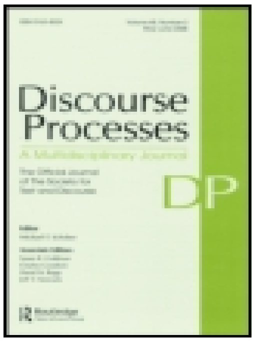

Discourse Processes

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/hdsp20

Locating tasks in psychology and education

Peg Griffin , Michael Cole & Denis Newman

DISCOURSE PROCESSES, 5, 111-125 (1982)

Locating Tasks in Psychology and Education*

PEG GRIFFIN, MICHAEL COLE AND DENIS NEWMAN
Laboratory of Comparative Human Cognition D003
University of California, San Diego
La Jolla CA 92093

The issue is raised: How does performance on one occasion relate to performance on another occasion? This is a problem both for practitioners and researchers working in classrooms and for psychologists concerned with the relation between results obtained on laboratory tasks and everyday performances. Naturally occurring events lack the constraints of the laboratory, making it difficult to compare one event to the next or to evaluate performance across events on a standard scale. In the study reported, attempts were made to instantiate the same task in a series of classroom lessons and activities. Two "process" coding schemes were tried out as ways of evaluating performance. The strengths and weaknesses of the Stallings Five Minute Observation system and Blank's system for coding cognitive demand are discussed with respect to specifying cognitive performance across situations.

The work we will describe in this article[1] is part of a general effort by members of the Laboratory of Comparative Human Cognition to develop an overall theory of the way in which the cultural organization of experience influences cognitive behavior. In this particular study, we are interested in how different ways of structuring classroom events influence the cognitive activities and learning achievement of different children.

From the point of view of the practicing teacher, the phenomenon we see as basic to our inquiry is the frequent intuition that a child knows more, is more capable, than she/he shows in a given evaluative context. In an example from our videotapes of the activities of third and fourth graders, there is a child who seems to know a great deal about the social organization of Native Americans. He volunteers relevant information, answers questions effortlessly. He "knows it all." But when a seemingly trivial task requiring that the child fill in an incomplete chart containing the information he has just discussed is presented, the child fails

*We wish to thank the classroom participants in our study and our colleagues: J. Bagneris, P. Bengel, R. Brown, C. Drale, K. Fiegener, M. Gearhart, L. Hirsch, J. Ideone, J. Levin, H. Mehan, R. Olascoaga, B. Pepper, R. Souvinney, K. Trauptmann, J. Wellenkamp.

[1]This is abstracted and condensed from a paper prepared for the National Institute of Education, sponsors of our current research. (NIE G 78 0159)

effortlessly to demonstrate what "we know he knows." There is no scientific apparatus to account for such observations. It is to provide such an apparatus, in a sufficiently explicit and detailed form that it can be applied to a variety of issues that grow from this familiar observation, that we have undertaken our current research.

The practicing teacher's dilemma is a version of a problem that faces research in a variety of areas. The most general statement of the issue is: How does performance on one occasion relate to performance on another occassion?

ANOTHER LOCATION OF THE PROBLEM

The general issues can be seen in studies considering the cognitive consequences of education, including investigations we have undertaken with colleagues in the Yucatan (Cole, Gay, Glick, & Sharp, 1971) and in Liberia (Cole, Sharp, & Lave, 1976). By and large, and as long as we stay with evidence from commonly used tests of cognitive behavior, the findings are that schooling has produced a very significant transformation in the way people think.

There are good reasons, however, to be suspect of the face value of this evidence. On logical grounds alone, we might be concerned that all we have done is to show that special practice produces specialized learning. That is, the commonly used tests of cognitive behavior have a special relationship to what people do during their schooling; thus, conclusions drawn on the basis of such tests are problematic. The issue can be posed best by stepping outside of the school to consider a test for the consequences of career training. Cole, Sharp, and Lave (1976) provide the following example and discussion:

> Suppose, for example, that we wanted to assess the consequences of learning to be a carpenter. Sawing and hammering are instances of sensorimotor coordination. Learning to measure, to mitre corners, and to build vertical walls requires mastery of a host of intellectual skills which must be coordinated with each other and with sensorimotor skills to produce a useful product (we are sensitive to this example owing to our own lack of success as carpenters!) To be sure, we would be willing to certify a master carpenter as someone who had mastered carpentering skills, but how strong would be our claim for the generality of this outcome? Would we want to predict that the measurement and motor skills learned by the carpenter make him a skilled electrician or a ballet dancer, let alone a person with 'more highly developed' sensorimotor and measurement skills?
>
> Lest it be thought that the example is too absurd to merit juxtaposition with the outcome of schooling, consider psychological experiments in light of the contexts from which their procedures have been derived and the domains in which they are routinely applied.
>
> Some version of virtually every experimental task reported in this monograph can be found in Alfred Binet's early work on the development of behavior samples which would predict children's success in school. The inspiration for their content came from an examination of the school curriculum, combined with Binet's sage guesses about the fundamental principles that underlie success in mastering that curriculum. The correlation between successful performance on Binet's tasks and success in school was a tau-

> tology; the items were picked because they discriminated between children at various levels of academic achievement. Might we not be witnessing the converse of that process when we observe people with educational experience excelling in experimental tasks whose form and content are like those they have learned to master in school? Is there any difference in principle between their excellence in recalling word lists, and the master carpenter's ability to drive in nails quickly? After all, practice makes perfect; if we test people on problems for which they have lots of practice, why should we be surprised when they demonstrate their competence? Conversely, what leads us to conclude that they will be equivalently good at solving problems for which they have no specific practice? (p. 227)

This work reveals as a problem the close tie between the strategies available for psychological studies and the tasks embedded in educational curricula. It also makes clear the need for developing new strategies in order to investigate psychologically interesting phenomena related to education. Specifically, we need to be able to locate and study behavior in tasks other than those found in standard experimental studies, in order to understand whether the performance difference noted in studies of the cognitive consequences of education is anything other than a function of the school populations' prior exposure to test tasks in school.

In order to understand on a more basic level how performance in culturally organized educational settings is related to an individual's cognitive behavior in other settings, or in general, we also need strategies to determine (1) how schools do the reorganization of thinking that they seem to do, (2) how the competencies promoted by schools are related to the competencies demanded by other parts of life, and (3) whether students transfer learning from the school to the non-school setting.

All of these questions presuppose for their answer knowledge of how school- and non-school-like tasks come about, inside and outside of schools. However, none of these methodological pre-requisites exist currently in the social sciences, from which it follows that we are not in a position to make professional statements about the effects of education on human thinking, let alone the impact of different kinds of schooling on different children.

In an effort to discover how to answer questions raised by our cross-cultural work, we began a series of investigations of U.S. children. In an initial study, we looked at a group of children in a range of activities: formal tests, various kinds of school activities, and after-school clubs (Cole, Hood, & McDermott, 1978). How and why our current study differs from this first attempt is basically the story of this article.

That first time, we looked for several cognitive activities that we believed would occur in the everyday world as well as in school: remembering, noticing similarities, reasoning, and so on. In the tests, we found these cognitive activities to be dense and visible; in the classroom, we found them to be scattered but more or less visible; but the club was different: Except on rare occasions, it was very difficult to identify any of the cognitive tasks that we had posed for the children in their testing situations and seen during our observations of the classroom. Some-how, cakes were getting baked, plants grown, rat mazes constructed, and electric circuits lit without anyone doing anything that a cognitive psychologist could recognize as thinking. On the rare occasions when we thought that a cognitive task-like problem had occurred, we found it virtually impossible to specify how a particular child had responded.

Perhaps our difficulties should not have come as a surprise to us, but in terms of our goal of building a technology whereby cognitive tasks could be discovered and their sequalia studied in non-classroom and non-test situations, we were (and remain) in deep difficulty.

Crudely speaking, our data indicated to us that the source of the difficulty resides in the social constraints operating on people during the conduct of a problem. The psychologist's task (classifying, paired associate learning, logical reasoning) is not a physical object in the world. It is, rather, a set of activities (perhaps involving physical objects) the goal of which is specified by the psychologist, along with a set of constraints that must be honored in meeting that goal. One of our difficulties when moving from club to school to test was that the larger social context within which "the same task" was embedded placed very different constraints on the individuals participating in the scene. As a consequence, the individuals were more or less free to change the conditions of the task, even to the point of making it go away, depending upon what social context it occurred in.

A second problem concerned the specification of goals. Even casual analysis of a single testing situation quickly reveals that an enormous amount of "social work" goes into maintaining the psychologist's task as a focus of attention. Subjects often are as anxious to demonstrate their friendliness or intelligence, or simply to get-it-all-over-as-quickly-as-possible, as they are to "think hard." Test situations are designed to minimize the impact of these alternative goals, of course, and large groups of subjects are usually run on quantifiable tasks so that "valid" inferences regarding thinking can be achieved.[2] What is crucial to point out is that in non-test settings including the school and the club, the multiple goals that occupy an individual at any single point in time are very difficult to ignore, because the settings are rarely constrained to the extent that they keep people from working to achieve several goals at the same time. That means that we have some difficult problems of "task analysis" to deal with in order to specify *real* task similarity across contexts. And without task similarity we cannot get far with an investigation of how performances on different occasions are related.

---

[2]We don't wish to question the validity of this approach (See Cole, Hood, & McDermott, 1978, for an extended discussion). Of course it is possible to claim, often with great justification, that the psychologist has been unsuccessful in creating a properly constrained model task environment and that the subjects are *not* simply trying to achieve the goal specified for them. Elsewhere we have developed the implications of this critique (Cole, Hood, & McDermott, 1978) for experimental psychology. Here we want to point out that even with maximal constraints erected to permit flawless inference about "intent," the flaws remain.

ASSISTANCE FROM PSYCHOLOGICAL RESEARCH

A useful starting point for thinking about making different cognitive tasks occur in a range of different contexts is to consider the procedures that psychologists use to maximize the probability that the *same task* will re-occur in the *same context*. The key idea goes under the label of standardization. Materials are pretested so that the subjects can plausibly be said to be working at a uniform level of difficulty; instructions are given in a standardized manner; restrictions specifying what the subject may *not* do are emphasized; the time taken to deal with any part of the materials is fixed; scoring procedures are rigidly adhered to so that only "relevant" parts of the subject protocol are included.

Even under these circumstances, all psychologists recognize that the same task is never repeated in all of its details. Instructions are sometimes garbled; subjects ask questions for which no standard answer exists; a subject with a cold keeps taking time out to blow his nose. To accommodate this recognized variability, the psychologist works with a model of "the same task" that permits him/her to proceed with the work. The model assumes that the variability in what happens from one experimental/test session to another is randomly distributed with respect to the essentials of the task. This assumption is built directly into the statistical tests that are used to evaluate psychological tests; these statistical tests include the assumption of random error by making each subject's score consist of two components—the true score and an error term. Discussions of test reliability looked at in this way reduce to tests of the size of the random error component relative to the true effect. It is also important to note that this model of standardization relies on testing relatively large numbers of subjects so that the real effects can be distinguished from error.

This model is not without its critics. Although the argument is made in a variety of ways, the basic point boils down to the contention that the error term in the standard model is not random. For example, Cicourel and his students (Cicourel et al., 1974; Mehan, 1973) show that experimenter/testers systematically provide information for some subjects that they do not provide for others, thereby inadvertently changing the difficulty of the task. When this observation is combined with the work of Labov (1970), who makes a convincing argument that some subjects view their task in some standardized tests as that of self defense against an antagonistic adult (Labov, 1970) or a variety of demonstrations that the contents of tests are subtly non-equivalent for different subject populations (see Houts, 1977, for a summary of such criticisms), one comes to appreciate that the sense in which many psychological experiments and tests represent instances of making the same task happen over and over again is a very technical sense indeed.

It is also essential to consider and to make explicit the basic procedures by which cognitive psychologists make plausible their claims that a particular task *has* occurred in the first place. Cognitive tasks don't just happen; they are made to happen. Speaking schematically, the psychologist creates an environment for action and observes the actions that follow as they relate to the hypothetical structure of the environment-as-constructed. Psychological tasks constructed in this way are virtually never one-shot affairs. Rather, the psychologist does a good deal of "pilot" work. This is the part of the study where the experimenter's intuitions about the task that he or she has constructed are tried out. The psychologist looks to see if the environment-as-constructed seems to be the environment-as-responded to. It is crucial that the subjects be responding to the stimuli (including the instructions) in the manner prescribed, and that the subject not engage in behaviors that are considered to fall outside the limits of the task. In effect, the psychologist creates a model system and studies behavior within it. The goal of theorizing is to account for as many details of the subject's behavior within the model environment as possible. In this approach, the psychologist's theory is simultaneously a theory of what the task is, what the relevant behaviors are, and the relation between elements of the tasks and elements of behavior. As a general rule, the psychologist's theory of the task-behavior interaction he or she has set up will prove faulty in one or more of its details. This leads to the construction of a new task environment that differs in some principled way from the first, but is similar to it in many respects. The new task presents the subject with somewhat modified conditions for action, and again the psychologist sees how well his theory can account for the pattern of behavior-environment relations that results. Excellent accounts of this research process are to be found in Estes (1975–79).

Two major characteristics of cognitive psychological research can be abstracted from the forgoing discussion. First, cognitive tasks are constructed, not discovered. Their construction involves the design of a functional system (Luria, 1979) which provides for the structure of micro-environments. Within this system, subjects are constrained in a variety of explicit and implicit ways. These constraints are essential to the analysis because only when they are met to a reasonable degree can we think it plausible to conclude that we have identified the task that the subject is engaged in. Second, the procedures for constructing a task also yield a set of rules about what counts as relevant behavior for analysis. In effect, the psychologist has constructed a coding scheme for behavior. It includes many subcategories of relevant behaviors and a catch-all category called error.

In order to tie classroom observations to laboratory-generated cognitive psychology in anything other than a casual way, we must find ways of specifying tasks and establish the relevant behaviors in a manner that can yield the same kinds of statements about task-behavior interactions. Insofar as we fail in this enterprise, we are subject to virtually unlimited uncertainty about the validity of our claims concerning such matters as cognition, transfer, and learning.

Our current work with third and fourth grade children grew out of the earlier observations of children in tests, classrooms, and clubs. But now our collection of data is designed so that we can construct something like the psychologists' model systems within which we can study the children's behaviors in interaction with various environments. What we have done is introduce a "tracer" element, in the form of a topic or a problem, that confronts the children in lessons, and in peer work situations, and in tutorials, and in clubs. The topic or problem can serve as a trace of the task being searched for, as it appears in different guises, under different constraints, in different settings, and as it evokes different behaviors from varying participants. We can locate recurring situations where a goal can be isolated (that is, we can identify it via its relation to the tracer) and the interactions of that goal with other co-occurring goals can be studied. By having a tracer element, we have a clearer chance to see what is varying; we can see how the researcher/teacher/club leader's plans concerning the task are transformed to create the task or tasks that the participants perform. Some of the problems revealed by our earlier attempt to look at children seem to be resolvable using this technique.

ASSISTANCE FROM EDUCATIONAL RESEARCH

If our tracers work, then we will have engineered the appearance of the "same" behavior (or at least "same in some respects") in a variety of differently organized events. But what we will need, in order to see if we have been successful, is something corresponding to the other aspect of the cognitive psychologist's work: that is, we need a way to identify what is to be counted as relevant for analyses of how performance on one occasion (in one setting/by one person) is related to performance on another occasion. The construction of the highly constrained task settings produces, in effect, a coding system for the psychologist; we need the same effect in order to analyze the behaviors in our less constrained and more varied task settings. In educational research, there is a history of concern with this issue. The *process-product* paradigm in educational research has developed as a way to investigate how the conduct of teaching sessions (i.e., the processes involved in education) are related to what the students end up knowing (i.e., the products of education).

Some problems in the process-product paradigm have been discussed in detail elsewhere (Koehler, 1979; Soar & Soar, 1976; Fenstermacher, 1978; Borich et al., 1978). Several inter-related problems arise in connection with the product measure, the test. First are the difficulties of the sort we have already referred to: The non-random errors that can be argued to occur in psychological experiments can also be found in tests; the close relationship of tests and school tasks leaves us with the problems we encounter in our work on the cognitive consequences of education.

Another problem, central to our general line of work, is that the reliance on product measures often includes an implication that what is measured (cognitive level, academic skill) is an entity apart from its use. An analogy is in order: a tape measure can be used to determine whether two pieces of wire are of equal length. Say that one piece of wire is hair-thin tungsten, the other coiled nichrome. While there is a standard somewhere, a theory of tape measures and of the materials tape measures are made of, that will tell us that the measure is adequate or to what degree it might vary over occasions of use, this theory may not be sufficient for all purposes. In fact one must decide what length means for tungsten and for nichrome (how much stretching or uncoiling is valid) in order to use the measure at all, and one must decide this in relation to the use to which the lengths of wire will be put. Length, like cognitive development or academic skills, is as much the product of an occasion of use of a measure as it is of the measurement's measuring theory; and length, like cognitive development or academic skill, is related to the use to which it is put outside of the measuring environment. A product measure used in educational research may in fact vary with respect to how generalizable it is to other contexts where the skills measured might be used and may also vary related to differences among the children. For some children, it may elicit the highest performance they have ever had available to them; for others, it may elicit the lowest.

A final problem with the process-product research, from our point of view, is that the paradigm assumes that the children's cognitive or academic skills cannot be measured *in situ;* but this assumption flies in the face of what the classroom teacher does everyday. (Cf. Mehan, 1979; Cazden, 1977; Griffin & Humphrey, 1978). In effect, process-product studies write off the chance to answer teachers' need to know how to work with the evaluation aspects of daily instructional encounters: to what degree, and how, can you find out where a child is (and gets to be) during the day or a unit; what kinds of inferences about children's capabilities in what situations are warranted, and what kinds of situational variation in displays of competence is one likely to encounter?

In spite of these uncertainties, our research has much to gain from an examination of work carried out in the process-product paradigm. The strengths of the paradigm include: (1) reliance on observable behavior in naturalistic settings as data, i.e., non-test tasks, are considered (2) treatment of at least some of the environmental variations (e.g., teacher behaviors, time on task) related to children's behaviors, and (3) an interest in differentially evaluating performances by various children and/or from various settings. A great many different kinds of process measures have been developed that reflect these strengths. Process measures characterize what goes on in educational settings, often by categorizing the kinds of questions teachers ask and the kinds of responses children provide.[3]

VIEWS OF A SCENE

We are interested in seeing how much further we can stretch the coding systems that serve as process measures, to see if they can meet our needs. With our "tracers" we have a chance to locate tasks (and "same" tasks in differently organized settings). Our question is whether existing coding systems can help us locate dif-

[3]Elsewhere (D'Andrade, 1974; Borich et al., 1978; Griffin & Mehan, 1979) there are extensive discussions of the problems that arise in many coding systems of these types.

ferences in the achievement of children, and among children, and perhaps differences in the tasks the children encounter as the tasks that we present to the children are operated on by the varying contextual constraints. In short, we want to know if existing measures can (or can be adapted to) show when children appear to be more and less smart, as well as when their work appears to be easier, and when harder.

We have chosen two coding systems to illustrate the kinds of variations that can be noted and the kinds of problems encountered when the systems are used. Stallings' system, developed for the Follow Through evaluation project, is the first example (Stallings, 1973); the second is Blank's system, which has been used both for studies of children and studies of lesson appropriateness (Blank et al., 1978). Both systems have been applied to one of our lessons from a unit, the topic of which was cultural variations among six groups of Native Americans. The lesson we chose is a good candidate for this purpose because of the variations noticeable. It was conducted as a small group lesson with a teacher and five children. Within the lesson, the participants make overt mention of how much some of the children know, and they also notice negatively the performance of another child. Furthermore, the lesson has clear cut phases or episodes that serve as mini-contexts in which the "same" task can reoccur. We coded ten phases: Phase 1—getting organized; Phase 2—reviewing the unit without visual aids; Phase 3—reviewing by reading from a filled in chart; Phase 4—reviewing characteristics of *bands* and *tribes* covered in detail the day before; Phase 5—reviewing the concept of *states*, covered a week before; Phase 6—getting organized for role play regarding the characteristics of *states;* (Phase 6A—the role playing, we omit from consideration here); Phase 7—drawing conclusions about the characteristics of *states;* Phase 8—review by filling in a partially empty chart; Phase 9—discussion of various forms of social organization that occur in the children's daily lives; Phase 10—minimally supervised short matching item test. We will concentrate on what the two coding systems can reveal about two of the child participants in this lesson, Chuck and Angelica.

Stallings' Five Minute Observation system (Stallings, 1977), while clearly descended from Flanders' scale (Flanders, 1970), differs from it by taking individual children as the unit of analysis rather than the whole class. Each turn in the lesson discourse is coded separately. Speakers and addressees are coded for every utterance in *Who* and *To Whom* categories; a *How* category, capturing basically emotive aspects, is optionally coded and proved difficult for us to apply consistently to our video-taped data. Most important for our purposes are the *What* categories that classify each turn according to the kind of task that is involved. There are thirteen classifications altogether; three levels of questions are included.

> *Code 1* asks for a response free of argument or speculation. There is one expected, acceptable response that is to be carried out, verbally or non-verbally . . . [e.g.] "Draw a line"

> *Code 1Q* questions elicit the following responses: statements of preference, statements of fact, itemizing, classifying and definitions . . . [e.g.] "If you had two pears and three apples, what would you have five of?"
>
> *Code 2* questions encourage responses that require: interpreting ideas, cause and effect establishing relationships, making comparisons, reasoning, applying previously learned materials to a new situation and describing a process . . . [e.g.] "Tell me how an electric train works." (Stallings, 1977, p. 269)

The other *What* codes include differentiations among responses, non-responses, informative statements, and evaluative statements. We adapted Stallings system by creating an additional code to distinguish among correct, incorrect, and irrelevant turns. We also added a dimension to the cumulative scoring system so that we could notice variations occurring among the lesson's ten phases.

The coding provides pictures of Chuck and Angelica, the two children that we are focusing on, that are best described in terms of their similarities and differences. Stallings' code without our adaptation shows striking similarities between the children: neither asks open-ended questions or fails to respond; they respond about the same number of times; the teacher asks each of the children about the same number of questions, and only one of her questions addressed to each child is a higher level question (category 2). Most of the responses given by both children are to questions that are not specifically addressed to them. Angelica makes more requests and issues more evaluations than Chuck. The teacher accepts and praises Chuck 10 times, while only accepting and praising Angelica 5 times. By using our adaptations, we can display an interesting difference: Chuck is correct for 70% of his 34 responses, Angelica is correct for only 46% of her 37 responses.

By looking at the ten different phases in the lesson, we can see that Chuck gives more of his correct answers during the discussion, Phase 9, than anywhere else; Angelica gives as many to her peers in the mini-tests as she gives to the teacher in the discussion. During Phase 7, drawing conclusions, both children give five correct responses, but Angelica also gives five other responses to Chuck's one. In general, the modifications seem to work and Stallings' coding system seems to capture some of the situational variation as well as making a differentiation between a more competent Chuck and a less competent Angelica.

There is one clear problem, in principle, related to the use of a system like this that is peculiar to the nature of turn taking in a small group situation. As it happened, only one higher level "open question" was specifically addressed to each child. This should not be understood to mean that there was only one response by each child to a "higher" level question. In fact the preponderance of the teacher's questions in the lesson were Code 2. Notice further that one of the ways that small groups differ from large group lessons is in the potential for questions being on the floor without the answer-turn having been allocated by the teacher to a particular child. (Cf. Griffin & Humphrey, 1978; Mehan, 1979, for discussions of the strict turn allocation procedures that work in keeping large group lessons together.) In dyads, the addressee is specified automatically. But in small groups, the turn-allocation machinery is not called upon as regularly as it is in large groups and there is no automatic indication of who should answer questions as there is in tutorial dyads. On many occasions in small groups, children can self-select to answer. However, in a system like Stallings', the only way to derive that a child has answered a question of a particular type or at a particular level is to locate the child's name in the *To Whom* category for questions of that type. Hence, when questions are addressed to members of the group at large, and Chuck or Angelica answer, we have no way of noticing the level of the question they are responding to. The nature of small group communication processes and the nature of this kind of coding system make systematically unavailable for analysis a reliable assessment of how hard the questions were that the children were answering or failing to answer. Although the system is handy for on the spot coding, there is no adaptation that we can imagine that would overcome this difficulty in using the Stallings system for our purposes.

Blank's system, specifically concerned with the issue of how "hard" or "easy" the demands on the child are, can be expected to avoid such a problem (Blank, 1977; Blank et al., 1978). The system, designed for use with pre-school children's language, codes the speaker on two levels: one is the social role, e.g., teacher vs. child; the other is the conversational role, e.g., initiator of an exchange vs. responder. The initiator's utterances are also coded on two levels: First, a determination is made whether the utterance puts an explicit demand on the responder to respond; if so, it is an *Oblige,* if not it is a *Comment.* Second, each Oblige and Comment is coded for a level ranging from (in the order given) less to more abstract, viz: *Matching Perception* (Level 1); *Selective Analysis of Perception* (Level 2); *Reordering Perception* (Level 3), and *Reasoning about Perception* (Level 4). (See Blank et al., 1978, pp. 8–21.) Blank describes an underlying model of cognition and language that assumes that acquiring language is a matter of mapping from one representational system (the child's conceptual notions) into the language system. Thus, an account is provided for the ordering of the levels which "reflect increasing distance between the perceptual style with which the children view the world and the language that they apply to these perceptions." (Blank et al., 1978, p. 15) An additional hierarchy (from Fully Adequate to No Response) is provided for coding the utterances of the responder.

Our adaptations for using the system with older children followed from suggestions made by Blank. She speculated that no interesting differentiations would be shown between tasks at Level 1 and those at Level 2 for older children, and that the interesting differences would be between Level 3 and Level 4 tasks. We therefore decided to adapt the Level 1 and Level 2 codes to capture particular aspects of our videotaped lessons that seemed to require special treatment. Our first adaptation was to code as Level 1 any utterances to which an adequate response could be made based on what was available in writing or pictures at the time of the expected response, regardless of what level the utterance could have been assigned on other grounds. We suspected that the presence of these kinds of environmental supports in a third/fourth grade classroom should lower the difficulty level of the task. This tactic is reasonable, given the matching aspect of our special Level 1s, and the matching aspect of Blank's original Level 1 code. Our second adaptation was to distinguish as a special category those utterances that are related to the elements of the domain that had been drilled in the lesson just prior to the lesson being coded. Such utterances were coded at Level 2. Again, this adjustment seems defensible: these special 2s set up a demand for the selective analysis of the previously drilled materials analogous to the original level 2s. Our third adaptation was a direct recommendation by Blank, the establishment of an *Adequate plus* category for exceptionally good, relevantly elaborated responses.

This coding system shows few similarities between Chuck and Angelica; the differences between the children are most striking. The teacher asks Chuck more questions than he asks Angelica: 86% of Chuck's codable units are questions from the teacher (22) while only 34% of Angelica's are questions from the teacher (15). Only 9% of Chuck's units are repetitions, while 28% of Angelica's are. Chuck's repetition in answer to a question occurs at level 4; Angelica repeats in responses at levels 1, 2, and 3. The only indication of Angelica being higher than Chuck is that she issues two level 4 comments while he issues none. In general, the picture of Angelica is one of a child less advanced than Chuck, and the fact that she performs adequately in response to only 47% of the obliges she engages in deepens the contrasts with Chuck, who is adequate 87% of the time.

However, there is an even more interesting contrast. Chuck's adequacy decreases gradually as he is asked higher level questions (from 100% at level 1 to 89% to 80% to 50% at level 4). Chuck is a model child for Blank's notion of levels. Angelica, on the other hand is almost directly opposite: at level 1, she gets only 33% correct, she climbs to 50% at level 2, to 43% at level 3, and is correct the one time she is asked a level 4 question. Angelica seems to behave contrary to Blank's expectations.

An examination of how this coding system operates in the different mini-contexts related to the phases of the lesson as described above shows an interesting relationship between Blank's level of difficulty and the phases in this lesson. Twice the phases progress in level of difficulty in a way that fits Blank's notion of the progression that should take place in lessons: The teacher asks only level 1 obliges in Phase III, only level 2 obliges in Phase IV, and only level 3 obliges in Phase V. A similar progression occurs after the getting organized phase: Phase VIII has only level 2 obliges, and Phase IX starts with level 3 and goes on to level 4 obliges. It seems that Blank's notion of levels redundantly specifies the phases of this lesson.

Overall, Blank's system seems to be an interesting base for our work and amenable to adaptations related to the "mini-contexts" of a lesson and to the specifics of the lesson topic. However, the problem with Angelica points to a major concern. Children like Chuck "fit" the assumptions underlying the work. Children like Angelica do not "fit" the assumptions. One of the phases, Phase 7, is also odd—it is not a part of the lesson's progressions in difficulty; in fact, it does not have a consistent level of difficulty like the other phases do. The system could allow us to draw a conclusion that the differences are quantitative (less developed child, inadequate teaching) rather than qualitative (Angelica and Phase 7 have a different, perhaps more complicated, relationship to level of difficulty and demonstration of strengths and weaknesses than the relationship that other people and/or situations have). The way that Angelica and Phase 7 diverge from the norm of Blank's system are related to the way they are inadequately accounted for in the theories (folk and formal) of education, cognition, and discourse that interact with the system in various ways. Angelica is from a Spanish speaking background and has been using English for only a few years. Phase 7 calls for some mixture of what might be called convergent and divergent thinking. The demands on theories to respond to these kinds of variations have not been met adequately by theories available to Blank, or to us—which brings us full circle to the general work of our group.

PROBLEMS REMAIN

There are two general problems with coding systems for educational talk and tasks that bear an interesting relationship to the problems that we encountered in our earlier attempt to locate cognitive tasks in the clubs. The first we shall call the point of control problem.

Most coding systems assume that there is a point of control standardly locatable, and that the category to which the control utterance is assigned affects the categorization or understanding of utterances which it controls. For example, most coding systems derive the cognitive level category of a child's response from the cognitive level category of the adult's question. We know enough about teachers' differential expectations of children and about the chance of these being evident in teachers' questioning behaviors (Cherry, 1978) to suspect that using the teacher's question as a way to describe the cognitive level of the child respondent will systematically distort the data. Some children will be pictured with inflated levels and others will be underestimated. We may have a better picture of the teacher's expectations than of the children's capabilities. While we do not dispute that the teacher and teacher questions have a lot to do with constraining the tasks that children perform in classrooms, (and that the academic or cognitive tasks that are our primary concern are very heavily influenced by the teacher), we must not ignore the facts that *more* than teacher questions can be involved in specifying the tasks the children undertake, and that teacher questions do other things beside specifying tasks.

We are here in the opposite corner from the one our earlier look at children in varied settings had painted us into. There we were concerned because the multiple goals in the settings made it too hard to see the kinds of cognitive tasks that could be seen in the engineered system of the experimental setting. We moved to a solution of that problem by specifying tracer elements that would let us highlight certain goals in our settings. Now, if we use coding systems that award control of our understanding of the child's performance to some preceding question, we will find ourselves assuming away the issue of multiple goals and how they interact. We have planted tasks so that the children's behaviors that we capture on video tape will be a little bit more comparable with the experimental subject's behaviors; the tasks are cultivated in the richly varied real world of the children; we cannot afford to have an analytic tool that works by looking for seeds.

The second general problem with many coding systems is related to the sequential nature of discourse. The best illustration of it is commonly available by noticing one of the best indicators of an easy task: children raising their hands, shouting for a turn. In many lessons, including many of ours, there is a point close to an end boundary where everyone wants a turn. The same question is a different question by virtue of its placement in the sequential development of the lesson. In one instance, as in the chart used in the lesson described above, there may be six slots to be filled and six fillers to use and so, of course, when five have been covered, the sixth is quite easy. One need only attend to what hasn't been said. In other circumstances, answering late may face the child with a more difficult task because all of the "easy" answers are used up. What has been said can make subsequent questions easier or harder or perhaps even different than earlier questions. Testers and teachers and peers all use the sequential nature of discourse events to be co-constructors of an answer with the person who appears to be the primary performer. This point of view is adequately argued and demonstrated in the work of classroom discourse analysts. But, in commonly used coding systems, an utterance is categorized uniformly whenever it occurs; changes in the constraints on it occasioned by the situation in which it is embedded are disregarded.

Once again, we return to the problems we found in our earlier attempt to locate cognitive tasks: the constraints on behavior differed so radically from situation to situation that we found it difficult to locate tasks. If the coding system we use in analysis assumes away the constraints imposed by the sequential nature of the discourse, then it may make it easier to locate tasks; but we have little faith that the tasks so located will be the tasks the participants were engaged in. We cannot afford to base our analysis on "same" tasks that are the same primarily by fiat of a coding system that ignores the influence of constraints operating on those tasks.

REFERENCES

Blank, M. Language, the Child and the Teacher: A Proposed Assessment Model. In H. L. Hom, Jr. & P. Robinson (Eds.), *Psychological Processes in Early Education*. New York: Academic Press, 1977.

Blank, M., Rose, S. A., & Berlin, L. J. *The Language of Learning: The Preschool Years*. New York: Grune & Stratton, Inc., 1978.

Borich, G. D., Maletz, D., & Kugle, C. L. Convergent and Disrciminant Validity of Five Classroom Observation Systems: Testing a Model. *Journal of Educational Psychology*, 1978 *70* (2), 119–128.

Cazden, C. B., Cox, M., Dickinson, D., Steinberg, Z., & Stone, C. "You all gonna hafta listen": Peer teaching in a primary classroom. In W. A. Collins (Ed.), *Children's Language and Communication*. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Cherry, L. J. A Sociolinguistic Approach to the Study of Teacher Expectations. *Discourse Processes*, 1978 *1*, 373–394.

Cicourel, A. V., Jennings, K., Jennings, S., Leiter, K., MacKay, R., Mehan, H., & Roth, D. *Language Use and School Performance*. New York: Academic Press, 1970.

Cole, M., Gay, J., Glick, J. A., & Sharp, D. W., *The Cultural Context of Learning and Thinking*. New York: Basic Books, 1971.

Cole, M., Hood, L., & McDermott, R. *Ecological Niche Picking: Ecological Invalidity as an Axiom of Experimental Cognitive Psychology*. Unpublished manuscript, University of California at San Diego & The Rockefeller University, 1978.

Cole, M., Sharp, D. W., & Lave, C. The Cognitive Consequences of Education: Some Empirical Evidence and Theoretical Misgivings. *The Urban Review*, 1976 *9* (4), 218–233.

Estes, W. K. *Handbook of Learning and Cognitive Processes* (Vols. 1–6). Hillsdale, NJ: Lawrence Erlbaum Associates, 1975–1979.

Fenstermacher, G. D. A Philosophical Consideration of Recent Research on Teacher Effectiveness. *Review of Research in Education*, 1978 *6*, 157–184.

Flanders, N. A. *Analyzing Teaching Behaviour*. Reading, MA: Addison-Wesley, 1970.

Griffin, P., & Humphrey, F. Task and Talk. In Shuy, Roger and Griffin (Eds.) *The Study of Children's Functional Language and Education in the Early Years*. New York, Arlington, VA: Center for Applied Linguistics, 1978.

Houts,. *The Myth of Measurability*. New York: Hart, 1977.

Koehler, V. *Methodology for Research on Teacher Training*. Paper prepared for the National Invitational Conference, "Exploring Issues in Teacher Education: Questions for Future Research". (Sponsored by the R&D Center for Teacher Education, the University of Texas at Austin, and the National Institute of Education, January 1979.)

Labov, W. The Logic of Non-Standard English. In F. Williams (Ed.), *Language and Poverty*. Chicago, IL: Markham, 1970.

Luria, A. R. *The making of mind: A personal account of Soviet psychology* (M. Cole & S. Cole, Eds.) Cambridge, MA: Harvard University Press, 1979.

Mehan, H. Assessing Children's Language Using Abilities. J. M. Armer & A. Grimshaw (Eds.), *Methodological Issues in Comparative Sociological Research*. New York: Wiley, 1973.

Mehan, H. *Learning Lessons*. Cambridge, MA: Harvard University Press, 1979.

Soar, R. S., & Soar, R. M. An attempt to identify measures of teacher effectiveness from four studies. *Journal of Teacher Education*, 1976 *27*, 261–267.

Stallings, J. A. *Follow Through Program Classroom Observation*. Unpublished manuscript, Sanford Research Institute, Menlo Park, CA, 1973.

Stallings, J. *Learning to Look*. Belmont, CA: Wadsworth Publishing Company, Inc., 1977.