveyed much of the current thinking and practice about the relationship between instructional resources and student opportunity to learn: what likely contributors to student achievement are important and how they are best measured. Evidence from previous research suggests that variation in educational achievement related to instructional resources may be caused by more than just differences in the type and level of resources available, but also by differences in the frequency and nature of use. This suggests three important areas for NCES consideration.

First, we note that questions seeking to enumerate quantities of specific resources available are present only in the longer and substantially more detailed CIE study. That study queried the specific number and types of computers and peripherals such as printers and modems. In other studies there were no questions on the amount of basic materials available in the school or classroom. Such items might question the number of texts available for each course, the number of books in the library, or the number of globes in classrooms. Except in extreme circumstances, data on materials, such as the number of books per student, may show such limited variability that its usefulness as an indicator of opportunity to learn would be nonexistent. Also, as discussed earlier, the nature of use is more informative than quantity available

Second, questions about the availability of materials. the type of materials, and their appropriateness to the curriculum, while basic, may retain merit in special situations as indicators of a school's readiness to implement curriculum reform. Direct evidence for this comes from the RUC study in which researchers concluded that the materials needed to enact current reform strategies may not even be available yet: ". . . instructional materials to support the curriculum reform of ambitious content for all students are simply not available (Porter et al., 1993). " Furthermore, in some large urban districts, the textbooks may not be available in sufficient numbers. Because teachers continue to rely heavily on student texts as a primary teaching resource, a scarcity of texts can create further problems with classroom instruction and homework. In this respect, the important information is not so much the specific number of items, but simply whether appropriate materials are available to do the job.

Third, as discussed throughout this paper, information on the nature, extent, and timing of the use of classroom and other instructional materials is important in understanding the relationship between instructional materials and student opportunity to learn. By collecting a set of national indicators about elements of resource use, NCES would be in a position to contribute substantially to an understanding of the status of instructional materials, their relationship to student achievement, and their change over time.

References

Anderson, R. (Ed.). (1993). Computers in American schools 1992: An overview. Minneapolis: University of Minnesota.

Becker, H. J. (1985). How schools use microcomputers. Summary of the first national survey. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Dunham, P. H. & Dick, T. P. (1994). Research on graphing calculators. The Mathematics Teacher, 87(6), 440-445.

IEA Computers in Education Study. (1992). Computer Coordinator Questionnaire.

IEA Computers in Education Study. (1992). Principal Questionnaire.

IEA Computers in Education Study. (1992). Student Questionnaire.

IEA Computers in Education Study. (1992). Teacher Questionnaire.

Kozol, J., (1991). Savage Inequalities. New York: Crown Publishers, Inc.

Porter, A. (1991). Creating a system of school process indicators. Educational Evaluation and Policy Analysis, 13(1), 13-29.

Porter, A., Kirst, M., Osthoff, E., Smithson, J., & Schneider, S. (1993). Reform up close: An analysis of high school mathematics and science classrooms. Madison, WI: Wisconsin Center for Education Research.

Third International Mathematics and Science Study. (1994). Teacher Questionnaire. Population 2. Mathematics. Review Version. Survey of Mathematics and Science Opportunities Research Center, College of Education, Michigan State University. East Lansing, MI: Author.

Woronov, T. (1994, September/October). "Six myths (and five promising truths) about the uses of educational technology." The Harvard Educational Letter, 10(5), p.1.

Appendix A
Selected Instruments Used to Assess Opportunity to Learn and Instructional Resources

Name of Study	Principal Investigators	Study Questions & Units of Analysis	Instruments
Reform Up Close (RUC) 1993	Porter, Kirst, Osthoff, Smithson, Schneider	• What are policies and practices governing math and science instruction at state, district, & school levels? • What is the enacted curriculum? Units of Analysis: Particular section of particular course in a particular year; also, various levels of aggregation (e.g., school, district, state)	• Daily teacher log form • Weekly questionnaire • Prelog Survey • Log entry directions • Classroom observation form • Clsrm obs report outline • Dept. chair intrvw protocol • Counselor intrvw protocol • Principal/VP intrvw protocol • Ast Supt/Curric intrvw protocol • Math/sci curric. specist intrvw protocol • Dir Testing intrvw protocol • State math/sci specist intrvw protocol • Math/sci teacher intrvw protocol
		[In Porter, *Defining and Measuring Opportunity to Learn*]	• Mathematics content codes • Science content codes • Teacher presentation codes • Lesson cognitive demand (on sts) codes • Science content taxonomy
		[From Center for Policy Research in Education]	• Mathematics/Science teacher questionnaire
Third International Math & Science Study (TIMSS) 1994-1995	Schmidt, Project Director	• What are the extent and nature of achievement differences in science and math among countries, schools, and students? • What factors explain these differences? 1. What are students expected to learn in math and science? 2. Who teaches? What is their background? 3. How is instruction organized? 4. How do 1, 2, & 3 relate to student achievement?	• Math coverage goals survey • Sci coverage goals survey • Math OTL solution set--tchr • School survey--tchr char'tcs • Tchr survey--collegiality • Principal questionnaire items • Student questionnaire items • Proposed math OTL items--tchr • Proposed sci OTL items--tchr • Sci OTL solution set--tchr • Sch survey--tchr organization • Tchr questionnaire items • Routine instruc'l materials use survey
Computers in Education (CIE) 1989, 1992	Anderson, Lundmark, Magnan, Beebe, Palmer	• What are existing forces and trends in educational computing?	• Principal questionnaire • Computer Coordinator questionnaire • Teacher questionnaire • Student questionnaire

Recent Approaches to Measuring Curriculum Content, Instruction, and Classroom Resources as Elements of Opportunity to Learn

Subject or Focus of Inquiry	Observation Reports	Interviews	Survey Data	
		Teachers	*School Staff*	*Students*
*C. Materials/Equipment *:* 1. computers/related equipment			CIE RUC TIMSS	CIE TIMSS
2. materials		RUC	RUC TIMSS	
3. facilities (room, lab space)	RUC	RUC	TIMSS	

**C. Materials/Equipment:* What do students and teachers have available to use and do use in the classroom?

Appendix B
Instructional Resources as an Element of Opportunity to Learn
Instructional Materials and Facilities

Focus		Study Questions	Data Collection Methods		
			Surveys	Interviews	Observation
Instructional Materials and Equipment	Texts and Coverage	2. What arc your primary instructional materials?		RUC-T	
		63. What is the primary text you used? (Title. author/publisher. publication date)	RUC-T		
		64. Approximately what percentage of the textbook will you cover in this course?	RUC-T		
		18. What is the primary textbook your students use? (Title. author/publisher. year. other)	TIMSS-T		
		66. Please list any other materials that you used in your mathematics/science class.	RUC-T		
	Frequency and Nature of Use	1. Provide information regarding the availability and usc of rheas materials-chemicals. glassware. batteries, weighing scales. charts. models, sinks. etc-in your target class. (Each student has own. groups of students must share. whole class must share. not available. Used in my teaching, never used in my teaching.)	TIMSS-T		
		21. Whets planning lessons, how much do you rely on your student textbooks, other textbooks or resource books. etc.	TIMSS-T		
		2b. Estimate the percent of your weekly mathematics teaching that is baaed on the text. (0-25%. 26-50%. 51-75%. 76-100%)	TIMSS-T		
	Selection Process	17. How arc instructional materials allocated at your school?		RUC-T	
		10. In your school. who has primary responsibility for (purchasing supplies, determining which text books arc used)?	TIMSS-A		
		65. Indicate the persons or groups who helped determine that you would use this particular textbook in this mathematics class.	RUC-T		
		33. How much control do you feel you have in your classroom over each of the following areas in your planning and teaching (selecting textbooks and other instructional materials)?	RUC-T		

Focus		Study Questions	Data Collection Methods		
			Surveys	*Interviews*	*Observation*
Instructional Materials and Equipment	Material Shortages	60. Use the scale below (1 = does not exist. 6 = excellent) to rate the instructional resources in your school-materials. budding, labs. software. library. etc.	TIMSS-A		
		17. To what extent is the instructional program of the school affected by shortages of budget for consumables. instructional materials. library materials? (Not at all. a very little. to some extent. to a great extent.)	TIMSS-A		
		12b. How does your level of resources compare with other schools in the district?		RUC-A	
Facilities		9. Give a physical description of the classroom. Include descriptions of availability and quality of bulletin boards. teaching materials. lab equipment. supplementary aids. etc.			RUC-Observation
		17. To what extent is the instructional program of this school affected by shortages of science/math laboratories. instructional space. adequate school buildings and grounds. adequate heating/cooling and lighting? (Not at all. a very little. some. a great extent.)	TIMSS-A		
		17c. Are lab facilities adequate for teaching your course?		RUC-T	

Key: A = Administrator/School representative
T = Teacher

Appendix C
Instructional Resources as an Element of Opportunity to Learn
Calculators, Computers, and Related Equipment Items

Focus		Item	Survey Population*
Calculators		59. How does this science class use ... calculators? (Teacher demonstrations. writing programs, learning science content, lab tool, drill/practice, games. testing, homework. other?)	RUC-T
		60. During the last week of instruction. how many minutes did a typical student spend working with ... calculators as part of this science class? (None. 1-14 minutes. 15-29 minutes...more than 60 minutes.)	RUC-T
		5b. For which of the following do students in the target class usc calculators? (checking answers. tests and exams. routine computation. etc.)?	TIMSS-T
		2. If your students usc calculators. what type do they use? (four function. scientific. graphing, programmable)	TIMSS-T
		22. Do your students have calculators? (Almost all, about half. a few. none have calculators) Are the calculators school or student owned?	TIMSS-T
		1. For which of the following were students in the target class allowed to use calculators? (Calculators not available. students not allowed to use calculators, checking answers in class. tests/exams. homework. extended projects. routine computation in class.)	TIMSS-T
		3. Do you ever usc an overhead projector or calculator during the lesson?	TIMSS-T
		80. We use hand calculators in mathematics. biological and/or physical science classes.	TIMSS-S
Computers	Number Type and Location	1. Please provide a count of all computers ● vailable to teachers or students. according to type listed below.	CIE-CC
		1b. How many of the above computers arc laptop. notebook, or portable computers?	CIE-CC
		2,2a,b. Are any of the computers listed in Q.1 not being used at all at this time? How many? Why? (broken, no software. no space)	CIE-CC
		3. What, then, is the total number of computers available for teacher or student use that are being used by teachers and/or students?	CIE-CC
		4. How many computers arc used by students and teachers in classrooms. computer labs, offices. and other types of rooms? (type of room. number)	CIE-CC
		5. How many computers remain in one room (for at least a semester), and how many move from room to room?	CIE-CC
		6a, b. Altogether, how many different rooms in the school arc equipped with computers for teacher or student use at any one time? Arc one or more of these rooms exclusively used by teachers?	CIE-CC
		7. In the table below, please provide information about the 3 rooms at your school that have the greatest number of computers or terminals for teacher or student use. (number of computers, number of teachers who use, number of hours used in typical day)	CIE-CC

Focus		Item	Survey Population*
Computers	Number Type and Location	17. Approximately how many computers or terminals did your school have for teachers or students to use 1, 2, 3, and 4 years ago?	CIE-CC
		18. How many computers do you expect your school will obtain in the next year, given realistic assumptions about hardware cost, space, and software availability?	CIE-CC
		19. Which models of computers are you most likely to obtain during the next year (including more of the models you already have)? For each model listed below, indicate the number of computers you expect to obtain.	CIE-CC
		9. How many microcomputers or terminals are usually available for use by this class?	CIE-T
		10. Which type of computer is used by this class for the work they do during this class?	CIE-T
		16. Computers - number available for teacher/student use, number available for teacher/administrator use, number used regularly by teacher for instruction, number used regularly by students?	TIMSS-A
		58. Which best describes the availability of computers for use with this science class for teacher demonstrations, student use in classrooms and student use in labs? (Not available, available but difficult to access, readily available.)	RUC-T
		59. How does this science class use computers...? (Teacher demonstrations, writing programs, learning science content, lab mol, drill/practice, games, testing, homework, other?)	RUC-T
		59. How many computers are available in your school for use by teachers or students?	TIMSS-A
	Frequency and Nature of use	4. For which of the following activities have students in the target class used computers? (practical exercises, problem solving, etc.)	TIMSS-T
		60. During the last week of instruction, how many minutes did a typical student spend working with computers and calculators as part of this science class? (None, 1-14 minutes, 15-29 minutes...more than 60 minutes.)	RUC-T
		81. We use computers in mathematics, biological and/or physical science classes.	TIMSS-S
		29. For each subject below, check the number of teachers using computers and circle the specific grade levels at which they use the computers.	CIE-CC
		32. How often are computers in your school used by a teacher for each of the following activities: (exchanging messages or files, giving a lesson, training, etc.)	CIE-CC
		31. How often are computers in your school used by a student for each of the following activities? (work at home, exchange messages, computer club, etc.)	CIE-CC
		30. Roughly how much (none, little some, much, half, most, all) of all student use of computers will involve each of the following activities? (variety of computer skills, computer use in content areas)	CIE-CC
		12. During a typical week in which computers are used by students in this class for reading, writing, or language arts, how many days does one student in this class use computers?	CIE-T

Focus		Item	Survey Population*
Computers	Frequency and Nature of use	13. For the two types of instruction below, estimate roughly what fraction of the time students do that kind of work using computers. (75%, 50%, 25%, 10%, never, n/a) (grammar, writing)	CIE-T
		15. How many days did a typical student in this class spend using a word processing program for writing essays and reports for this class during the school day?	CIE-T
		16. Since this school year began, how often has a typical student used each of the following kinds of computer programs on school computers? (spell check, outlining, programming, etc.)	CIE-T
		21. How many times did you use school computers in these subjects during this school year? (frequency, subject)	CIE-S
		26, 27. In which school contexts do students get a substantial amount of instruction focusing on computers (computer literacy, programming, computer applications skills)? (pull-out, computer period, learn about subject areas). In which context do students get most of their instruction about computers?	CIE-CC
		20. For how many years have you used computers for professional activities such as writing, keeping records, or making materials?	CIE-T
		20. In which grades have you used computers in school?	CIE-S
		22. During this school year, how many rimes did you use computers in school in each of the following different ways? (learn something new, lab experiments, taking tests, playing games, etc.)	CIE-S
	Peripherals and Software	7e. Indicate the software you plan to use. (Title, Author(s), Publisher,	RUC-T
		10. How many of the following types of printers are available for usc with the computers used by teachers or students?	CIE-CC
		12. Which of the following peripherals are available for usc wish at least one computer at your school? (joystick, modem, optical scanner, CD-ROM, etc.)	CIE-CC
		9. How many of your school's computers arc equipped with each type of video screen listed below? (monochrome, color, t.v. projection screen)	CIE-CC
		8. How many of the microcomputers use each of the following "media" for storing programs and files? (network)	CIE-CC
		13. Please indicate whether your school has the following "internal" networks. (frequency of use) (link in same classroom or across classrooms)	CIE-CC
		14. Please indicate whether your school has the following "external" computer links. (school/central office, school/network)	CIE-CC
		15. Does anyone at your school participate in the following networks (list)	CIE-CC
		13b. If you use any computers as network servers, please list the brand and model, along with the total number you have.	CIE-CC
		21. Which of the following software is available for teaching and learning purposes at your school? (drill-practice, spell check, database, gradebook, etc.)	CIE-CC
		23. Consider all of the instructional software that is in use at your school. Roughly, what portion of that software was produced noncommercially?	CIE-CC

Focus		Item	Survey Population*
Computers	Peripherals and Software	8. How many of the microcomputers use each of the following "media" for storing programs and files? (diskettes, hard disk, cartridges, cassettes)	CIE-CC
		22. For which of the following subjects is software available in your school or easily accessible through agencies outside the school, for teaching and/or learning activities? (content areas)	CIE-CC
	Equipment Shortages	17. To what extent is the instructional program of the school affected by shortages of computers for instruction, software, audio visual resources, calculators? (Not at all, very little, some extent, to a great extent.)	TIMSS-A
		12.(33). Listed below arc some problems that can affect your school's ability to usc computers effectively. Please read each alternative and indicate how serious the problem is for your school. (too few printers or other peripherals/software: not enough, not sound, too complicted, not useful, not enough information/can't fit into curriculum, inappropriate, teachers not interested/not enough space.)	CIE-A CIE-CC
		26. Listed below are a number of problems that students have reported in using computers in schools. For each problem, indicate how often (never, sometimes, often, very often) this has been a problem for you in school during this school year. (not available, too hard, not interesting, no help, broken)	CIE-S

*Key:
A = Administrator/School Representative
CC = Computer Coordinator
T = Teacher
S = Student

MEASURING OPPORTUNITY TO LEARN: ADVANCING THE STATE OF THE ART

Mary S. Leighton

with the assistance of
Brenda J. Turnbull
John E. Mullens

October 1994

Prepared for the U.S. Department of Education under a subcontract of Contract RN 9306100 between The National Center for Education Statistics and MPR Associates, Inc.

MEASURING OPPORTUNITY TO LEARN: ADVANCING THE STATE OF THE ART

Since undertaking the task of determining how NCES could assess school quality, beginning with opportunity to learn (OTL), this team has gathered and analyzed instruments from ten major studies and consulted with researchers to learn what they have discovered about the strengths and weaknesses of instruments assessing curriculum content, instruction, and classroom resources. In several cases, new information has emerged in the last few months as the researchers analyzed and reported on the results of their studies, and in other cases, final reports on findings and the quality of instrumentation have not yet been released. In some subject areas, we have identified pools of items that have been used to learn what was taught and how.

Here we propose four potentially viable projects for data collection and analysis through which NCES could learn more about the extent and nature of opportunity to learn in American schools. We begin by discussing briefly the purposes of measuring opportunity and describing four dimensions of the educational context that govern decisionmaking in this arena. Then we explain how each of the four suggested projects might seine NCES purposes. One idea takes advantage of current NCES work-Schools and Staffing and Teacher Followup Surveys-to improve the store of information about opportunity to learn. Three others are small pilot studies that could improve understanding of educational processes and outcomes while speeding the development of survey items that could eventually be included in regular NCES data collection activities.

Purposes of NCES Studies of OTL

In the long term, NCES hopes to collect data that can portray key dimensions of school quality in sufficiently accurate and comprehensive terms to inform analysis and provide a context for understanding student achievement. The goal of a new study would be to identify and/or develop and test sound measures of OTL, defined as an amalgam of the enacted curriculum, instructional practices, and resources used for learning.

- Enacted curriculum[1] is the knowledge and skills the teacher presents for the students to learn during their classroom experiences. Taken as a whole, the studies reviewed for this p reject characterize the enacted curriculum in a given subject area as a function of topics covered, time spent and emphasis on topics, targeted student mastery levels and forms, students' prior learning, and reachers' knowledge, among ocher things.

- Instructional practices are defined by current studies as the activities of teachers and students that flesh out the meaning of the lesson's content. Presentation format, the forms and extent of student engagement, grading practices, allocation of time within a period, homework, and attitudes influence how students and teachers develop lesson content.

- Learning resources are materials and equipment routinely woven into lessons in ways that contribute to content, for example computer demonstrations and applications, manipulatives, laboratory experiences, data bases, texts, or other instructional equipment and supplies. (Other things may reasonably be considered resources, of course, but in this context we consider only those chat bear immediately on lessons, what Porter tails "instructionally-embedded resources.")

This OTL construct is one focus of several major studies, and it remains one of the best predictors of student achievement, according to Porter (1991). Valid and reliable information about opportunity to learn provides immediately useful insight about aspects of education that are amenable to change. Although measurement problems in this area persist, work recently undertaken and partially completed already offers a strong foundation for further study. Such study could serve as a natural extension of present NCES efforts to provide policy-relevant information at the federal level.

Key Dimensions of Context

Four dimensions of the context of OTL measurement influence the usefulness of work in this area:

- Scope
- Subject matter

---

[1] Here we borrow a term from TIMSS researchers, who have identified three forms or stages of curriculum: The intended curriculum is spelled out in policy documents and frameworks that guide content decisionmaking at all levels. The enacted curriculum is what teachers present in classroom. The attained curriculum usually goes by the name of "achievement"—it is what students actually learn.

- Technical goal
- Achievement context

Some of the ways these dimensions shape decisions about measuring OTL are described below.

Scope

The scope of an OTL study can be either general or subject specific. Research on teaching has identified a number of instructional features that influence opportunity to learn across subjects. NCES could learn about these general features of instruction using the same set of items and instruments for all K-12 teachers. However, scholars in this field have argued persuasively that opportunity to learn has characteristically different traits in each discipline. Although the most valid and reliable information may come from studies of OTL within subjects, such information necessarily has limited generalizability across the curriculum. A clearer picture of a narrower slice of educational experience may be more revealing and ultimately more useful than a broader view with few details.

Subject Matter

If one subject is to be chosen for a study, the choice must accommodate the demands of both clarity and importance. Clarity requires a degree of consensus about subject-matter boundaries, in order to keep the data collection activities manageable and comparisons apt. Current instruments depend on widely-accepted topic lists to generate accurate accounts of what is taught. Generating these lists requires agreement among experts about the major and minor topics in each discipline and what therefore should be included in various courses. Such agreement is sometimes hard to get. For instance, courses offered by different schools under the title "biology" tend to include unique subsets of a long list of "biology" topics. Comparisons among courses or analyses of what has been learned--as demonstrated on a given achievement test-in light of what has been taught is often difficult in biology. However, most courses called "physical science" overlap to a great extent in what they cover, and students enrolled in such courses might reasonably be expected to recognize most of the items on an achievement test purporting to assess knowledge in physical science.

In addition, the subject targeted for study should be generally considered important. Developing instruments and collecting data on a national sample is too costly an enterprise to undertake in subjects that are easily circumscribed but of marginal interest. Philosophy may have much to offer but it is not currently a central achievement concern, whereas math and reading are. Selecting a subject-matter focus in which both educators and the larger community have a strong interest increases the educational value and the practical uses of the findings.

Technical Goal

The choice of subject (and grade level) should be compatible with the technical goal of an OTL study. For example, pushing the frontiers of assessment technique requires choosing a subject in which much is already known about instrumentation. The next step is then to assemble the best items and instruments, make the adjustments that the most recent findings suggest, and try the new, advanced system on a new group of respondents. If, alternatively, the goal is to broaden application of recently-developed techniques to new subject matter, then choosing a subject and grade where less has been done and investing in adaptations of new valid and reliable assessment techniques for that setting makes more sense.

Achievement Context

The choice of setting for studies of OTL should reflect whether having relevant achievement data as a context for understanding OTL is a high or low priority. If the purpose of a study includes illuminating the relationship between OTL and achievement, then the study should focus on a setting where relevant achievement outcomes are available or could be obtained without additional undue expense or response burden. The achievement data should describe performance on subject matter that closely matches the type of OTL assessed. If the purpose of a study is only to improve assessment methods, by either pushing the frontiers of technique or applying advanced techniques to new subject areas, then whether achievement data are available or relevant to the OTL assessed is immaterial. On one hand, having a well-defined achievement context may deepen understanding of the results of assessments of students' opportunity to learn. On the other hand, finding a suitable setting may be difficult and using such a setting may raise other issues about accountability that could skew response patterns by introducing a motive for biased responses.

The following sections describe briefly four promising possibilities for future efforts to assess opportunity to learn: K-12 instruction in general, eighth-grade math, fourth-grade language arts (reading, writing, literature), and U.S. history. Table 1 summarizes the features of each with respect to the dimensions listed above.

Table 1: Summary of Features of Proposals for NCES Pilot Studies of OTL

Study Focus	Scope	Subject Matter		Technical Goal	Achievement Context	
		Clarity	Importance		Possible Data Source	Match with OTL Focus
K-12 Instruction	General	Low	High	Broaden Application	State, local, NAEP	Probably Distant
Grade 8 Math	Specific	Med/High	High	Improve Technique	State, local, NAEP	Close
Grade 4 Language Arts	Specific	Low	High	Broaden Application	State, local, NAEP	Probably Distant
U.S. History	Specific	High	Med/High	Broaden Application	State, local, NAEP	Possibly Close

Focus: K-12 Instruction

Rationale

Two main factors support the idea of conducting a study of the most general aspects of opportunity to learn in K-12 classrooms. First, a growing body of evidence indicates that certain global features of curriculum content, instructional practices, and resources are associated with improved achievement across subjects and grades. Second, such a study could be added to existing NCES data collection activities involving the appropriate population samples.

The broad scope of such a study could reveal general trends, even though it would not add much detailed knowledge about differences in opportunity for different students. Because all subjects would be included, the study could create a crude map of the content territories covered, which could be helpful in understanding variations in general achievement. Although underachievement in math and science is headline news, concern about overall educational attainment is widespread, which lends significance to this focus. A K-12 study would broaden the application of recently-developed methods. If having achievement data as an indicator of attained curriculum is determined to be desirable, existing sources of such data could be used, although these outcomes might not match the inputs assessed in a broadly aimed set of OTL assessment items. If, on the other hand, simple description is the goal, such a study has much to offer.

Design

PSA has proposed[2] items to be included in the next Teacher Follow-up Survey (TFS) to assess this general form of opportunity to learn. Items related to pedagogy cover teacher and student actions, instructional organization, student assessment (including the use of portfolios), and homework. An additional item about teachers' informal professional development activities--professional reading, collegial collaboration, and Curriculum development-has been suggested. In the area of resources, we recommended collecting data on texts and other instructional materials and the nature and frequency of calculator and computer use across disciplines. The items have been adapted from instruments in current or recent use. (See Appendix A.) In combination with other TFS items, the new items would permit analysis of the extent to which reform recommendations in curriculum and instruction are influencing students' opportunity to learn and the type and dimensions of differences among opportunities provided to different student populations.

Focus: Eighth-Grade Math

Rationale

If technical advancement at the furthest frontier is our goal, then eighth-grade math presents one potentially productive setting for this study. While the contest varies considerably from remedial topics in arithmetic through algebra, the domain is well-mapped. Several existing content coding systems, carefully created and well tested, are available for use. Furthermore, the subject is of great interest to policy makers and planners. In most systems, eighth grade is the last relatively affordable opportunity for students to finish preparing for algebra in high school, and most educators view algebra as a gateway course, facilitating access to academically rigorous and engaging secondary school experiences. The curriculum covers relatively advanced arithmetic and algebraic concepts that are unlikely to be learned explicitly in out-of-school experiences, so what is taught in school may be especially important. Many states administer a benchmark assessment in eighth-grade math, as will both TIMSS and NAEP in the near future. Because many of the well-funded, highly visible recent studies have covered math, a rich bank of items and instruments is available. In addition, the existence of state benchmark and other assessments gives us an opportunity to compare our findings with standardized test results, which could improve our understanding of our findings about instructional content.

---

[2] Under a subcontract of contract RN93140001 between NCES and the American Institutes for Research

Design

This study would include a survey of a diverse sample of schools–perhaps nine-with case studies nested within the sample. Measurement strategies and data analysis would include the following:

Measurement strategies. Reviews of recent major studies suggest that the most cost-effective strategy for getting high-quality information about instructional content includes a combination of teacher surveys, logs, and/or weekly reports on instruction; classroom observations coded like the surveys; artifact collection, including student work; and interviews with teachers to clarify ambiguities in data and analysis.

- Survey items: TIMSS, RUC, and the RAND/UCLA studies of opportunity to learn developed new items and item formats in recent months, basing revisions on analyses of data from interviews, observations, teacher logs, and other artifacts, as well as of curriculum materials currently in broad use. For eighth-grade math, items and instruments with well-documented validity and reliability are available.

- Teacher logs and/or weekly reports: RUC and RAND/UCLA teams have developed daily log forms that provide valid and reliable data with a brief investment of time. The RUC team's forms can be used in conjunction with observations tied to the same reporting framework. Whether to gather log data for a whole year or a semester or a few weeks (variations used by RUC and RAND/UCLA) is an issue that will be resolved in part by findings documented in the final RAND/UCLA reports, expected to be published soon. The content/activity codes on the RUC forms were found to be limited with respect to descriptors for instructional activity, so those would need to be rewritten in light of RUC's experience. The present RAND/UCLA log form seems too sketchy to be helpful.

- Observations: RUC researchers reported getting little significant new information from observation reports (see attachments). However, the RAND/UCLA team chose to avoid observations on the grounds that they would be too time-consuming to be undertaken extensively on a national scale, and then found that without observations their data remained too ambiguous. Case studies should use an observation form that ties to logs and survey items, using RUC's experience to improve the data benefits.

- Artifacts: RAND/UCLA and TIMSS use different forms of artifact and/or materials analysis to enrich portrayal of lesson content. TIMSS does in-depth analysis of the primary texts and curriculum guides, and RAND/UCLA collected student work, text pages, assignments–a stack of documents related to lessons. The RAND/UCLA team found that its strategy provided data that could be used iteratively, beginning with a modest sampling of each collection and elaborating with new items until they achieved a stable interpretation. This seems more efficient for getting a clear picture of classroom experience in settings where we expect to find substantial variation. The

TIMSS approach seems better suited to situations where many of the members of the sample are working in countries with less curricular variation. The RAND/UCLA team has agreed to advise on what constitutes a sufficient artifact base for cross-checking content. (They gathered and analyzed more artifacts than they needed to establish a reliable measure.) Case studies will include collection of whatever artifacts proved productive as well as copies of portfolios of student work completed during the period of the study.

- Interviews: Conducting interviews of teachers and students will be part of case study work. Interviews will probe on teachers' completion of log entries, student portfolio elements, artifacts, and survey responses. Preliminary analyses of survey and case study data will be reviewed with groups of informants at each participating school to determine the extent to which what the data suggest coincides with their understanding of what happened during the period of study.

Data analysis. Adaptations of the analytic procedures that the RUC and/or the RAND/UCLA teams developed for integrating data will be used co develop portrayals of the opportunities to learn during the period of study. These procedures are well developed, and the research teams are willing to share them. However, they are not reported in current documents. In eighth-grade math, the recently-used procedures will require little adjustment.

Focus: Fourth-Grade Language Arts

Rationale

We agree with Porter's (1991) suggestion that a major focus on English language arts (reading, writing, and literature) would provide an appropriate counterweight to recent studies of math and science and balance attention within the core curriculum. While it constitutes a large and usually recognizable part of daily instruction, however, its boundaries are not clear. Indeed, advocates of whole-language approaches to literacy development and integrated instruction often weave it so artfully into every subject that it may diffuse itself entirely for scheduling purposes. BY sixth or seventh grade, reading instruction may not be offered on a regular basis (although it is still available as a remedial class, usually j. Because English language arts is a major focus in elementary school and acquiring mastery of basic skills and knowledge is most important for later school success and most easily achieved with the resources of elementary school, targeting that subject at the fourth grade level seems valuable. Such a focus would involve broadening the application of current techniques to include a new content area. Achievement data are readily available at that age. Unfortunately, the lack of clarity about subject domain may make it difficult to relate the findings about OTL specifically to achievement levels.

Design

Beginning with the NAEP framework for reading, any draft documents available from associations of language arts specialists, and a few language arts frameworks and/or curriculum guides from states or districts that are advanced in this area (California frameworks come to mind), a topic/subtopic list could be created in a form similar to those used by RUC and RAND/UCLA in science and math. Porter suggests limiting the lists to a few topics to keep the task manageable (1991). Using the document produced by this exercise as a guide, a study along the lines of that proposed for eighth-grade math could be conducted. This would include surveying a sample of schools and nesting within the larger sample a strategically sampled group of case studies. Surveys, teacher logs, observations, artifact collection, and interviews with students and teachers would contribute to the data base.

Focus: U.S. History (Grades 4-12)

Rationale

McDonnell and Burstein, chief researchers in several of the RAND/UCLA studies, recommended shifting the focus again to history or social studies, targets of some of their early work with curriculum indicators.[3] NAEP has a framework for U.S. history, and the National Council for Social Studies is about to publish K-12 curriculum standards similar in purpose to those for math. Documents such as these provide the grounds for fairly high levels of domain clarity, and the prominence of the subject matter in existing development efforts indicates its importance. History would be a new area in which to test new assessment techniques. If having related achievement data is a priority, it may be possible to obtain data reasonably well-matched to the OTL dimensions that are assessed. Furthermore, available documents map the subject across grade levels, which extends the potential scope of the study as broadly as possible within the confines of a discipline.

Design

Using existing frameworks, a topic/subtopic list could be created for U.S. history, as the starting point for a study similar to that proposed for eighth-grade math. The study would use a

[3] Personal communications, May 16-17, 1994

comprehensive survey of a broader sample of schools and case studies nested within this sample. In this case, the pilot study populations should be selected from districts or states with strong support for social studies education. An earlier PSA study of the status of curriculum standards development across the country (Pechman & Laguarda, 1993) indicates that in several states curriculum frameworks for social studies have already been developed and their use is supported by accountability systems that improve the likelihood that content coverage is made a priority. The demonstrated interest of these sites in promoting opportunities for students to learn history may ensure willingness co participate in a study such as the one proposed.

Final Thoughts: Opportunities to Collaborate

First, NCES may be able to use work developing its OTL assessment capacity to cultivate additional opportunities to support or take advantage of present initiatives of the U.S. Department of Education, LEAs, and SEAs. For example, pilot studies could be located in states whose Goals 2000 planning proposals require establishing some baseline data on school conditions and distribution of learning resources. These states may be glad to collaborate in order to facilitate development of approaches that can eventually be converted to their purposes, to *create* local indicators. Likewise, districts deliberating on whether to concentrate Chapter 1 funds into schoolwide projects may be looking for pinners. While Porter cautions (appropriately, in our view) against using OTL assessment in an accountability system, we found that the studies show great promise for enabling educational agencies to characterize some dimensions of their productivity that are generally under their control. We think that the process of developing a combined survey/case study assessment strategy will generate a lot of useful information, and, hence, that approaching pilot sites as potential collaborators makes sense. New studies of OTL represent opportunities for reinvention.

Second, opportunity co learn studies that are thoughtfully designed to capitalize on the findings of recent research can provide rich and renewable data bases for informing analysis of reform. Such studies can use the qualitative data collection co home in on practical definitions of response burden as well as survey focus. They can develop the kinds of survey items that engage the willing and reflective responses of participants. Researchers with extensive experience in this field reported that participants were perfectly agreeable about spending time on surveys that caused them co consider their work in new and useful ways and to comment on aspects of their work that they found to be important. What bothers respondents are questions that seem pointless no matter how little time it takes to answer them. Observations, interviews, focus groups, and artifact analysis have helped researchers to frame questions and develop response formats that were interesting as well as efficient. The RAND/UCLA team made it clear that extended, personal contact with a small sample of respondents had the potential to be an ongoing source of insight into the meaning of survey data responses and inspiration for improvements in survey contents.

Finally, given the choice to conduct one study, we would choose U.S. history. Math and science are well-studied, and more news is already in the pipeline from RAND/UCLA and TIMSS, among others. The recent controversies about language arts frameworks bode ill for having a curriculum standard document soon. The pedagogical features of U.S. history have been the focus of considerable research (by Wilson and Wineburg, with recent notable success) and are being documented in the work of the National Board for Professional Teaching Standards. The subject is of high interest, but not of high visibility in the national discussion of achievement, and including it would broaden the application of new techniques. We think history would offer a welcome change of pace and provide educators with an interesting and stimulating opportunity to learn about American educational process in a uniquely American process.

APPENDIX A

Draft

QUESTIONNAIRE ITEMS FOR THE TEACHER FOLLOWUP SURVEY

EXPLORING OPPORTUNITY TO LEARN

John E. Mullens

August 1994

Prepared for the National Center for Education Statistics under contract RN 93141001 with the American Institutes for Research. The views expressed are those of the author, and no endorsement by the Center should be inferred.

This section asks about the teaching strategies, instructional practices, and organizational techniques you use in teaching. The information you provide is intended to describe students' educational experiences and inform future national surveys of school processes.

Please answer the following questions thinking of the class for which you had primary responsibility last semester or grading period. If you were responsible for a single group of students ail day (such as an elementary teacher might have been), think of them as the designated class. If you were responsible for multiple classes or groups of students (such as a content area or special education teacher might have been), select your first instructional class or group of the day (not homeroom). Think of this as the "DESIGNATED" class.

1. In what grade level(s) were the students in your designated class? Circle all that apply.

Ungraded	6th
Prekindergarten	7th
Kindergarten	8th
1st	9th
2nd	10th
3rd	11th
4th	12th
5th	Postsecondary

2. Which one or the following best describes your designated class? Circle all that apply.

heterogeneous

homogeneous

remedial

special education

gifted

academic/college preparatory

advanced placement/college credit

honors course

vocational

bilingual

none of the above

3. Write in the percentage of students in your designated class who were at each level of academic ability for their age and grade. (Numbers should total 100.)

much above the NATIONAL average	______
somewhat above the NATIONAL average	______
at the NATIONAL average	______
somewhat below the NATIONAL average	______
much below the NATIONAL average	______

4. Write in the percentage of students in your designated class who were at each level of academic ability for their age and grade. (Numbers should total 100.)

much above the SCHOOL average	______
somewhat above the SCHOOL average	______
at the SCHOOL average	______
somewhat below the SCHOOL average	______
much below the SCHOOL average	______

5. Over the past semester, how often did YOU use each of the following instructional strategies with your designated class? The strategy need not have taken the entire class period. Circle one response on each line.

	Almost every day	Once or twice a week	Once or twice a month	Once or twice a semester	Never
a. Provide instruction to the class as a whole	4	3	2	1	0
b. Facilitate a discussion	4	3	2	1	0
c. Demonstrate a concept using the board or overhead projector	4	3	2	1	0
d. Work with individual students	4	3	2	1	0
e. Demonstrate a concept using a computer or videotape	4	3	2	1	0
f. Lecture	4	3	2	1	0
g. Work with small groups of students	4	3	2	1	0
h. Administer a test (full period)	4	3	2	1	0
i. Lead question-and-answer session	4	3	2	1	0
j. Demonstrate a concept using manipulatives, models, other tools or objects	4	3	2	1	0
k. Administer a quiz (less than a full period)	4	3	2	1	0

6. In class over the last semester, how often did planned activities require that STUDENTS:

	Almost every day	Once or twice a week	Once or twice a month	Once or twice a emester	Never
a. Respond orally to questions testing recall	4	3	2	1	0
b. Use school- or student-owned calculators	4	3	2	1	0
c. Lead whole group discussions	4	3	2	1	0
d. Listen to or observe teacher presentations	4	3	2	1	0
e. Use hands-on materials or objects	4	3	2	1	0
f. Complete a worksheet or workbook emphasizing routine practice	4	3	2	1	0
g. Use a textbook	4	3	2	1	0
h. Engage in discussion primarily with the teacher	4	3	2	1	0
i. Use school computers for writing	4	3	2	1	0
j. Use supplementary printed materials other than textbooks	4	3	2	1	0
k. Engage in discussion primarily with other students	4	3	2	1	0
l. Respond orally to open-ended questions	4	3	2	1	0

. Indicate the frequency with which STUDENTS did the following in your designated class during the last semester. Circle one response on each line.

	Almost every day	Once or twice a week	Once or twice a month	Once or twice a semester	Never
a. Explained how what they learned in class related to the real world	4	3	2	1	0
b. Worked individually on projects or preventions	4	3	2	1	0
c. Worked on projects that required at least one week to complete	4	3	2	1	0
d. Evaluated and improved their own work	4	3	2	1	0
e. Worked on problems for which there were several appropriate answers	4	3	2	1	0
f. Worked on problems for which there were several appropriate methods of solution	4	3	2	1	0
g. Worked as part of a group on projects or presentations to earn individual grades	4	3	2	1	0
h. Evaluated the work of other students	4	3	2	1	0
i. Worked as part of a group on projects or presentations to earn a group grade	4	3	2	1	0
j. Put events or things in order and explained why they were organized that way	4	3	2	1	0
k. Discussed with the whole class solutions developed in small groups	4	3	2	1	0
l. Conferenced with other students about their work	4	3	2	1	0

8. Over the last semester, how often did you emphasize the following with these students? Circle one response on each line.

	Almost every day	Once or twice a week	Once or twice a month	Once or twice a semester	Never
a. Generalizing from patterns or examples	4	3	2	1	0
b. Analyzing and interpreting information	4	3	2	1	0
c. Organizing, summarizing, or displaying information	4	3	2	1	0

9. During a class discussion if a student gave an incorrect response how frequently did you do each of the following? Circle one response on each line.

	Always	Often	Some-times	Rarely	Never
a. Call on other students to get their responses and then discuss what is correct	4	3	2	1	0
b. Ask the student another question to help him or her get the correct response	4	3	2	1	0
c. Call on another student likely to give the correct response	4	3	2	1	0
d. Provide the correct response yourself	4	3	2	1	0

10. The following is a list of ACTIVITIES TO COMPLETE AT HOME or homework you might have assigned your students. Although the list is not exhaustive, most activities could be considered variations of those listed below. For each type described below, indicate the frequency with which you assigned each over the last semester. Mark "never" for activities you did not assign during the last semester. Circle one response on each line.

	Almost every day	Once or twice a week	Once or twice a month	Once or twice a semes	Never
a. Write a journal entry	4	3	2	1	0
b. Prepare a written report	4	3	2	1	0
c. Work on problems for which there is no obvious method of solution	4	3	2	1	0
d. Read the textbook or other assigned reading	4	3	2	1	0
e. Apply concepts or principles to different or unfamiliar situations	4	3	2	1	0
f. Read supplementary material	4	3	2	1	0
g. Complete routine exercises or problems from worksheet, workbook, or text	4	3	2	1	0
h. Work on a project, gather data, conduct an experiment	4	3	2	1	0
i. Prepare an oral report	4	3	2	1	0
j. Complete a short writing assignment	4	3	2	1	0

11. When students were assigned written homework or activities to complete at home, how often did YOU do each of the following? Mark as "never" activities you did not assign during last semester. Circle one response on each line.

	Always	Often	Sometimes	Rarely	Never
a. Record only whether assignment was completed	4	3	2	1	0
b. Collect, correct, and keep assignments	4	3	2	1	0
c. Collect, correct, and return assignments to students	4	3	2	1	0
d. Have students exchange assignments and correct them in class	4	3	2	1	0
e. Have students correct their own assignments in class	4	3	2	1	0
f. Use assignment as a basis for class discussion	4	3	2	1	0
g. Use assignment as a basis for grading students	4	3	2	1	0
h. Use assignment as a basis for lesson planning	4	3	2	1	0

12. Estimate the amount of time in minutes an average student in your class spent doing homework or activities you assigned students to complete at home during an average WEEK.

0 minutes ____

1 - 30 minutes ____

31 - 60 minutes ____

61 - 90 minutes ____

91 -120 minutes ____

121 - 150 minutes ____

151 - 180 minutes ____

181 - 210 minutes ____

210 - 240 minutes ____

13. How often do you use assessment information for the following purposes? Circle one response on each line.

	Always	Often	Some-times	Rarely	Never
a. Determining student grades or ocher formal progress reports	4	3	2	1	0
b. Providing feedback to students	4	3	2	1	0
c. Diagnosing student learning problems	4	3	2	1	0
d. Reporting to parents	4	3	2	1	0
e. Assigning students to different programs or tracks	4	3	2	1	0
f. Planning for future lessons	4	3	2	1	0

14. In determining student grades or other formal progress reports, indicate the importance you gave to each of the following. Circle one response on each line.

	Extremely important	Very important	Somewhat important	Not important
a. Effort	3	2	1	0
b. Individual improvement or progress over past performance	3	2	1	0
c. Absolute level or achievement	3	2	1	0
d. Achievement relative to the rest of the class	3	2	1	0
e. Class participation	3	2	1	0
f. Regular completion of homework assignments	3	2	1	0
g. Consistent attendance	3	2	1	0
h. Results of standardized tests produced outside the school	3	2	1	0
i. Results of tests with open-ended items	3	2	1	0
j. Results of tests with multiple choice or true-fake items made by you or other teachers	3	2	1	0
k. Performance on projects or practical exercises	3	2	1	0
l. Your own observations of students	3	2	1	0
m. Items collected in student portfolios	3	2	1	0

The following questions ask about your classroom use of student PORTFOLIOS. Portfolios are a collection of student-generated artifacts that provide evidence over the semester or year about the range and extent of individual student performance and growth. Please answer the following questions about portfolio use last semester or grading period with your designated class.

15. In what content areas were PORTFOLIOS used with your designated class? Check all that apply. If portfolios were not used with your designated class, check "none".

English/language arts	____	music	____
math	____	home economics	____
reading	____	foreign language	____
social studies	____	other	____
science	____	none	____
art	____		

16. What types of student work were included in portfolios? Check all that apply. If portfolios were not used with your desireated class, check "none".

worksheets	____	self-reflective writing	____
open-ended problems	____	narrative writing	____
exploratory investigations	____	audio or video examples	____
long-term projects	____	group work	____
inter-disciplinary problems	____	independent work	____
journal entries	____	none	____
regularly assigned homework	____		

17. How were decisions made about the types of items that went into a student's portfolio? Selecting from those options listed below, indicate the source of directives and suggestions that guided the selection process for portfolios in your designated class. Mark as "n/a" those that do not apply. Mark one response on each line.

	Directive	Suggestion	n/a
a. State administration	____	____	____
b. State committee or task force	____	____	____
c. District staff	____	____	____
d. District committee or task force	____	____	____
e. School administration	____	____	____
f. School committee or task force	____	____	____
g. Classroom teacher	____	____	____
h. Other	____	____	____

18. How often did you use student portfolios in your designated class last semester for the following purposes? Circle one response on each line.

	Almost every day	Once or twice a week	Once or twice a	Once or twice a [illegible]	Never
a. Training students to reflect upon and/or assess **each piece** of work	4	3	2	1	0
b. Training students to reflect upon and/or assess their **overall** progress	4	3	2	1	0
c. Communicating student progress to parents	4	3	2	1	0
d. Determining student grades or other formal progress reports	4	3	2	1	0
e. Planning for future lessons	4	3	2	1	0
f. Diagnosing student learning problems	4	3	2	1	0
g. Informing decisions about student placement	4	3	2	1	0
h. Informing decisions about student graduation	4	3	2	1	0
i. Providing information for program or school accountability	4	3	2	1	0

19. Listed below are statements about portfolio use in the classroom. For your designated class last semester, please indicate whether you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with each statement. Circle one response on each line.

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. Criteria about types of work to be included or excluded in the portfolio were explicitly defined and were known by students	3	2	1	0
b. Criteria and process for evaluating work in the portfolio was explicitly defined and was known by students	3	2	1	0
c. Process to encourage students to reflect upon and revise work included in portfolio was explicitly defined and was known by students	3	2	1	0
d. Process to encourage student and teacher to work collaboratively on portfolios was explicitly defined and was known by students	3	2	1	0
e. Process to identify the amount and type of support student receives in completing each piece was explicitly defined and was known by students	3	2	1	0

Please answer the following questions while thinking of a SPECIFIC LESSON you conducted with your designated class last semester. Mentally select a recent class period that you can retail in some detail and that was fairly typical of what normally occurs in your classroom—i.e., a class period that was not affected by special events such as assemblies, guests, or any other unusual circumstances.

20. Circle the one option below that best describes the content of that specific lesson.

English/language arts	health
math	home economics
reading	industrial arts
social studies	vocational education
science	foreign language
art	business
music	physical education
ESL	other
integrated content	

21. Check the one option below that best describes the stage of instruction of that specific lesson.

review of topic taught in the past	___
introduction of new topic	___
continuation of a previous lesson on a recently introduced topic	___
concluding coverage of this topic	___

22. The following presents a list of activities that might have occurred during that SPECIFIC lesson. Although the list is not exhaustive of what is done in a classroom, most activities could be considered variations of those listed below. Using this list, indicate how that specific lesson developed. In the blanks,

a) write in the order in which the activities you used in the lesson took place (1 = first, 2 = second, and so on), and

b) estimate the amount of time you spent on each one.

Mark as "not applicable" (n/a) activities you did not use.

	Order	Minutes	n/a
• Review of previous lesson(s)	______	______	___
• Review or correction of previous homework	______	______	___
• Introduction or the lesson topic	______	______	___
• Development of the topic (extending depth and coverage)	______	______	___
• Summary of main points of the day's lesson	______	______	___
• Assignment of student homework	______	______	___
• Work on homework in class	______	______	___

23. Indicate the professional activities related to your present assignment in which you participated during the most recent typical (i.e., not unusual) WEEK by noting the amount of time in hours spent on each (0 = did not do) and whether that time was too much, adequate, or too little for your needs.

	Hours	Too much	Adequate	Too little
a. reading professional materials (e.g., journals)	____	____	____	____
b. lesson planning with colleagues	____	____	____	____
c. other professional conversation with colleagues	____	____	____	____

24. Indicate the professional activities related to your present assignment in which you participated during the most recent typical (i.e., not unusual) SEMESTER by noting the amount of time in hours spent on each (0 = did not do) and whether that time was too much, adequate, or too little for your needs.

	Hours	Too much	Adequate	Too little
a. long-range curriculum development planning with colleagues	____	____	____	____
b. your instruction was observed (for pur~~poses~~ other than formal evaluation)		____	____	____
c. you observed someone else's instruction (for purposes other than formal evaluation)		____	____	____
d. conference or workshop	____	____	____	____
e. on-going inservice or university course	____	____	____	____

Listing of NCES Working Papers to Date

Number	Title	Contact
94-01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05	Cost-of-Education Differentials Across the States	William Fowler
94-06	Six Papers on Teachers from the 1990-91 SASS and Other Related Surveys	Dan Kasprzyk
94-07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrel Kindel
95-01	Schools and Staffing Survey: 1994 papers presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk

Listing of NCES Working Papers to Date (Continued)

Number	Title	Contact
95-04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NILS: 88 Seniors	Jeffrey Owings
95-06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08	CCD Adjustments to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph