The will to learn is an intrinsic motive, one that finds both its source and its reward in its own exercise. The will to learn becomes a "problem" only under specialized circumstances like those of a school, where a curriculum is set, students confined, and a path fixed. The problem exists not so much in learning itself, but in the fact that what the school imposes often fails to enlist the natural energies that sustain spontaneous learning—curiosity, a desire for competence, aspiration to emulate a model, and a deep-sensed commitment to the web of social reciprocity.

—Jerome Bruner, Toward a Theory of Instruction
CHAPTER 2

The Big Picture: Language and Learning in Hundreds of English Lessons

Martin Nystrand and Adam Gamoran

FEW GOOD MODELS currently exist for understanding how social processes affect student learning in classroom settings. There is widespread consensus that research must focus on teacher-student and peer interaction as it affects learning in order to be sensitive to the social context of learning. Yet most conceptions of instruction view learning as the result of what teachers plan and provide for students, that is, what teachers do to students. Adherents of this approach see instruction as a one-way transmission of knowledge from teacher and texts to students, and they typically assess students' knowledge for its congruence with curricular aims and objectives.

In our research in secondary school English classes, however, we had in mind not what teachers “do to students” but rather what teachers and their students do together, that is, what Michaels (1987) calls “the day-to-day practice of a curriculum” (p. 323). In this sense, teacher and students negotiate the actual curriculum—as opposed to the ideal or intended curriculum (for example, as written up in a curriculum guide). Superficially, this negotiation is visible in the give-and-take of classroom talk (Flanders, 1970). However, such ostensible interaction clearly is pedagogically less significant than the cognitive interaction that occurs—or does not occur, as the case may be—between teacher and students. When minds meet in this way, the result is a sequence of shared understandings of subject matter among members of the class, and the course of instruction, whether considered on any given day or examined over an entire school year, may be analyzed in terms of how classroom talk and activities modify and expand these understandings. In this sense, dialogic instruction is a negotiation of meaning by and between teacher and students.

We observed hundreds of eighth- and ninth-grade lessons over 2 years. In all our observations, we never found the “perfectly dialogic classroom”; indeed, such a perfect classroom probably does not exist in the real world. Rather, we did a large empirical study to examine the general effects of dialogic practices on achievement and learning. The scope of our study, under-

taken with a large sample of students in a large and diverse sample of classes, schools, and communities, enabled us to systematically test hypotheses about such practices.

This chapter reports details and findings from this study, sketching an overall portrait of classroom discourse in middle and high school English classes, and helping us address several important questions about classroom talk and student learning.

• What is instructional discourse generally like in eighth- and ninth-grade English and language arts classes?
• How much instruction is recitation? How much consists of discussion and small-group work?
• How much instruction is organized dialogically? Monologically?
• How does classroom discourse vary from middle to high school? By ability group? By subject? Among urban, suburban, and rural schools?
• How do these practices affect student learning about literature? Which interactions are appropriate and productive?

Our study, which shows the role that large-scale data analysis can play in the investigation of classroom interaction and its effects on learning, found that generally students learn more in classrooms organized more dialogically than monologically. For an overview of the entire study, see Figure 2.1.

STUDYING CLASSROOM DISCOURSE: DESIGN AND METHODS

In order to develop as comprehensive an understanding of classroom discourse as possible, our study coordinated three separate but related investigations:

1. Surveys and interviews. We sought first to determine practices and attitudes toward classroom discourse by surveying both students and teachers on various classroom practices; in addition, we interviewed the teachers to learn about their instructional methods and the context of instruction.

2. Class observations. Through direct observation, we investigated teachers' allocation of class time to various types of classroom discourse, including recitation, discussion, and small-group work. We collected data allowing us to gauge the quality of teacher-student interaction by focusing on the characteristics of questions asked by both teachers and students.

3. Hypothesis testing. Using statistical techniques, we examined the general effects of classroom practices and the organization of instruction on student achievement.
Figure 2.1A. Synopsis of Study—Design

The purpose of our research was to investigate the effects of instructional organization on student learning, contrasting the epistemologies of recitation and discussion. This work was conducted in a 2-year study (1987–1989) in 16 middle and junior high schools and 9 high schools in eight midwestern urban, suburban, and rural communities. Participating were 58 eighth-grade and 54 ninth-grade language arts and English classes, involving more than 1,100 students each year. Each class was observed four times, twice in the fall and twice in the spring, providing observational data for more than 200 lessons each year.

TRACKING

The study encompassed both middle and high school, following a subset of students as they moved into high school, in order to understand the mechanism and effects of placement and tracking in high school. The study was designed to provide systematic contrasts of instruction and learning in high- and low-track classes.

INSTRUCTIONAL DISCOURSE

Instructional discourse was studied in two ways. First, observers timed instructional activities in order to determine the allocation of class time to various activities, for example, question-answer, discussion, small-group work, seatwork, and other activities. In addition, observers recorded and coded both teacher and student questions for dimensions of dialogic instruction, including (1) authenticity (whether or not questions had "prespecified" answers), (2) uptake (incorporation of previous answers into subsequent questions), and (3) level of evaluation (extent to which the teacher allowed a student response to modify the topic of discourse). More than 23,000 questions were coded. The observational data were supplemented by teacher and student survey data and end-of-year teacher interviews.

LEARNING

Learning about literature was tested with a written examination in the spring based on several works of literature studied during the year. The test involved a set of increasingly more probing questions, ranging from naming and/or describing as many characters from each story as the student could remember, and explaining the ending of each story to briefly explaining the themes and conflicts of each story and relating theme, conflict, and ending. All students answered the same general questions, although the details of the tests varied depending on the titles studied and selected. For the ninth-grade test, students wrote a brief essay on some character from their readings whom they admired, and explained their admiration.

Figure 2.1B. Synopsis of Study—Results

OVERALL

Classroom discourse was overwhelmingly monologic. When teachers were not lecturing, students mainly were either answering questions or completing seatwork. The teacher asked nearly all the questions, few questions were authentic, and few teachers followed up student responses. On average, discussion lasted less than 50 seconds per class in eighth grade and less than 15 seconds per class in ninth grade. Small-group work in eighth grade took only about half a minute each day, and only a little more than 2 minutes a day in grade 9.

PROCEDURAL VARIABLES

The study found a modest effect for time spent on homework, no effect for asking questions in class, and a negative effect for level of activity during recitation.

DIALOGIC INSTRUCTION

Results provided support for dialogic instruction, indicating that time devoted to discussion, authentic questions, uptake, and high-level teacher evaluation had a strong positive effect on achievement. Discussion in grade 8 had a particularly large effect.

IMPORTANCE OF CONTENT

Results of the grade 9 study were consistent with grade 8 results but only when controlled for content. Discussion and authentic questions unrelated to literature had a negative effect on learning.

GROUPWORK

Groupwork was successful to the extent that teachers clearly defined goals and tasks at the same time that they encouraged students to generate conclusions, solve open-ended problems, and address authentic questions rather than simply manipulate information and answer study questions. Most small-group work in study classes was in fact "collaborative seatwork," however, which had a negative effect on learning.

TRACKING

Instruction was more fragmented, contrived, and monologic in low-track than in high-track classes. In grade 8, teachers lectured 40% more in low-track classes than in high-track ones, and low-track discussion time was only half that of high-track groups. In grade 9, seatwork was nearly four times more frequent in low-track than in high-track classes.
Table 2.1. Characteristics of School Sample

<table>
<thead>
<tr>
<th>School District Type</th>
<th>Total</th>
<th>Middle Schools</th>
<th>High Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parochial</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Public</td>
<td>17</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Small town/rural</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Suburban</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Urban</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Research Sites and Participants

Our study lasted 2 years: Eighth-grade classes were observed during 1987–88, ninth-grade classes during 1988–89. We collected data in eight midwestern communities, including rural, urban, and suburban sites, in both public and parochial schools. Six of these communities were public school districts; the other two were Catholic high schools with students from a number of urban and suburban K–8 feeder schools. Unlike ninth-grade classes, which were all called English, eighth-grade classes were variously called language arts, English, reading, communications, literature, and so on; we selected the eighth-grade classes that focused most on reading. Table 2.1 provides a breakdown of the community and school types that participated in our study.

In each school we observed four English classes. In the smaller schools, we observed all the English classes; in the larger schools, we selected a representative sample of different ability groups as defined by the school (honors or accelerated, regular or average, basic or remedial). In all we made 451 observations in 58 eighth-grade classes in 16 middle and junior high schools, and 54 ninth-grade classes in nine high schools (which were fed by the junior high and middle schools in our eighth-grade study). Between 1,100 and 1,200 students participated each year; of all eligible students, about 10% were lost through absence or refusal. About one-third of all students participated in both years of the study. Table 2.2 summarizes these data.

Observational Procedures

Each class was visited four times by a trained observer, twice during fall semester and twice during spring semester. Within these parameters, observations were scheduled at the mutual convenience of teachers and observers. On these occasions, the observer noted the time spent in different instructional activities, and recorded and coded all teacher and student questions for a set of variables contrasting monologic and dialogic instruction (coding is explained on pp. 37f).

Table 2.2. Scope of Study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Grade 8</th>
<th>Grade 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>1,041</td>
<td>1,100</td>
</tr>
<tr>
<td>Number of classes</td>
<td>58</td>
<td>54</td>
</tr>
<tr>
<td>Number of times each class observed</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of observations</td>
<td>227</td>
<td>224</td>
</tr>
<tr>
<td>Number of coded questions</td>
<td>12,033</td>
<td>11,043</td>
</tr>
</tbody>
</table>

Discourse Episodes and Segments

Data from each class session were organized according to episodes and segments. An episode was defined as a coherent classroom activity centering around a particular objective or purpose. A new episode was marked when the teacher addressed a new objective. Like the start of a new paragraph, each such shift usually was evident in the teacher’s initiation of a new topic. Usually episodes consisted of two or more activities. For example, in addressing a particular objective, a teacher might initiate a question-and-answer session that then would be interrupted by brief periodic lectures and culminate in a homework assignment. When something like this happened, we divided the episode into segments, defined as any coherent part of an episode that differed from other activities of the episode. Instructional activities were classified with durations in minutes and seconds for the following:

1. Classroom management activities
   Classroom procedures
   Directions
   Discipline
2. Direct instruction
   Lecture, film
   Question-answer
   Discussion
   Student presentations
   Students reading aloud
3. Seatwork
   Supervised with teacher helping
   Supervised with teacher monitoring
   Unsupervised
   Small-group work
4. Tests and quizzes

When the teacher did one thing (e.g., lecture) and some students were allowed to do another (e.g., when the teacher lectured to part of the class but
We defined discussion as the free exchange of information among students and/or between at least three students and the teacher that lasted at least a half minute. Typically discussions came about during question-answer exchanges when a student would volunteer an observation (rather than ask a question) that the teacher allowed to substitute for normal evaluation. These discussions, which interrupted or violated the normal initiation-response-evaluation (IRE) sequence of recitation, included few questions, and those that were asked typically clarified ideas and information (“By that do you mean...?”).

Questions

Bakhtin’s conception of discourse encompasses far more than just questions and the utterances immediately preceding and following them: As we saw in Chapter 1, Bakhtin’s chains of utterances also encompass frames of mind and core beliefs, including those related to students’ experience out of school. Dialogic analysis is most directly accomplished through close analysis of transcripts of individual lessons and other forms of qualitative analysis, especially of the sort Dyson undertakes in her Social Worlds of Children Learning to Write (1993). Although our study undertook some such analysis (see Chapter 3), the main focus of our research was a comprehensive analysis of classroom discourse, generally with a special focus on teacher and student questions. We did this for several reasons. First, we wanted to capture the general dimensions of instruction, requiring us to examine hundreds of instructional episodes and lessons. We focused on question-answer exchanges because they provide an effective method for such a general analysis, and because they are so central to instruction, occupying 30% of class time in the eighth-grade classes we studied and 42% in the ninth-grade classes. Question-answer exchanges between teachers and students clearly dominate instruction for most students. They play a key role in both accommodating and excluding student voices in the public, authoritative discourse of the classroom, and they are the central instructional mechanism in American classrooms for assigning epistemic roles to students. As such, they significantly regulate the extent to which teacher-student interaction can be dialogic.

Two features of questions were of particular interest to us: authenticity (whether or not teacher questions had “prespecified” answers) and uptake (incorporation of previous answers into subsequent questions). Each is a critical variable affecting the salience of student voices in classroom discourse, and reflecting how far the horizons of classroom discourse extend beyond the question-answer sequences themselves to draw on student experience central to student engagement. While these data cannot detail the dynamics of particular teacher-student interactions, they nonetheless provide a powerful index of the extent to which teachers open their classes to student voices when they ask the questions they do.

Questions are not everything, of course, and authentic questions, we found, do not invariably produce learning. Nonetheless, one must not underestimate the role teachers’ questions play in shaping the character of classroom discourse as it affects learning. Questions presume answers. As negotiations of sorts, question-answer sequences reveal important features of teacher-student interaction and hence the character of instruction. Much can be learned about teacher-student interaction and talk in a classroom by determining the source of questions, the extent of authenticity and uptake, the level of cognitive activity that questions elicit, and so on. Even the pace of a teacher’s questioning can be revealing: Carlsen (1991) cites studies finding that a slow pace of teacher questioning and extended wait times correlate with greater numbers of student responses (Honea, 1982), as well as more sustained student responses of greater complexity and higher-order thinking (Fagan, Hassler, & Szabl, 1981).

Early in our study, we learned that the vast proportion of teacher questions (a) are test questions, (b) get a response, (c) do not involve uptake, and (d) elicit a report of what is already known. Indeed, this is the very profile of monologic classroom discourse, and we soon began to describe such questions, unfortunately, as normal teacher questions. The following questions, all from Mr. Schmidt’s lesson described in Chapter 1, are examples:

- “According to the poet, what is the subject of The Iliad?”
- “Where does the action of the first part of Book I take place when we enter the story?”
- “What is the result of the quarrel between Agamemnon and Achilles?”

We used data about questions to build profiles of instruction and classroom discourse, coding more than 23,000 questions, and examining each question in the context of the whole lesson at the time it was asked. Ninth-grade classes were tape recorded. Whenever observers were uncertain about how to interpret classroom activities and code questions, they consulted with the teacher after class. Questions were coded for

- Source: Teacher or student
- Response: Yes or no
- Authenticity: Whether or not an answer was prespecified
- Uptake: Incorporation of a previous answer into a subsequent question
The type of cognitive demand made by the question

• Cognitive level: The type of cognitive demand made by the question

• Level of evaluation: Whether the teacher valorized and elaborated the students’ responses

Coding reliabilities were based on paired readings of a sample of questions.3 During the ninth-grade study, observers collected data using a specially written computer program, CLASS 2.0,4 during class. This program helped with question coding, as well as the allocation and timing of various instructional activities, which the observer recorded by selecting from a menu and then briefly described. Every 5 minutes during question–answer exchanges and every 2 minutes during seatwork and lecture, the program prompted the observer to record the number of students obviously off task,5 as well as the number actively participating, and to make adjustments in the number of students in case any recently had entered or left the classroom.

During question–answer exchanges, the observer typed in and coded the questions that teachers and students asked during instruction. As observers entered each question into computer memory, the program prompted them for codings. When data collection was completed, CLASS-EDIT 2.0, a companion program to CLASS 2.0, allowed proofreading, editing, and revising of each file for inappropriate codings, and then computed basic statistics for each episode. Generally, we were satisfied that these procedures were minimally intrusive to instruction during our observations.

Authenticity. Authentic questions are questions for which the asker has not presupposed an answer and include requests for information as well as open-ended questions with indeterminate answers. Dialogically, authentic teacher questions signal to students the teacher’s interest in what they think and know and not just whether they can report what someone else thinks or has said. Authentic questions invite students to contribute something new to the discussion that can change or modify it in some way.

By contrast, a test question allows students no control over the flow of the discussion. Because authentic questions allow an indeterminate number of acceptable answers and open the floor to students’ ideas, they work dialogically. By contrast, a test question allows only one possible right answer, and is hence monologic (in Lotman’s terms, univocal; see Chapter 1).

Before we started our observations, we worried that determining the authenticity of questions might be complicated and unreliable since such determination requires assessing teachers’ intentions; authenticity cannot be determined from words alone. For example, “Who won the World Series in 1928?” can be either a test question or an authentic question depending on (a) whether the asker knows the answer and wants to see if the person asked also knows, in which case it is a test question, or (b) whether the asker doesn’t know and wants to find out by asking someone who does know, in which case it is authentic.

The nature of the activity, that is, the genre of classroom discourse, we discovered, is the most reliable indicator of authenticity. Hence, when teachers began a lesson by saying, “Okay, class, let’s check the answers to your study questions,” we quickly learned that the questions were invariably test questions (although follow-up discussions of students’ answers were sometimes authentic). By contrast, when teachers asked about students’ personal experiences as lead-ins, for example, to open-ended discussions of a poem or short story, we found that these questions were authentic. In Chapter 1, Ms. Turner’s initial question, “Can you recall things from Huckleberry Finn that, um, seemed racist to you?” is an example. Questions asked during discussions, for example, the question Tom asks John in Ms. Lindsay’s class, “Is it Mr. Hollings’s store? Is that it?” are also authentic since their purpose is not testing someone’s knowledge but rather exchanging only that information the person asking the question actually needed to know.

Whenever the authenticity of a question was unclear or ambiguous to us, we consulted the teacher. In practice (and somewhat to our initial surprise), coding authenticity proved to be generally quite straightforward. We quickly learned that most classroom discourse is not subtle: The vast proportion of questions teachers ask are test questions, whereas student questions are virtually always authentic (except when students role play teachers; then they ask test questions!).

Uptake. Uptake occurs when one conversant, for example, a teacher, asks someone else, for example, a student, about something the other person said previously (Collins, 1982). Here is an example of uptake from a ninth-grade lesson on The Iliad: The teacher asks, “What do they have to do to Polyphemus?” A student replies, “Blind him.” The teacher then follows up, asking, “How come the plan is for blinding Cyclops?” Uptake occurs here when the teacher picks up on the student’s response, asking about “blinding” him. Uptake often is marked by the use of pronouns, for example, “How did it work?” “What caused it?” “What city grew out of this?” In each of these questions, the italicized pronoun refers to a previous answer. Uptake also may be characterized by ellipsis. In Ms. Turner’s class, for example, when Linda says that racism in The Adventures of Huckleberry Finn makes her “ashamed,” Ms. Turner’s reply, “In what way?” exhibits uptake since her question makes Linda’s answer the momentary topic of discourse. Teachers use uptake whenever they follow up on student responses. As an essential dialogic resource facilitating the negotiation of understandings, uptake plays a prominent role in discussion as conversants listen and respond appropriately to each other.
Cognitive Level. Our project also sought to assess whether the cognitive level of questioning affects student learning. In this way we examined whether instruction stressing higher-order thinking is necessarily dialogic. We therefore coded the level of cognitive functioning that each question sought to elicit, judging it high to the extent that the question could not be answered "through the routine application of previously learned knowledge" (Newmann, 1990, p. 44; see also Polanyi's (1958) distinction between routine performances and heuristic acts). Like authenticity, the cognitive level of questions cannot be judged altogether from words alone. For example, if the teacher expected students to answer questions by reciting information found in textbooks, we coded questions as reports regardless of their linguistic structure. Hence, although a why-question normally will elicit an analysis, it will elicit a report if the teacher's focus is the recitation of a textbook's analysis rather than the class' reflection; then "Why?" really means, "According to your text, why did it happen this way? Do you remember?" The results of our research found most classroom discourse to be overwhelmingly monologic. In this regard, our study replicates numerous previous analyses. We coded questions as records if they elicited descriptions of what students were observing, feeling, or thinking at the time of the question. Examples include: "Any questions on that?" and "What or why are you thinking about that?" If the question required students to think and not just report something already known or previously thought by someone else, then we scored cognitive level higher than 2. Determining the level involved judging whether the student answering the question was building up a generalization, in which case we scored it a 3, or breaking down an argument, in which case we coded it as an analysis and rated its cognitive level as 4. Generalizations display inductive reasoning, building up ideas rather than breaking them down. They address questions such as: "What happens?" and "What do I make of what happens?" They tie things together; they are not restatements of information. Analyses display deductive reasoning, breaking concepts, ideas, and arguments down rather than building up ideas. To be scored as analyses, questions had to require more than restatements of known information. Questions were judged to be lower order (i.e., eliciting records or reports) if they elicited old information, or higher order (i.e., eliciting generalizations, analyses, or speculations) if they elicited new information and could not be answered through the routine application of prior knowledge.9 Unless texts explicitly stated the answers to teachers' questions, we judged most questions about literary texts to be either generalizations or analyses. Hence, from a ninth-grade class session on *To Kill a Mockingbird*, the question, "How does Tom die?" elicited a report since the answer is stated in routine cognitive operation, we coded questions as eliciting reports. We defined prior knowledge as "prior to the previous night's homework."8

1. **Source of the question.** The same question that elicits an analysis from a person who has to figure things out may well elicit a report from another, more knowledgeable individual who already knows and simply needs to explain. For example, "Why did Odysseus and his men plan deliberately to blind Polyphemus?" may elicit an analysis from students (assuming, of course, that they have to figure out the answer and not merely recite their textbook on the point), but most likely will elicit a report if a student asks a teacher who already knows the answer. When we were unclear, we asked about it after class.

2. **Experience, ability, and prior knowledge** of the person answering the question, whether student or teacher. If student answers seemed to require routine cognitive operation, we coded questions as eliciting reports. We defined prior knowledge as "prior to the previous night's homework."8

3. **Nature of the instructional activity.** When an episode was devoted to review, our normal expectation for responses was a report, even if questions had the linguistic form of higher-level questions (e.g., "What's the difference between a symbol and an image?" when asked as a study question).

4. **Source of information** required by the question. Information sources include prior experience, textbooks, and previous teacher lectures.

Level of cognition elicited by questions was measured on a five-point linear scale calibrated for level of abstraction and derived from Applebee (1981). Britton, Burgess, Martin, McLeod, and Rosen (1975), and Moffett (1968). Levels were as follows:

1. **Record** of an ongoing event: What's happening?
2. **Recitation and report** of old information: What happened?
3. **Generalization**: What happens?
4. **Analysis**: Why does it happen?
5. **Speculation**: What might happen?
ous studies documenting the historical and widespread prevalence of recitation in American schools. Indeed, as early as 1860, Morrison complained that "young teachers are very apt to confound rapid questioning and answers with sure and effective teaching" (cited in Hoetker & Ahlbrand, 1969, p. 153). In 1912, Stevens complained that the widespread practice of recitation made "the classroom the place for displaying knowledge instead of a laboratory for getting and using it" (p. 16). In 1919, Colvin estimated that only "about five percent [of the teacher questions he studied] could be considered in any way genuine thought questions" (p. 269). Writing about the same time, Miller (1922) complained that teachers were unable to "endure the silence that must prevail while the pupil is thinking and organizing his material" (quoted in Hoetker & Ahlbrand, 1969, p. 154). Thayer (1928) claimed that recitation was a progressive reform enabling teachers to gauge the mastery of large groups of children by checking the knowledgeability of relatively few. Corey (1940), Bellack, Kliebard, Hyman, and Smith (1966), and Hoetker (1967) all found that teachers talked about two-thirds of all instructional time and that more than 80% of all teacher questions sought to elicit recall in a recitation format. Recent studies continue to find similar results; see Duffy (1981), Durkin (1978-79), Hoetker and Ahlbrand (1969), Goodlad (1984), Sarason (1983), and Tharp and Gallimore (1988).

Class Time

Our study generally replicates these depressingly enduring findings. Recitation and lecture were common. When teachers were not lecturing, students mainly were either answering questions or engaged in seatwork. Indeed, on average, 85% of each class day in both eighth- and ninth-grade classes was devoted to a combination of lecture, question-and-answer recitation, and seatwork. Discussion and small-group work were rare. On average, discussion took 50 seconds per class in eighth grade and less than 15 seconds in grade 9; small-group work, which occupied about half a minute a day in eighth grade, took a bit more than 2 minutes a day in grade 9.11

In grade 8, more than two-thirds of all classes had at least 10 minutes of seatwork daily, and 31% had 20 minutes or more daily; only one class had no seatwork at all. Half of all classes had at least 10 minutes a day of question-and-answer activity, and 17.2% had 20 minutes or more. All but two classes routinely involved lecture; teachers in four of the 58 classes lectured 21 minutes or more each day. Overall means for all eighth-grade class activities are summarized in Table 2.3.

In ninth grade, time spent on lecture increased from 5 to 8 minutes per day (26% of all class time), and question-answer recitation increased from 12 to almost 18 minutes per day. Lecture was so ubiquitous that less than 2% of

---

Table 2.3: Grade 8 Allocation of Instructional Time to Various Activities (mean minutes per class day)

<table>
<thead>
<tr>
<th>Activity</th>
<th>All Classes</th>
<th>Rural Classes</th>
<th>Urban Class</th>
<th>Suburban Class</th>
<th>Low-Track Classes</th>
<th>High-Track Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>5.33 (2.23)</td>
<td>4.63 (2.27)</td>
<td>4.47 (2.78)</td>
<td>4.19 (2.54)</td>
<td>3.11 (3.30)</td>
<td>3.47 (4.30)</td>
</tr>
<tr>
<td>Question-Answer</td>
<td>4.84 (2.91)</td>
<td>3.96 (2.91)</td>
<td>4.63 (2.68)</td>
<td>4.47 (1.97)</td>
<td>2.52 (3.30)</td>
<td>3.11 (4.30)</td>
</tr>
<tr>
<td>Discussion</td>
<td>1.30 (1.99)</td>
<td>1.21 (1.96)</td>
<td>1.30 (1.65)</td>
<td>1.21 (1.65)</td>
<td>1.21 (1.65)</td>
<td>1.21 (1.65)</td>
</tr>
<tr>
<td>Student presentations</td>
<td>0.86 (1.99)</td>
<td>0.86 (1.99)</td>
<td>0.86 (1.99)</td>
<td>0.86 (1.99)</td>
<td>0.86 (1.99)</td>
<td>0.86 (1.99)</td>
</tr>
<tr>
<td>Student working students alone</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
</tr>
<tr>
<td>Teacher helping</td>
<td>1.37 (1.37)</td>
<td>1.37 (1.37)</td>
<td>1.37 (1.37)</td>
<td>1.37 (1.37)</td>
<td>1.37 (1.37)</td>
<td>1.37 (1.37)</td>
</tr>
<tr>
<td>Teacher working students alone</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
</tr>
<tr>
<td>Tests and quizzes</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
</tr>
<tr>
<td>Small-group work</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
<td>0.20 (0.00)</td>
</tr>
</tbody>
</table>

Notes: Statistics in parentheses are standard deviations. Times given for tests and quizzes are underestimates because we tried not to schedule observations on days when they were given.
all classes had none at all, and half the classes heard 8 minutes or more of lecture daily (one class actually had an average of more than 27 minutes a day). All classes involved question-and-answer recitation; 50% had at least 16 minutes daily (two classes averaged more than 30 minutes of recitation each day). Only two classes had no seatwork at all; at least 10 minutes daily was common for nearly a third of all classes. Overall, results on use of class time document the clearly monologic character of classroom discourse in the classes we observed. Overall means for use of class time in grade 9 are summarized in Table 2.4.

Questions

The character of instructional questions in the classes we observed was consistent with the monologic organization of class time. More commonly than not, students were treated as "empty vessels" to be "filled" by teachers. In virtually all classes, the teacher asked nearly all the questions; few about literature were authentic, and equally few followed up on student responses. In the eighth-grade classes (see Table 2.5), about 35 questions on average were asked during each class, 92% of them by the teacher. Only 12% of the teacher questions were authentic, and only 11% exhibited uptake. In the ninth-grade classes (see Table 2.6), teachers asked more than 52 questions each class period on average, or about 50% more than in the eighth-grade classes. 54% of all questions involved recitation. These results help explain National Assessment of Education Progress (NAEP) results, which perennially show that American students are far more proficient at literal comprehension than at analysis and critical thinking (see, e.g., Applebee, Langer, Mullis, Latham, & Gentile, 1994).

Given the infrequency of authentic questions in eighth grade, we were surprised to find that the proportion in ninth-grade classes was twice as high. Indeed, authentic questions were asked in all the ninth-grade classes we observed; half the classes routinely had 25% or more. In subsequent analysis, we discovered that much of the increase was due to the use of authentic questions to inquire about nonacademic topics. Uptake also was more common in ninth-grade classes, exhibited by 26% of all questions. Like eighth-grade instruction, however, most ninth-grade classroom interaction tended to be monologic, permitting little opportunity for substantive exchange.

Track Differences

The near-universal preference for "recitable information" afflicted low-track classes even more than regular- and high-track classes since low-track classes typically got a refracted, watered down, fragmented rendition of the

---

Table 2.4 Grade 9 Allocation of Instructional Time to Various Activities (mean minutes per class day)

<table>
<thead>
<tr>
<th>Activity</th>
<th>All Classes</th>
<th>Low-Track Classes</th>
<th>High-Track Classes</th>
<th>Urban Suburban Classes</th>
<th>Rural Suburban Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>8.42 (5.88)</td>
<td>8.58 (6.31)</td>
<td>7.46 (7.38)</td>
<td>7.58 (7.35)</td>
<td>8.58 (2.99)</td>
</tr>
<tr>
<td>Question-Answer</td>
<td>17.28 (5.55)</td>
<td>15.62 (4.27)</td>
<td>18.11 (6.58)</td>
<td>20.20 (6.45)</td>
<td>17.57 (7.14)</td>
</tr>
<tr>
<td>Discussion</td>
<td>0.24 (0.50)</td>
<td>0.51 (0.50)</td>
<td>0.37 (0.25)</td>
<td>0.09 (0.25)</td>
<td>0.28 (0.47)</td>
</tr>
<tr>
<td>Student presentation</td>
<td>1.24 (2.69)</td>
<td>4.87 (4.90)</td>
<td>1.50 (3.33)</td>
<td>1.02 (3.23)</td>
<td>1.92 (3.91)</td>
</tr>
<tr>
<td>Student reading aloud</td>
<td>2.98 (3.15)</td>
<td>5.97 (3.11)</td>
<td>2.52 (2.11)</td>
<td>5.70 (3.39)</td>
<td>2.09 (3.40)</td>
</tr>
<tr>
<td>Teacher helping</td>
<td>0.37 (0.36)</td>
<td>0.77 (0.11)</td>
<td>0.17 (0.33)</td>
<td>0.05 (0.11)</td>
<td>0.04 (0.11)</td>
</tr>
<tr>
<td>Supervised</td>
<td>0.11 (0.26)</td>
<td>0.03 (0.10)</td>
<td>0.07 (0.33)</td>
<td>0.01 (0.11)</td>
<td>0.04 (0.11)</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>1.87 (1.09)</td>
<td>2.52 (1.10)</td>
<td>3.18 (1.10)</td>
<td>5.55 (1.04)</td>
<td>1.29 (1.54)</td>
</tr>
<tr>
<td>Tests and quizzes</td>
<td>1.28 (1.69)</td>
<td>1.00 (1.50)</td>
<td>1.20 (1.50)</td>
<td>0.55 (1.50)</td>
<td>1.34 (1.50)</td>
</tr>
<tr>
<td>Small-group work</td>
<td>2.25 (4.44)</td>
<td>2.71 (2.95)</td>
<td>2.19 (4.51)</td>
<td>2.76 (2.95)</td>
<td>2.41 (2.94)</td>
</tr>
</tbody>
</table>

Notes: Statistics in parentheses are standard deviations. Times given for tests and quizzes are underestimated because we tried not to schedule observations on days when they were given.
Table 2.5. Properties of Classroom Questions in Grade 8 Literature Classes

<table>
<thead>
<tr>
<th>Class Type</th>
<th>Average Number of Questions per Class Session</th>
<th>Proportion of</th>
<th>Proportion of</th>
<th>Mean Cognitive Level</th>
<th>Proportion of Students Off Task</th>
<th>Average Class Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>All classes</td>
<td>34.62 (17.37)</td>
<td>0.92 (0.16)</td>
<td>0.11 (0.07)</td>
<td>0.10 (0.11)</td>
<td>0.02 (0.03)</td>
<td>0.36 (0.15)</td>
</tr>
<tr>
<td>Low track</td>
<td>36.91 (18.71)</td>
<td>0.97 (0.05)</td>
<td>0.11 (0.08)</td>
<td>0.12 (0.12)</td>
<td>0.04 (0.04)</td>
<td>0.37 (0.12)</td>
</tr>
<tr>
<td>High track</td>
<td>41.97 (20.91)</td>
<td>0.93 (0.09)</td>
<td>0.16 (0.07)</td>
<td>0.12 (0.10)</td>
<td>0.02 (0.03)</td>
<td>0.40 (0.14)</td>
</tr>
<tr>
<td>Urban</td>
<td>29.22 (18.25)</td>
<td>0.89 (0.20)</td>
<td>0.10 (0.06)</td>
<td>0.13 (0.10)</td>
<td>0.03 (0.03)</td>
<td>0.36 (0.15)</td>
</tr>
<tr>
<td>Suburban</td>
<td>42.52 (14.99)</td>
<td>0.96 (0.04)</td>
<td>0.14 (0.07)</td>
<td>0.12 (0.14)</td>
<td>0.02 (0.03)</td>
<td>0.40 (0.14)</td>
</tr>
<tr>
<td>Rural</td>
<td>41.11 (9.22)</td>
<td>0.98 (0.02)</td>
<td>0.11 (0.08)</td>
<td>0.06 (0.05)</td>
<td>0.02 (0.03)</td>
<td>0.25 (0.12)</td>
</tr>
</tbody>
</table>

Notes: N = 58 classes, including 15 low-track and 13 high-track classes, and 39 urban, 8 suburban, and 11 rural classes. Numbers in parentheses are standard deviations.

Table 2.6. Properties of Classroom Questions in Grade 9 Literature Classes

<table>
<thead>
<tr>
<th>Class Type</th>
<th>Average Number of Questions per Class Session</th>
<th>Proportion of</th>
<th>Proportion of</th>
<th>Mean Cognitive Level</th>
<th>Proportion of Students Off Task</th>
<th>Average Class Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>All classes</td>
<td>52.77 (20.25)</td>
<td>0.91 (0.10)</td>
<td>0.26 (0.12)</td>
<td>0.27 (0.19)</td>
<td>0.03 (0.03)</td>
<td>2.91 (0.38)</td>
</tr>
<tr>
<td>Low track</td>
<td>48.62 (22.35)</td>
<td>0.94 (0.05)</td>
<td>0.27 (0.13)</td>
<td>0.25 (0.19)</td>
<td>0.03 (0.04)</td>
<td>2.95 (0.35)</td>
</tr>
<tr>
<td>High track</td>
<td>60.05 (11.61)</td>
<td>0.89 (0.12)</td>
<td>0.26 (0.11)</td>
<td>0.28 (0.18)</td>
<td>0.02 (0.02)</td>
<td>2.85 (0.37)</td>
</tr>
<tr>
<td>Urban</td>
<td>53.78 (20.80)</td>
<td>0.92 (0.09)</td>
<td>0.27 (0.13)</td>
<td>0.33 (0.19)</td>
<td>0.03 (0.04)</td>
<td>2.99 (0.41)</td>
</tr>
<tr>
<td>Suburban</td>
<td>45.50 (15.28)</td>
<td>0.82 (0.16)</td>
<td>0.24 (0.12)</td>
<td>0.22 (0.16)</td>
<td>0.02 (0.01)</td>
<td>2.51 (0.28)</td>
</tr>
<tr>
<td>Rural</td>
<td>54.83 (22.12)</td>
<td>0.96 (0.04)</td>
<td>0.22 (0.10)</td>
<td>0.09 (0.04)</td>
<td>0.02 (0.02)</td>
<td>2.89 (0.46)</td>
</tr>
</tbody>
</table>

Notes: N = 54 classes, including 9 low-track and 13 high-track classes, and 35 urban, 8 suburban, and 11 rural classes. Numbers in parentheses are standard deviations.
regular curriculum; it was as if low-track students were to understand a book by dealing only with the index (Page, 1991). Our data also show that low-track students, in contrast to high-track students, engaged in far more clerical as opposed to compositional tasks; indeed many of their so-called “writing” tasks, such as filling-in-the-blanks, were not discourse at all. Their writing was more formulaic, and the level of response to their writing was low. In these low-track classes, the terms of reciprocity were limited mainly to procedures.

To examine differences between high- and low-track classes, we compared the two groups on allocation of instructional time and all discourse variables. In eighth grade, both groups spent 40% or more class time doing seatwork. The biggest differences were in time devoted to lecture—the teacher lectured to low-track students 40% more than to high-track students—and to discussion, which occupied nearly twice as much time in high-track classes than in low, hence giving a more dialogic tone to the high-track classes (even though the proportion of authentic teacher questions and uptake did not differ much). In ninth grade, seatwork occupied almost 13 minutes a day (29% of instructional time) in the low-track classes but less than 4 minutes (8%) in the high-track classes. Both groups spent close to 9 minutes in lecture and less than 1 minute in discussion, but the high-track classes spent more time doing small-group work and answering questions. (Tables 2.3 and 2.5 show some of these results for grade 8; Tables 2.4 and 2.6 provide a breakdown for grade 9.)

These results are consistent with our earlier research (Nystrand & Gamoran, 1988), which found that students in low-achieving classes are far more likely than their higher-achieving counterparts to be involved in fragmented, contrived learning. Using survey data, we found that students in low-achieving eighth- and ninth-grade English classes

- Did grammar exercises 2.6 times as frequently as did their high-achieving counterparts
- Did reports 2.4 times as frequently
- Filled in blanks 5 times as often
- Answered true-false questions 4 times as frequently
- Completed multiple-choice questions 4.1 times as often

In their responses to the papers of students in low-achieving classes, teachers commented

- 2.3 times as much about spelling (in marginal and terminal comments)
- 1.8 times as much about punctuation
- 2 times as much about grammar

In their responses to high-achieving students’ papers, however, teachers commented

- 1.7 and 1.9 times as much about content (in marginal and terminal comments, respectively) compared with teacher comments on low-track papers

Teachers held writing conferences with low-ability students about as infrequently (about once a month on average) as with high-achieving students. However, in these conferences they discussed spelling 2.6 times as much with students in low-achieving classes, and they discussed content 1.9 times as frequently with high-achieving students.

Why these differences? To answer this question, we looked closely at interviews of the teachers in our study and then compared what they said with what we observed in their classes. First, despite considerable lip service to “discussion,” we observed little discussion in any classes in the sense of in-depth exchanges of ideas in the absence of teacher evaluation. Most teachers who spoke readily of their value for “discussion”—and indeed there were a great many—really enacted some version of recitation. Discussion almost always turned out to be what one teacher described as “question-and-answer discussion” involving a prescribed, teacher-set exchange. Such discussion was rarely collaborative, thoroughgoing, pushed-to-the-limit sharing and exploration of student ideas unfolding in class—what Britton (1970) described as a “struggle to organize . . . thoughts and feelings, to come up with words that . . . shape an understanding, [a] struggle to rise above the limitations of [the] language” (p. 12). What mainly varied in the lessons we observed was the length of students’ responses as they answered teachers’ questions.

Some other teachers expressed a conception of discussion best described as forensic. One teacher, who taught academically talented ninth-grade English in a large urban high school, believed that schooling too much favors docile, cooperative students; in contrast, he liked aggressively expressive, openly assertive students who could readily state and defend their points of view and were willing to argue in class, even with him. Right answers weren’t enough in his classes, he said; students had to be able to support them and prevail. Needless to say, this conception of discussion as debate favored the most confident, verbally articulate, and competitive of students. This was a view most commonly heard in suburban schools and was expressed almost exclusively by male teachers (most of the suburban teachers in our study were male; most of the urban teachers were female).

Given these conceptions of discussion—one recitation, the other debate—it is perhaps not surprising that students tracked into low-ability classes proved hesitant or “reticent” in the classroom, as their teachers often de-
she said, to engage the high-track students in discussion. She did not seem to realize that these differences in pedagogy and interaction afforded the two groups of students significantly different learning opportunities, especially favoring the high-ability students at the expense of the low (Gamoran, Nystrand, Berends, & LePore, 1995).

Far more teachers supported ability grouping than opposed it. Those opposing it were far more likely to be in suburban and rural schools than in urban schools, where opposition was very rare. Rural teachers, who seemed to view school as part of extended families in their small towns, were more concerned than the other teachers about possible negative effects on the learning and self-image of their slower students, who in tracked groups would not "have anyone to model or peers with good behavior and good skills," as one teacher put it. Several suburban teachers particularly were disaffected with high-ability students, whom one described as "pseudo-intellectuals and pseudo-sophisticates" who tended to be "abrasive." Another teacher said, "Sometimes the higher-level kids get an attitude that 'I'm superior—I don't have to work hard.'" Some suburban teachers felt that mixed grouping could benefit average students. They thought it could improve student writing by expanding the stock of shared experience in the class: "Writing is based on experience," said one. Others thought it could stimulate discussion and improve literature instruction. Yet these teachers were in the minority. We interviewed only one teacher who claimed that while mixed-ability groups might make it more difficult to teach, they also made it "more fun," "They're more real," she claimed.

Most teachers, however, were content to teach tracked classes. Some candidly said they were better able to engage the high-ability students than the low simply because they had more in common with the high-ability students; they had more to talk about together. One such teacher taught eighth-grade reading to "basic" students and literature to "academically talented" students in a midwestern urban junior high. At her school, IQ scores played a major role in assigning students to different ability groups, and she agreed that her academically talented students were "innately brighter," especially able to manipulate multiple points of view, and more likely than her basic students to go beyond the literal meaning of their readings. She believed that mixing students of different abilities shortchanged everyone: If teachers pitched things toward the higher-ability students, she explained, they lost the slower students; if they made adjustments for the slower students, their classes were not as stimulating for the faster students. Hence, she felt such students needed to be separated from the basic students "for their own good." She preferred teaching the academically talented because she could do more things with them—"more fun things and creative things." She spent "too much time" on discipline in the basic class and found that the students "hated"
Another teacher said her school system had been very remiss in coming up with a plan for these basic kids. In fact, there isn't one. So all they have done is throw these materials at us, and no one has ever sat down and said, “Look, this is what you should cover. This is what these kids should know,” or whatever. They pretty much leave it to the teacher. And if you went to other schools in the district, they would be doing something totally different. It's a very poor setup for these kids. And they are the ones that need the most structure, and they are not getting it.

Discouraged, these teachers often set remarkably low expectations for their students. They said they needed books with shorter chapters, lower-level vocabulary, and higher interest levels. One teacher said that morning was the best time to teach because her students' concentration was highest and she had fewer behavioral problems then. Students were more likely to be alert and fresh; there were “no clowns to disturb anyone.” Even though her first-period class did not do much work, she nonetheless judged it to be a good one because the students were obedient and “did what they were told,” and many of the troublemakers were “not alert enough to be disruptive.” Absenteeism was high, and one-third of her students were usually late. Fridays were always the hardest day of the week. Discussion was hard. Most readings in the curriculum were irrelevant to students' lives. More than anything, this teacher said, she wanted her students to know she cared about them. After that, discipline — coming to class, doing the work, and following directions — was her most important goal.

Yet not all urban teachers had given up. One teacher, specially trained to teach Title I students, was upbeat. She liked her students and found the remediation techniques she had learned highly effective. She reported no significant discipline problems and found parents cooperative. She especially favored small-group work because there “you can really discuss.” She liked using newspapers because “you could discuss recent developments,” and she claimed success when she asked students to write questions about what they were learning. Another teacher emphasized that it was up to her as the teacher to “discern gifts, bring out students, make them grow,” and that she was able to do this. She looked at students as individuals — “people working with me”— and her students often came back just to “talk about life.” She allowed regular time for “impromptu sessions,” and said students became engaged when she took the time to ask, “What do you think of this?” Another teacher stressed the importance of respect: “Teaching has got to be respectful. And you have to establish that rapport early because the anger and hostility are very difficult to wear off once they start.”
Nonetheless, most of the teachers in the urban schools said discussion and small-group work were impossible. They generally had resigned themselves to worksheets and newspapers as the best they could offer. Most just tried to keep order in their classrooms, especially through seatwork, which was the dominant mode of instruction, averaging more than 21 minutes a day. On top of this, eighth-grade urban teachers lectured about 6 minutes a day. Even question–answer exchanges involved too much interaction for these teachers, and it averaged only about 9 minutes a day versus 14 minutes for suburban students and nearly 19 minutes for rural students. Unsurprisingly, discussion was rare in these classes, averaging only 45 seconds a day.

Eighth-grade urban teachers asked about as many authentic questions (13%) as suburban teachers did (12%) but twice as many as rural teachers (6%). In ninth-grade classes, these differences were even more pronounced: 33% of teacher questions asked in urban classes were authentic compared with 22% in suburban classes and 9% in rural classes, although in a subsequent analysis (see p. 58), we discovered that many authentic questions did not concern literature or anything academic. Tables 2.5 and 2.6 summarize these and other data.

In grade 9, seatwork in urban classes diminished to about 7 minutes a day, whereas lecture increased to almost 9 minutes a day and question–answer increased to 18. This was the case even though, unlike the eighth-grade classes, the average number of students in ninth-grade urban classes (about 25) was notably higher than either suburban (22) or rural (19) classes. Tables 2.4 and 2.6 summarize these and additional data.

Suburban Classes. Most suburban teachers in our sample said they prized student autonomy: the ability and willingness to articulate and defend ideas. The suburban teachers were clearly more up-to-date on leading-edge pedagogy and professional buzzwords than were the urban teachers. The fundamental province of their instruction, as they saw it, was the life of the mind. Discussion commonly was cited as a goal for all classes; one teacher represented the expressed views of the others when he said, “It is important for students to participate, to feel free to voice their opinions, to present divergent opinions without feeling intimidated, to do the work, be motivated, and be free and willing to discuss literature.” His goals for students were excitement and depth of understanding. Another teacher wanted all his students to “discuss, contribute, and offer original ideas.”

In fact, we observed more teacher–student interaction in the suburban than in the urban schools, especially in eighth-grade classes. The suburban teachers did less lecturing than the urban (4 minutes a day versus 6), and their students did less seatwork (15 minutes a day versus more than 21). Yet “discussion” in these schools was rarely open-ended and generally took the form of question-and-answer recitation: Eighth-grade suburban classes typically had more than 14 minutes a day of question–answer activities versus less than 10 minutes in urban classes, and actual discussion was about 50 seconds a day on average, not much more than in urban schools. In ninth-grade suburban classes, we encountered not even a second of discussion.

If suburban teachers prized student autonomy and independent thought, they also said they prized sharing and cooperation. Many teachers used small groups to promote such values, and in both eighth- and ninth-grade classes, suburban students spent twice as much time as urban students in small groups.

Reading skills of suburban students were stronger than those of urban students, and this difference was evident in the fact that suburban students spent less time reading aloud in class. The eighth-grade classes spent about 2½ minutes a day reading aloud compared with more than 3 minutes a day in the urban classes; in the ninth-grade classes, this difference became even more pronounced: The suburban classes spent only 43 seconds reading aloud, whereas urban classes spent more than 2 minutes. While urban teachers often sought simplified reading materials, suburban teachers sometimes did just the opposite. As one told us, “Last year I used a basic reader. . . . There were good stories, but they were so watered down that I said I can't use this book. I realize that the stuff I use now, the kids are frustrated with the vocabulary, but that's life.” The suburban teachers said they spent more time on interpretation than on reading skills. They told us they encouraged students to give opinions, explain why, and go “beyond remembering” to “using” texts. Yet the cognitive level of questions asked in eighth-grade suburban classes was no higher than in urban classes, and in ninth grade it was actually lower. Nor did suburban teachers ask more authentic questions: In eighth grade, about 12–13% of all teacher questions were authentic in both suburban and urban classes; in ninth grade, urban teachers asked about 50% more. Tables 2.3–2.6 summarize these and other data.

Rural Classes. Schools in small, rural towns were a lot like the towns themselves—places where everybody knew everybody and where secrets could sometimes be hard to keep. Teachers often knew the siblings of their students and kept in close touch with the parents: One teacher regularly recognized his students as “chips off the ol' block.” Because the same students often spent the day together through all their classes, they often knew each other's grades, and this was sometimes a problem. Also, because the schools were so small, there was often no ability grouping; everyone was just mixed together. Virtually all the teachers approved of this. They felt that “the slower
kids don’t learn as much if they’re grouped,” and believed that mixed groups provided these students with role models. They also said it was important for the smarter kids to “learn the other side of life.”

Teachers told us they prized “question-and-answer discussion,” which averaged approximately 18 minutes a day in both eighth and ninth grades. We found more open-ended discussion in eighth-grade rural classes than in any of the other schools, on average a little less than 2 minutes a day. Also, seatwork in grade 8 occupied much less time in rural than in urban and suburban schools. Tables 2.3–2.6 summarize these and other data.

ANALYZING THE DATA: THE EFFECTS OF DISCOURSE ON LEARNING

How concerned should we be with these bleak figures? Our data enabled us to test specific hypotheses concerning the overall effects of dialogic elements on learning. There is clearly a trade-off between research, such as our study, that comprehensively depicts the big picture and examines general effects, on the one hand, and case studies that examine the dynamics of individual cases and episodes, on the other. Ideally these different perspectives complement each other. Our large study enabled us to test empirically many widely debated hypotheses about the effectiveness of different instructional practices and discourse environments (e.g., recitation, discussion, small-group work) for learning.

In a series of analyses, we examined the effects of many of these practices on literature achievement. To assess student learning, we administered a literature test in the spring to each class. The test required students to answer a series of questions about five works of literature (stories, novels, dramas, and short plays) they had read during the year.14 The questions ranged from simple recall (e.g., “Who were the main characters in Roll of Thunder, Hear My Cry?”) to ones requiring in-depth understanding (“Relate the conflict of Roll of Thunder, Hear My Cry to the ending and to the theme”). The same types of questions were asked of each class, but the stories varied, depending on what students actually read during the year. For the ninth-grade test students also wrote a brief essay on a character from their reading whom they admired. An example of an eighth-grade test is found in Appendix A.

The literature tests were scored for the following:

1. Extent of recall
2. Depth of understanding
3. Number of endings remembered
4. Relation of ending to denouement
5. Relation of conflict/and or ending to theme
6. Understanding of the internal motivations of characters
7. Interpretive treatment of the major selection
8. Level of discourse used to discuss theme and conflict

Readers read the entire test and then determined a single score for each of the above variables. Each student’s literature score was the sum of the individual scores. Each test was scored by two readers and the scores were averaged. The overall reliability of the assessment in grade 8, computed as a correlation of the two readings, was .90; in the grade 9 assessment, the reliability was .82. The rubric used for scoring the tests is shown in Appendix B.

We then examined the effects of instruction and classroom discourse on learning through a statistical technique called regression analysis. Regression analysis makes it possible to examine the effect of one condition (e.g., the amount of time spent in discussion) while statistically holding constant other important conditions (e.g., prior abilities in writing and reading, socioeconomic status, and characteristics of students in different classes). For example, if authentic questions are found to be related to higher achievement, regression analysis can reveal whether this is because teachers ask previously higher-achieving students more authentic questions, or whether authentic questions actually promote higher achievement. Regression analyses systematically estimate the effect of each variable while statistically holding constant each of the other variables. Our analyses controlled for the effects of both background variables (sex, race, ethnicity, family socioeconomic status) and prior achievement (as measured by fall tests of reading and writing skills).

Overall Results

In none of our analyses did we ever find that a higher cognitive level of instructional activities actually enhanced learning. Instead, we could explain the relative effectiveness of different instructional practices only when we examined the ways teachers and students interacted as evidenced by authentic questions, uptake, and especially discussion.

Eighth-Grade Classes. The results of our analyses of the eighth-grade classes, controlled for writing and reading ability, socioeconomic status (SES), race, and ethnicity, showed unsurprisingly that disengagement, including off-task behavior and failure to complete homework, adversely affected achievement. The results revealed a modest effect for time spent on homework, no effect for how many questions were asked in class, and a negative effect for level of activity during recitation. Results indicated that dialogically organized instruction, indicated by time devoted to discussion, authentic questions, uptake, and high-level teacher evaluation, had a strong, positive effect on
achievement. Discussion in particular had a large effect, which is especially striking when it is recalled that the average class engaged in less than a minute of discussion each day. Table 2.7 summarizes these results.

In our eighth-grade study, we also found that effective teachers of literature regularly assigned extended pieces of exposition (Nystrand, 1991c); this practice enhanced students’ recall and understanding of the literary works they read. The frequent assignment of short-answer exercises, however, actually degraded students’ overall recall and depth of understanding. This result is consistent with Applebee’s (1984) contention that, because writing tends to promote recall of what it focuses on, such “narrow-banded” activities as short-answer exercises are likely to hinder total recall—in other words, helping students to remember trees at the expense of understanding the overall shape of the forest. In addition, because they elicit cryptic, fragmented discourse, short-answer exercises promote superficial involvement with literature; in so doing, they trivialize students’ experiences with literature. All in all, students learn literature best in classes that encourage substantive and personal student response to literature in both classroom interaction and writing. Table 2.8 summarizes the results.

**Ninth-Grade Classes.** In analyzing data from the ninth grade (Gamoran & Nystrand, 1992), we sought to replicate our findings from the eighth-grade study, but we were initially frustrated to find that discussion had no effect on learning in ninth grade and that authentic questions appeared to have no effect or even a negative effect. We were further perplexed to discover that authentic questions had positive effects in high-track classes but negative effects in low-track classes. Table 2.9 summarizes these results.

Looking more closely at our data, we discovered that the two different tracks used authentic questions very differently. In the high-track classes, fully 68% of authentic questions concerned literature, whereas only 25% of authentic questions in low-track classes did. In low-track classes, teachers’ authentic questions often concerned such issues as, “How do most of you feel about tests?”; “What would your parents say if you got an A on next week’s test?”; “What things would you associate with lying in the sun?”; “Do you ever have to take notes?” Discussion broke down in a similar way so that discussion in the high-track classes tended to be about literature far more than did that in the low-track classes.

**Tracking, Instructional Discourse, and Learning**

Again and again, researchers have found that tracking and ability grouping promote inequality of achievement, as the gap between students in high-track and low-track classes widens over time (Gamoran & Berends, 1987).

---

Table 2.7. Effects of Disengagement, Procedural Engagement, and Substantive Engagement on Spring Literature Achievement in Grade 8 (metric regression coefficients)

<table>
<thead>
<tr>
<th>Model</th>
<th>Independent Variable</th>
<th>Procedural Engagement Variables Only</th>
<th>Substantive Engagement Variables Only</th>
<th>Full Model (All Variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (1 = female)</td>
<td>0.47</td>
<td>0.62*</td>
<td>0.59*</td>
<td></td>
</tr>
<tr>
<td>Race (1 = Black)</td>
<td>-2.67***</td>
<td>-1.55*</td>
<td>-1.75***</td>
<td>-1.10**</td>
</tr>
<tr>
<td>Ethnicity (1 = Hispanic)</td>
<td>-1.51*</td>
<td>-0.15*</td>
<td>-1.47*</td>
<td>-0.58*</td>
</tr>
<tr>
<td>SES</td>
<td>1.62***</td>
<td>1.10***</td>
<td>1.45***</td>
<td>1.05***</td>
</tr>
<tr>
<td>Grade (1 = eighth)</td>
<td>2.09***</td>
<td>1.10**</td>
<td>1.11*</td>
<td>0.16</td>
</tr>
<tr>
<td>Fall reading score</td>
<td>0.39***</td>
<td>0.30***</td>
<td>0.36***</td>
<td>0.30***</td>
</tr>
<tr>
<td>Fall writing score</td>
<td>0.93***</td>
<td>0.70***</td>
<td>0.73***</td>
<td>0.58***</td>
</tr>
<tr>
<td><strong>Disengagement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off task in class</td>
<td>-0.23***</td>
<td>-0.16***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading not completed</td>
<td>-0.02</td>
<td>-0.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing not completed</td>
<td>-0.03***</td>
<td>0.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonresponse to questions</td>
<td>-0.20***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Procedural engagement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active in class</td>
<td>-0.03***</td>
<td>-0.06***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asking questions</td>
<td>-0.02</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time on homework</td>
<td>0.45**</td>
<td>0.39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Substantive engagement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authentic questions</td>
<td>0.08***</td>
<td>0.04*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authentic reading</td>
<td>-0.01</td>
<td>0.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High evaluation of writing</td>
<td>-0.39</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uptake</td>
<td>0.14***</td>
<td>0.10*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coherence of reading</td>
<td>0.16***</td>
<td>0.11**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion time</td>
<td>0.34**</td>
<td>0.29**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small- group time</td>
<td>-0.19*</td>
<td>-0.23***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Dependent variable: Spring Literature Achievement Test. N = 924 students. Missing values deleted listwise. Numbers in parentheses are standard errors.

* p < .10 (marginal trend) ** p < .05 *** p < .01 **** p < .001
Table 2.8. Effects of Selected Instructional Variables on Difficulty of Recall and Difficulty of Understanding Literature in Depth in Grade 8
(metric regression coefficients)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Difficulty of Recall</th>
<th>Difficulty of Understanding in Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade (1 = eighth)</td>
<td>0.054</td>
<td>-0.183</td>
</tr>
<tr>
<td>Race (1 = Black)</td>
<td>0.246</td>
<td>0.065</td>
</tr>
<tr>
<td>Ethnicity (1 = Hispanic)</td>
<td>0.122</td>
<td>0.165</td>
</tr>
<tr>
<td>SES</td>
<td>-0.107***</td>
<td>-0.138***</td>
</tr>
<tr>
<td>Sex (1 = female)</td>
<td>-0.120**</td>
<td>-0.174*</td>
</tr>
<tr>
<td>Fall writing score</td>
<td>-0.073***</td>
<td>-0.105***</td>
</tr>
<tr>
<td>Fall reading score</td>
<td>-0.034****</td>
<td>-0.036****</td>
</tr>
<tr>
<td>Time on homework</td>
<td>-0.065***</td>
<td>-0.059*</td>
</tr>
<tr>
<td>Reading completed</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>Writing not completed</td>
<td>0.005**</td>
<td>0.002**</td>
</tr>
<tr>
<td>Participation in class</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>No response to teacher questions</td>
<td>0.008</td>
<td>0.026*</td>
</tr>
<tr>
<td>Amount of writing</td>
<td>-0.256**</td>
<td>-0.289**</td>
</tr>
<tr>
<td>Discussion time</td>
<td>-0.022</td>
<td>-0.023*</td>
</tr>
<tr>
<td>Authenticity of teacher questions</td>
<td>-0.0005</td>
<td>-0.006**</td>
</tr>
<tr>
<td>Authenticity of readings</td>
<td>-0.003</td>
<td>-0.018***</td>
</tr>
<tr>
<td>Uptake</td>
<td>-0.015****</td>
<td>-0.020****</td>
</tr>
<tr>
<td>Relating discussions to other discussions and student compositions</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>Relating readings to other readings</td>
<td>-0.046**</td>
<td>-0.018**</td>
</tr>
</tbody>
</table>

Notes: N = 762 students. Missing values deleted listwise. Numbers in parentheses are standard errors.

Table 2.9. Effects of Instruction and Engagement on Ninth-Grade Literature Achievement

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Regression Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (1 = female)</td>
<td>0.51</td>
<td>0.50</td>
<td>1.47**</td>
<td>0.37</td>
</tr>
<tr>
<td>Race (1 = black)</td>
<td>0.07</td>
<td>0.26</td>
<td>-0.46</td>
<td>0.72</td>
</tr>
<tr>
<td>Ethnicity (1 = Hispanic)</td>
<td>0.09</td>
<td>0.28</td>
<td>-1.56*</td>
<td>0.65</td>
</tr>
<tr>
<td>SES</td>
<td>-0.02</td>
<td>0.80</td>
<td>0.44</td>
<td>0.25</td>
</tr>
<tr>
<td>Fall reading score</td>
<td>31.88</td>
<td>5.34</td>
<td>0.40**</td>
<td>0.04</td>
</tr>
<tr>
<td>Fall writing score</td>
<td>5.71</td>
<td>1.28</td>
<td>0.90**</td>
<td>0.15</td>
</tr>
<tr>
<td>Ability groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors/accelerated</td>
<td>0.24</td>
<td>0.43</td>
<td>0.25</td>
<td>0.96</td>
</tr>
<tr>
<td>Basic/remedial</td>
<td>0.10</td>
<td>0.30</td>
<td>-0.09</td>
<td>1.13</td>
</tr>
<tr>
<td>Other</td>
<td>0.09</td>
<td>0.29</td>
<td>0.57</td>
<td>1.16</td>
</tr>
<tr>
<td>Psychological engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement scale</td>
<td>2.42</td>
<td>1.40</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>Behavioral engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing completed</td>
<td>87.88%</td>
<td>19.68</td>
<td>0.03**</td>
<td>0.01</td>
</tr>
<tr>
<td>Reading completed</td>
<td>83.04%</td>
<td>24.62</td>
<td>0.03**</td>
<td>0.01</td>
</tr>
<tr>
<td>Homework time (hrs/wk)</td>
<td>1.27</td>
<td>1.27</td>
<td>0.10**</td>
<td>0.03</td>
</tr>
<tr>
<td>Off task in class</td>
<td>3.22%</td>
<td>3.27</td>
<td>-0.12*</td>
<td>0.06</td>
</tr>
<tr>
<td>Instructional discourse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authenticity of readings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In honors classes</td>
<td>24.30%</td>
<td>11.41</td>
<td>0.10**</td>
<td>0.03</td>
</tr>
<tr>
<td>In regular classes</td>
<td>28.13%</td>
<td>18.81</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>In remedial classes</td>
<td>27.40%</td>
<td>18.86</td>
<td>-0.09**</td>
<td>0.04</td>
</tr>
<tr>
<td>In other classes</td>
<td>36.90%</td>
<td>26.03</td>
<td>-0.20**</td>
<td>0.03</td>
</tr>
<tr>
<td>Uptake</td>
<td>25.90%</td>
<td>11.26</td>
<td>0.09**</td>
<td>0.02</td>
</tr>
<tr>
<td>Discussion (min/day)</td>
<td>0.24</td>
<td>0.48</td>
<td>-0.18</td>
<td>0.40</td>
</tr>
<tr>
<td>Coherence^a</td>
<td>13.01</td>
<td>7.07</td>
<td>0.12**</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Notes: N = 971 students. Dependent variable: Spring Literature Achievement Test (mean = 21.82, SD = 7.66). R² = .52.

* Other classes include two classes in a school-within-a-school program and two classes in a heterogeneously grouped school.
** Engagement scale based on student responses to the following questions, coded on a weekly scale: In English class, how often do you: (a) Try as hard as you can? (b) Think what you are learning is interesting and worthwhile? (c) Find yourself concentrating so hard that time passes quickly?
*** Authenticity of readings measure based on teacher responses to the following questions, coded on a weekly scale: About how often do you and your class discuss things students have read? (b) About how often do you discuss writing topics with your students before asking them to write? (c) About how often do you and your class discuss the readings you assign? (d) When you ask students about their reading assignments in class, how frequently do you attempt to do each of the following: Ask them to relate what they have read to their other readings? (e) About how often do you and your class discuss things students have written about?

R² Source: Nystrand, 1991c.
Notes: N = 762 students. Missing values deleted listwise. Numbers in parentheses are standard errors.

*p < .10 (marginal trend)  **p < .05  ***p < .01  ****p < .001
A major purpose of our investigation was to determine whether inequality resulted from differences in the quality of discourse in high-track and low-track classes. This question is not easy to answer, for two reasons. First, prior to our study, quantitative measures of discourse quality had not been available. Second, it is hard to measure the effects of tracking because they are confounded with initial differences among students. Students assigned to different tracks have different achievement from the start, and it is important to take this into account when examining inequality.

We addressed the first problem by developing the scheme for measuring classroom discourse described in this chapter. We addressed the second problem with statistical controls for fall reading and writing skills, as well as controls for sex, race/ethnicity, and SES (see note 15). In addition, for this analysis, we used standardized reading and math test scores as indicators of an underlying "ability" construct, to add a stricter control for students' prior academic background (Gamoran et al., 1995). Even with these controls, we found that average achievement was almost two points higher in honors classes than in remedial classes, a statistically significant difference. One standard deviation on the test was 6.8 points, so the two-point gap was almost 30% of a standard deviation, a substantial amount (Gamoran et al., 1995).

Do differences in instructional discourse account for this achievement inequality? Yes, to a large extent (Gamoran et al., 1995). Part of the difference occurred because students in honors classes were more responsive—they completed more of their reading and writing assignments—which helped them learn more (see Table 2.10). Uptake and discourse coherence also helped students learn, but these did not account for track differences in learning, because they occurred at similar rates in honors and remedial classes (see Tables 2.5 and 2.6). Authentic questions also were comparably distributed across class types, but they were beneficial only in honors classes because there they pertained to literature (see Table 2.10). Hence, similar levels of authenticity led to inequality of learning, because the content of the authentic questions differed.

We further observed that students were more often off task in remedial classes, and this was detrimental to their achievement (see Table 2.10). In honors classes, where off-task behavior was less prevalent, its effects were less harmful. Conversely, discussion occurred more often in honors classes, where it brought benefits for literature achievement, and had no benefits in other classes, presumably because it focused on topics other than literature.

These results contradicted the usual rationale for ability grouping. Ability grouping is supposed to allow teachers and students to engage in whatever instruction is most beneficial in each context. In actuality, we found that authentic questions occurred at similar rates in all classes, but contributed to achievement only in honors classes; off-task behavior was most common where it was most harmful, in regular and remedial classes. Discussion fit the pattern of occurring more where its payoff was greater (in honors classes); this, too, contributed to inequality.

In light of these results, we concluded that two changes must be considered (Gamoran et al., 1995). Either ability grouping in eighth- and ninth-grade English should be eliminated, or it should be implemented very dif-

---

**Table 2.10. Maximum Likelihood Estimates of Background and Instructional Effects on Literature Achievement in Eighth- and Ninth-Grade Ability-Grouped English Classes**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Effect</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (1 = female)</td>
<td>1.188a</td>
<td>0.252</td>
</tr>
<tr>
<td>Minority (1 = Black or Hispanic)</td>
<td>-0.652</td>
<td>0.339</td>
</tr>
<tr>
<td>SES</td>
<td>0.155</td>
<td>0.174</td>
</tr>
<tr>
<td>Fall reading score</td>
<td>0.202a</td>
<td>0.024</td>
</tr>
<tr>
<td>Fall writing score</td>
<td>0.512a</td>
<td>0.103</td>
</tr>
<tr>
<td>Ability</td>
<td>0.121a</td>
<td>0.018</td>
</tr>
<tr>
<td>Instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completion of reading</td>
<td>0.022b</td>
<td>0.006</td>
</tr>
<tr>
<td>Completion of writing</td>
<td>0.025b</td>
<td>0.007</td>
</tr>
<tr>
<td>Off task in class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors classes</td>
<td>0.149</td>
<td>0.092</td>
</tr>
<tr>
<td>Regular classes</td>
<td>-0.193a</td>
<td>0.044</td>
</tr>
<tr>
<td>Remedial classes</td>
<td>-0.124a</td>
<td>0.028</td>
</tr>
<tr>
<td>Authentic teacher questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors classes</td>
<td>0.056c</td>
<td>0.022</td>
</tr>
<tr>
<td>Regular classes</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td>Remedial classes</td>
<td>-0.050c</td>
<td>0.017</td>
</tr>
<tr>
<td>Uptake</td>
<td>0.063a</td>
<td>0.013</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors classes</td>
<td>0.277c</td>
<td>0.129</td>
</tr>
<tr>
<td>Regular classes</td>
<td>-1.510c</td>
<td>0.591</td>
</tr>
<tr>
<td>Remedial classes</td>
<td>0.045</td>
<td>0.169</td>
</tr>
<tr>
<td>Discourse coherence</td>
<td>0.158c</td>
<td>0.022</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honors classes</td>
<td>-8.502a</td>
<td>1.385</td>
</tr>
<tr>
<td>Regular classes</td>
<td>-7.081a</td>
<td>1.207</td>
</tr>
<tr>
<td>Remedial classes</td>
<td>-7.144a</td>
<td>1.061</td>
</tr>
</tbody>
</table>

\[ \chi^2 (61, N = 1,564) = 86.33 \]

**Source:** Gamoran et al., 1995

- Coefficient is four times its standard error.
- Coefficient is three times its standard error.
- Coefficient is twice its standard error.

where it was most harmful, in regular and remedial classes. Discussion fit the pattern of occurring more where its payoff was greater (in honors classes); this, too, contributed to inequality.
ferently than typically occurs at present. To mitigate the inequality that results from grouping and tracking, teachers and students in regular and low-level classes need higher expectations and more engaging discourse that focuses on academic subject matter.

Small Groups

In the eighth-grade study, we initially expected small groups to be dependable sites of dialogic peer interaction, but this was naive. In fact, time spent in small groups, rare as it was—39 seconds a day on average—turned out to have a prominent negative effect on achievement. We had predicted that small-group work and discussion would enhance achievement by engaging students substantively, especially compared with lecture, drill work, and recitation, which, like short-answer study questions, typically involve abbreviated responses from students. In fact, increased time spent in small-group work seemed to result in lower achievement in literature. We decided to examine this finding in more detail in a follow-up study of ninth-grade literature classes. We were particularly interested in distinguishing various kinds of small-group work and to see whether some were more effective than others.

In the ninth-grade study, unlike the eighth-grade study, we audiotaped the classes we observed, and although we had not placed microphones in small groups, the tapes provided a record of how each class proceeded. For classes involving small-group work, we were able to determine: (a) what preceded and what followed this work, (b) the kinds of tasks undertaken in groups, (c) the instructions the teachers gave to the groups, and (d) the roles teachers played in setting up and running the groups.

We learned that small-group work involved a great range of activities. Some small-group work was so highly structured by the teacher—involving, for example, students completing worksheets together—that it might best be called “collaborative seatwork.” Other groups, which we called problem-solving groups, required students to come to consensus concerning some issue or question that the teacher defined. Yet other small-group work, which we described as “autonomous,” was even more open-ended, with the groups themselves defining as well as resolving the problems and issues they discussed. Figure 2.2 shows the continuum described by these categories.

Our study looked exclusively at the features of small-group work that promote thinking. Given our findings that ownership, coherence of discourse, and student production of knowledge are important features of effective instruction, we reasoned that to promote response to and thinking about literature, small groups also should manifest these traits. More specifically, we predicted that collaborative seatwork, which is essentially written recitation done by students working together, would be less effective than problem-solving and autonomous groups, which allow ownership and thereby maximize the possibility of coherent discussion. It may be that collaborative seatwork is effective for teaching facts and grammar, but our research did not examine this possibility.

As we examined the range of activities that we had coded as small-group work, we confirmed our initial impression that “small-group time” was a misleading category because of important differences among the activities that occurred in small groups. We can illustrate these contrasting activities by focusing on two dimensions of work in small groups: student autonomy and student production of knowledge. Small-group activities occurred in only 29 of the 216 class sessions we observed. The small-group sessions averaged about 15 minutes when they occurred, but because they were so infrequent, small-group time accounted for less than 2 minutes of the average 50-minute period among all the classes. The groupwork we observed was usually closer to collaborative seatwork than to student problem-solving and autonomous groupwork.

**Student Autonomy.** Teachers shape groupwork by assigning tasks and establishing parameters of interaction. In highly “prescripted” groupwork, the parameters are defined entirely by the teacher, and the task could just as easily be done without interaction among students. The group setting is gratuitous, used perhaps more for the teacher’s convenience. Here is an example from a ninth-grade English class.

Today while you are working in groups you will keep the same groups that you had yesterday. The same rules will apply, and those are: You must, as a group, form a tight circle; you must bring all of your materials with you to that group and you may not get up from your seat; your voice must stay at a whisper—if I can hear you above anyone else’s, that means you are too loud and your name will go up on the
board. You may do one of two things in your group. You may continue
to work on your paper — there are five, probably six people I have to see
yet in conference. If you are as far as you can go in your groups with
your papers — that is, editing, proofreading, all of those — then as a
group, I would like you to see if you can fill in the blanks on this hand-
out on five basic sentence patterns: how to find them, what questions
to ask. And we will go over this. Remember one section of your binders
should be sentence patterns.

This groupwork, so completely structured by the teacher, promotes neither
ownership nor coherent discussion.

In more autonomous groups, the teacher gives students some latitude in
their interactions with each other, and although students may remain on a
"short leash," the groupwork nonetheless displays spontaneous student inter-
actions concerning the substance of the lesson. In the most autonomous
groups, the teacher clearly defines group tasks but without prescripting
the groupwork. Typically the teacher (a) defines the goal of the group, for
example, arriving at a consensus concerning some controversial issue; (b)
outlines the tasks to be accomplished, for example, the group composition
of a letter outlining the group's views to a public official; and/or (c) assigns
roles to group members, for example, two group members argue one side of
the issue while two others argue the other side and one student acts as the
recorder. In the following transcript, for example, the teacher initiates
groupwork after students have written some original verse. Referring to the
students' poems, the teacher says:

If yours is the best it can be — instead of counting off, because we'll run
out of time, will you just group together with the three or four — three
or four maximum — people around you. Read them over, and choose
the one [poem] that . . . looks the most specific to you — see that person
that they're talking about.

A little later, this teacher reminds students that she wants
to hear what you've picked out in your groups so that you're hearing
some of the real strong images of these people. . . . I'm asking you to
brag; I want to hear the three or four real good ones, so people who
are having difficulty get a good sample before they go home.

This skillfully organized groupwork starts with expressive student writing,
thus promoting ownership. Not only because students must work together
to find good examples but also because the teacher encourages them to articu-
late what "good" means in this context, the ensuing group talk is more likely
to be coherent conversation than the first session, where students must "fill
in the blanks on this handout on five basic sentence patterns." Such discus-
sions were rare in our study: Only 11.1% of all small-group work was judged
either to be wholly autonomous or to display significant student interaction
in producing the outcome of groupwork, whereas 70.4% of all small-group
work was prestructured by the teacher.

The degree of student autonomy in the small groups we observed was
coded from audiotapes according to the following scale:

1 Teacher-structured groupwork: Task parameters entirely defined, or
"prescripted," by teacher. Task can be done without student interac-
tion (e.g., worksheets); group setting is gratuitous.
2 Prescripted task with obligatory student interaction.
3 Limited student interaction: Groupwork involves spontaneous student
interaction concerning substance but students are on "short leash." For
example, the teacher might define some general principle that
students in groups must then apply.
4 Significant student interaction defining shape of task and outcome, al-
though teacher might have been able to predict results before class.
5 Autonomous groupwork: Teacher sets up groupwork without pre-
scripting activities; significant student interaction define shape of task
and outcome. Results of groupwork cannot be predicted before class.

Student Production of Knowledge. In addition to promoting coherent
conversation, the second example encourages students to generate insights
and understandings far more than the first session. In activities such as com-
pleting worksheets and answering study questions, students are required
mainly to manipulate and master information provided by the teacher or a
textbook. This work calls only for lower-order cognitive activity, characteriz-
d by questions with prespecified right and wrong answers, and it tends not to
be very coherent because it does not build on student responses. It con-
sequently fails to result in student production of knowledge. Here is an
example.

Everyone put their name in the upper right. Put your name and your
period and the group number in the upper right-hand corner. Now I've
gone over two or three times your group number so you should remem-
ber it — when you get into your groups, if you have forgotten it, maybe
somebody in there will remember it. Now you are to combine eight sen-
tences to form the kind of sentence shown in parentheses, okay? And
I mean that . . . when I say a simple sentence, that's what I mean: a
simple sentence that has one subject and one predicate. When I say compound sentence, I mean two sentences of equal value and equal importance, put together with a conjunction or a semicolon or a transitive adverb, okay? Complex sentences: you will be asked to either write one with an adjective clause or an adverb clause. And because there were no noun clauses, I have given you a task down below here to write two sentences with noun clauses. Now, all of the sentences up above — the first eight — are all about Niagara Falls and the river and stuff like that, and though it isn't absolutely important that you focus on the Niagara River, it would be kind of nice if you did it, all right?

In another example, from another class, the teacher puts students to work in groups on a worksheet concerning Greek gods.

I know you won't finish the entire chart, but we will be starting . . . sharing, because what we want to be sure is that everyone has the same items on the chart. So I will give you the next 15 minutes . . . to just work on what you have. Be sure if you haven't—some of you maybe start from the back; we won't get them all checked today so that you have the chart. This is the chart that you will have to memorize, you will be tested on it . . . I want to make sure you all have the same information.

In both these examples, the groupwork involves little or no production of knowledge. In other groups, by contrast, students must sustain a coherent discussion in order to work out problems that generate new understandings. Typically, this groupwork proceeds in response to open-ended questions with more than one acceptable answer and involves higher-order cognitive activity. In the following transcript, for example, the teacher asks students to predict the ending of an Agatha Christie novel. To get them thinking, she first asks them to write a brief plot summary.

In a paragraph, write out what happened, what you think happened—and talk it out first, because there are a lot of lights going on in your eyes, and some of you are still saying, "I have no idea." Get together in your group, and on a sheet of loose-leaf tell me what happened.

As students work, the teachers says:

As I came around, a lot of you picked up a lot of the clues as to how she might have been killed and, so then, who would have done it. Whoever you're accusing, think about a motive—so some of you are saying, "Oh, the inspector did it," or "The colonel did it," or "Pollet did it," or "Ted did it," and that's all you're writing down. What I came around and asked you for was a motive, right? Miss Marple's got that down. So in your paragraph: who did it and a motive.

In another lesson, from another class, the teacher asks her students to analyze the characterization of Mr. Morrison in Roll of Thunder, Hear My Cry.

Here's what you need to do. . . . First of all, you want to name three outstanding character traits. Now, remember traits refer to personality, not physical characteristics. . . . Give supporting quotations for your ideas, one quote for each trait will be fine, and then give a warrant—that is, explain how the quote sets up that trait or how it establishes that trait, and then tell me what technique [the author] is using. Does she use a character's language, a character's actions, or do you see the character through the eyes of another character, or the reaction of another character?

In each of these latter two examples, students must not only identify some underlying principle—a motive for murder, a character trait—but also find supporting evidence for the interpretation. Once again, we discovered that such groupwork is infrequent; only about a quarter of the groupwork we observed involved discussion of open-ended questions with students actively constructing interpretations (rated a 5 on the scale in the list that follows). By contrast, two-thirds was a version of collaborative seatwork.

In our analysis of the extent to which students were required to actively construct new understandings during small-group work, we coded each session, using audiotapes, according to the following scale:

1 Collaborative seatwork. Students receive only prespecified information; student activities in small groups are tightly controlled. Worksheets are an example.

2 Students primarily receive prespecified information; occasionally their tasks involve open-ended questions. Student activities are highly orchestrated, with topic coverage defined by the teacher.

3 Teacher identifies issue or problem; students find examples and explore applications. Teacher provides time and encouragement for exploring new meanings and implications.

4 Students receive some information; much of the groupwork involves students using information with open-ended questions or with teacher-provided questions.

5 Students identify both problems and applications. Discussing open-
ended questions, students actively construct interpretations. Teacher sets the parameters of the groupwork, but it is mainly the students who work out ways to address issues and answer questions.

**Effective and Ineffective Groupwork.** Did small-group work affect achievement, holding other conditions constant? In our regression analysis, we held constant students' sex, race/ethnicity, sses, and fall reading and writing skills. We also took into account teachers' uses of authentic and follow-up questions, the amount of discussion time outside of small groups, as well as rates of student off-task behavior and homework completion.

We first discovered that, overall, classes spending more time in small groups produced lower achievement, a finding that replicated the surprising conclusion of our eighth-grade study. However, we were now prepared to go beyond simply asking how much time was spent in small groups: We could ask whether the effectiveness of small-group time depended on what was going on in the small groups. To do this, we used the two measures of the quality of group time discussed earlier: student autonomy and student production of knowledge. Although these measures differ in theory—students could be given autonomy but not take the opportunity to produce knowledge—in practice the two tended to occur together: the more student autonomy, the more production of knowledge (the correlation was .78).^1^6

Analysis of student autonomy showed that the higher the degree of autonomy, the more likely group time was to contribute positively to achievement. For example, a class that averaged 5 minutes per day of highly "prescribed," rigidly structured group time, with little autonomy for students, would actually score about 1 point lower on the test than a class with no group time at all. By contrast, a class that averaged 5 minutes of highly autonomous groupwork, in which students worked together to define the task, would gain almost 2 points on the test over a similar class with no groupwork. This is a significant effect: It could move a student from the fiftieth to the fifty-eighth percentile on the 32-point test. We found similar results for student production of knowledge: Collaborative seatwork actually reduces achievement, but groupwork in which students actively construct interpretations promotes achievement.

These results are consistent with Hillocks's (1986) study, which examined dozens of empirical studies conducted over a quarter century involving nearly 12,000 students and found that the most effective writing instruction involved peer-response groups with what Hillocks called an "inquiry" focus: assigned topics involving analysis of readings or other data and attention to rhetorical strategies.

By replicating our eighth-grade study and moving beyond it, we accounted for the unexpected results. Small-group work appeared ineffective because the groups were being used ineffectively; many of the assignments could just as well have been done individually. The follow-up study shows that when small-group time allows students to interact over the substance of their problem, defining tasks as well as solutions and constructing interpretations, students benefit from the opportunity to work in small groups.17

These findings reaffirm the view that effective small-group work requires coherent activities that result in the sustained production of student knowledge. To promote such activities, the teacher must not overly specify group tasks. In other words, effective teachers clearly define the general parameters of the tasks, but not the precise character of the activities themselves. Teachers who promote thinking about literature may present clear objectives to students for groupwork—for example, identifying the best poems written by students in each group, articulating character traits, and finding supporting quotations—but avoid telling groups exactly how to proceed; they do not, for example, specify a list of questions and topics students must answer in a particular order.

When teachers put groups of students together to work on some problem, they send students an important message that developing their own thoughts is important. The teacher above who told her students, "There are a lot of lights ... in your eyes," as she prepared them for groupwork sent exactly this message.

The benefits of direct instruction presumably result from higher on-task behavior when the teacher works directly with students. One could argue that this occurs when the instructional task is the same for both whole-class and small-group settings. For example, students may be more on task when they answer recitation questions as a whole class than when the recitation questions are assigned to small groups for written responses. Learning, in this example, may not differ or may be greater in the whole-class format. However, this format does not take advantage of the opportunities for intellectual collaboration that are made possible by the small-group setting. If the tasks are the same, one should expect little difference in achievement; the point of small-group instruction, however, should not be to assign the same tasks, but to design work that draws on the potential for cooperation and collaboration in the small group.

If successful small-group work depends on the teacher setting up open-ended rather than prescribed tasks and on the students having coherent conversations generating insights, then teachers must carefully match small-group work to suitable tasks. For example, if the objective for a given day requires presenting a lot of new information, a lecture is probably better than small-group work. If teachers want students to practice some particular skill, recitation and seatwork may be better than small-group work. On the other hand, if teachers want students to compare ideas, develop a train of thought,
air differences, or arrive at a consensus on some controversial issue, then the
forum of small groups may be just the right setting for most students to carry
on intensive conversation and discussion, especially students who are too shy
to say much in the larger setting of the whole class. Teachers must always
remember, however, that they cannot just put students in groups and expect
them to "go to it" with positive results. For groupwork to succeed, teachers
must carefully design collaborative tasks that are interesting to students (and
not just to the teacher). They also must be prepared to help students develop
effective group skills; an excellent guide for teachers is Cohen's Designing
Groupwork (1986).

THE BOTTOM LINE: LEARNING TO THINK REQUIRES
EFFECTIVE INTERACTION

Most of the time when teachers ask students questions, they are not asking
to be informed since they already know the answers themselves. When
students are asked to recite for teachers who have no need to be informed,
they produce "pseudo-discourse." Authentic discourse occurs only when
some information or interpretive stance is really at issue. Only authentic dis­
course can engage students.

Yet the results of our study suggest that authentic questions, discussion,
small-group work, and interaction, though important, do not categorically
produce learning; indeed we observed many classes where this was not the
case. We also found that recitation is not categorically ineffective; rather,
it’s effectiveness varies depending on whether and how teachers expand IRE
sequences. The underlying epistemology of classroom interaction defines the
bottom line for learning: What ultimately counts is the extent to which instruc­
tion requires students to think, not just report someone else’s thinking. As Leont’ev
(1981) might put it, the pedagogical usefulness of each interaction needs to
be gauged in terms of the particular activity or project involved and, more
than this, the system of social relationships that support it. Authentic ques­
tions, discussion, and small-group work have important instructional poten­
tial, but unless they are used in relation to serious instructional goals and,
more important, unless they assign significant and serious epistemic roles to
students that the students themselves can value, they may be little more than
pleasant diversions.

In one English class in our study, for example, students engaged in several
ostensibly open-ended, imaginative writing tasks requiring them to write
from the point of view of a pencil eraser or a bullet. Upon close inspection of
these exercises and especially the teacher's responses to the papers, we eventu­
ally came to understand that, from the teacher’s point of view, the content
of student responses to these prompts—imaginative or not—was irrelevant;
early the only thing the teacher responded to in his marginal comments was
whether or not all words had been spelled correctly. As it happened, the stu­
dents in this class understood the operational rubric for this exercise and
played their roles more or less appropriately. However, the ostensible purpose
(imaginative discourse) and the actual purpose (correct spelling) of the writ­
ting tasks significantly differed; the writers were not really speaking to a reader
who was listening to what they were trying to say. These writing tasks were
what Bloom and Argumeda (1983) call procedural displays. Students can be
involved fully and substantively in reciprocal instruction only when the
ostensible purpose of the discourse is the same as the actual purpose.

Generally, we may say that reciprocity in instruction occurs most often
when students, as well as the teacher, have some input into and control over
instructional discourse, and when their previous learning significantly affects
the course of subsequent learning. This concept has been implemented in
some elementary reading instruction by Palincsar and Brown (1984) in what
they call reciprocal teaching. In reciprocal teaching, students take turns being
the teacher. In other classrooms where students do not play the role of
teacher, the teachers nonetheless honor the terms of reciprocity when they
avoid prespecifying answers to their questions so that student answers poten­
tially can affect subsequent questions and discussion. When teachers ask gen­
tuine questions of this sort, they treat students as full-fledged conversants.

This is not to suggest that the dialogicality of instruction can be judged
in terms of the how of instruction—question-answer sequences evidenced
in face-to-face interaction—alone. The study found that the what of instruc­
tion—the content and subject matter—is critical to learning as well. Authen­
tic questions must challenge students to think and reflect on the conse­quences
of their ideas, not just remember their past experiences. Teachers must prize
vigor notice discussion in dialogic classrooms, encouraging what Bakhtin (1981)
calls the struggle of "contradictory opinions, points of view and value judg­
ments" (p. 281). Bakhtin teaches us that meaning and therefore learning—
understood as the expansion of a personally coherent interpretation of infor­
mation and events—are actively constructed and negotiated through lan­
guage use. Learning is a dialogic event, part the instructor's and part the learn­
ner's. If events and information are to acquire meaning and students are to
learn, then teachers must think of curriculum not just in terms of points to
be made, information to be conveyed, and abstract skills to be mastered.
Rather, they must engage students in activities and projects bridging the pur­
poses of their students and the goals of instruction. Yet instruction creating
such bridges, we found, was either rare or altogether missing in the classes
we observed, which overwhelmingly were given over to lecture, recitation,
short-answer questions, and seatwork.
The Bakhtinian explanation for the relative ineffectiveness of monologic instruction in promoting learning and conceptual change, compared with discussion and instructional conversation (even for basic objectives such as recall and literal comprehension), is that meaning "is realized only in the process of active, responsive understanding" (Vološinov, 1973, p. 102). Yet it is just such active, responsive understanding that teachers fail to practice — especially in low-track classes — when they determine prior to a given class the entire sequence of questions they will ask and what answers they will accept, and when they respond to correct student answers with a mere nod before moving on to the next question, often changing the topic of discourse. In doing so, these individuals make no attempt at active, responsive understanding; they "want, in effect, to turn on a light bulb after having switched off the current. Only the current of verbal intercourse endows a word with the light of meaning" (Vološinov, 1973, p. 103).
APPENDIX A

Sample Literature Familiarity Test

1. For each of the following stories that you have read this year, indicate, by checking the correct box, whether each had a happy ending or a sad ending.

   b. “The Fifty-first Dragon”
   c. “The Most Dangerous Game”
   d. *Of Mice and Men*

2. Name as many characters as you can remember for each story. If you can’t remember their names, briefly describe as many as you can.

3. For as many stories as you can remember, briefly explain how each story ended. Write no more than two sentences for each story.

4. For each story that had a conflict you can remember, briefly explain what this conflict was.

   Briefly relate the conflict in *Of Mice and Men* to the ending of the story.

   Briefly relate this conflict in *Of Mice and Men* to the theme of the story.

5. For each story that had a theme, or main idea, that you can remember, briefly explain what this theme was.

   Briefly give three reasons why you think the theme of *Of Mice and Men* is what you say.
APPENDIX B

Scoring Information for Literature
Familiarity Test

Categories A–H describe dimensions of interpretive ability. Information from the complete test should be used to derive a score for each.

INFORMATION

A. Extent of recall (absolute scale)

0  None — no detail recognizable from any stories
1  Vague — very few details recognizable for only one or two stories
2  Some details regarding some or most of the stories
3  Many details regarding more than half the stories; richly detailed for less than half the stories
4  Richly detailed — many details regarding more than half of all the stories

B. Depth of understanding (absolute scale)

0  Never goes beyond literal understanding; empty generalization
1  Goes beyond literal understanding for only one story
2  Some evaluation and interpretation of more than one story
3  Some interrelation of story elements in more than one story
4  High interrelation of story elements in most or all stories

C. Number of endings from stories studied in class remembered (absolute scale)

0  None
1  One story
2  Two stories
3  Three stories
4  Four stories
5  Five stories
LITERARY UNDERSTANDING

D. Relates ending to denouement

FOR STUDENTS ANSWERING FOR ALL STORIES
1 For either no stories or one (barely)
2 For either one really well or two
3 For three stories
4 For four or five stories

FOR STUDENTS ANSWERING FOR HALF OR LESS
OF THE STORIES
1 Almost never or not at all
2 Most of the time
3 Consistently

E. Relates conflict and/or ending to theme

FOR STUDENTS ANSWERING FOR ALL STORIES
1 For either no stories or one (barely)
2 For either one really well or two
3 For three stories
4 For four or five stories

FOR STUDENTS ANSWERING FOR HALF OR LESS
OF THE STORIES
1 Almost never or not at all
2 Most of the time
3 Consistently

F. Intuits internal motivations of characters

FOR STUDENTS ANSWERING FOR ALL STORIES
0 For no stories
1 For only one story
2 For less than half the stories
3 For more than half the stories
4 Four out of five or all

FOR STUDENTS ANSWERING FOR HALF OR LESS
OF THE STORIES
0 For no stories
1 For less than half the stories
2 For more than half the stories
3 Consistently

G. Interpretive treatment of major selection

0 No response
1 Literal/superficial
2 Literal/substantial
3 Some interpretation
4 Sophisticated and original/substantial

H. Level of discourse regarding theme and conflict

0 No response
1 Cliche or truism
2 Literal
3 General
4 Universal

Note: Score level of discourse the high point reached; that is, if one story is treated literally and two are treated in terms of cliche, record a score of 2 (for literal). Level of discourse should be scored as general if the student goes "beyond the information given" to treat characters, plots, conflicts, etc., in general terms, i.e., offers more than a literal rendition of events, but nonetheless limits these generalizations to the story itself (e.g., statements about what characters are generally like in the story). By contrast, level of discourse should be scored as universal if the student seeks to apply such generalizations beyond the story to experience in general and does so in well developed terms, making an explicit case for the universality of the generalization.
struggle and tension between self and other—"the differential relation between a center and all that is not that center" (Holquist, 1990, p. 18)—which in their view is why conflict is essential to the meaning we give our experience and the understandings we have of it. "Consciousness takes shape and being," Volosinov (1973) claims, "in the material of signs created by an organized group in the process of its social interaction" (p. 13).

8. The interactive character of discourse is evident even when the conversants fail to interact: in the conversational management of misunderstandings. When conversants misunderstand one another, they must backtrack to some point in the discourse where they were in synch with each other and properly repair the trouble source; for example, the listener asks, "What?" The ease with which speakers usually explain the "what"—readily, without first asking what was unclear ("What do you mean by 'what'?")—means that speakers consistently monitor their utterances in terms of what their listeners already know.


10. As Klnacker (1989) explains, "Bakhtin's crucial starting point—the diversity of practical languages rather than a unitary abstract structure—leads him to argue that every effort to impose unity on these languages is 'monologic.' The institutions of the school, the state, and the church enforce monologic languages as the voice of culture, the voice of authority, the voice of God ventriloquized through the literary critic, the politician, or the priest. His terms dialogic and monologic thus describe uses of language rather than inherent properties of language itself" (p. 85).

11. James Britton (1969, 1970) argued that conversants are most likely to discover insights when their talk is expressive.

12. "Stratification, diversity, and randomness [i.e., heteroglossia] is (sic) not only a static invariant in the life of language, but also insures its dynamics" (Bakhtin, 1981, p. 272).

13. This formulation follows Halliday's (e.g., 1978) conception of register as a unique configuration of tenor (relationship between conversants), field (what discourse is about), and mode (channel [i.e., written or spoken] and genre).

14. "The understanding of any sign [i.e., learning], whether inner or outer, occurs inextricably tied in with the situation in which the sign is implemented... It is always a social situation" (Volosinov, 1973, p. 37; emphasis in original).

15. In an analysis of 36 hours of instruction in six different classes, Corey (1940) found that the average student utterance was 11 words long (cited in Hoetker & Ahlbrand, 1969, p. 159).

16. Popper (1972) called this World-3 knowledge, distinguishing it from World-1 knowledge of the physical world and World-2 knowledge of experience and thought.

Notes

The section "Small Groups" is reprinted with editorial changes from "Using Small Groups for Response to and Thinking about Literature," by Martin Nystrand.
10. More than half of the eighth-grade classes (58.6%) spent no time at all in discussion; 20.7% spent 1 minute or more on average; only two classes of the 58 regularly spent 7 minutes or more. In ninth grade, 61.1% of all classes had no discussion at all, and only 5.6% had more than a minute daily; only one class of the 54 averaged more than 2 minutes.

11. In grade 8, 91.4% of all classes spent no time in small-group work; only 5 out of 58 classes involved small-group work, and this ranged from an average of 1.75–13.50 minutes. In grade 9, 63% of all classes had no small-group work at all; only 11.1% spent more than 5 minutes daily; and only four classes spent more than 10 minutes daily in small groups.

12. In 36.2% of the eighth-grade classes, teachers asked 5% or fewer authentic questions. Only 6.9% of classes involved more than 30% authentic teacher questions; one class of the 58 involved more than 50%. In 25.9% of the classes, uptake was present 5% or less. Only 8.6% of classes involved more than 20%; two classes involved 28%.

13. In 90% of all classes, one or more class teachers on average asked 83% authentic questions. Uptake was noted in all eighth-grade classes, although 5.6% of classes involved less than 10%. Half the classes had at least 25% uptake; and 13% of classes involved 40% or more uptake.

14. Four of these titles were chosen as a stratiﬁed sample to represent the kinds of literature each class had studied; if half of the titles studied were short stories, then two of the four were short stories, and so on. The fifth selected title was the one work the class had spent the most time on; typically it was either a novel, such as *To Kill a Mockingbird* or *A Tale of Two Cities,* or a drama, such as *Romeo and Juliet.*

15. Information on student characteristics came from student questionnaires. Race and ethnicity were coded when students identiﬁed themselves as Black or Hispanic. Family socioeconomic status was indicated by an unweighted additive composite of father’s and mother’s education, the higher of either parent’s occupational status, and a list of home resources, as reported on student questionnaires. Fall reading skills were measured by a *NAEP* multiple-choice test of reading comprehension, and fall writing skills were measured by a holistically scored writing sample. Fall writing skills were indicated by a writing sample that was scored by two readers, whose marks were averaged, for (a) level of abstraction, based on Britton and colleagues’ (1975) categories of transactional-informative prose; and (b) coherence and elaborateness of argumentation, based on the 1979/1984 *NAEP* criteria for informative writing (in Applebee, Langer, & Mullis, 1985). Each student’s writing score was the sum of these two measures. Interrater reliability of scoring this test was .68. (For further details on the background variables and achievement controls, see Gamoran & Nystrand, 1991).

16. The relationship between student autonomy and knowledge production has been found in many studies, including studies examining small groups; see King (1992) and Palincsar, David, Winn, Stevens, and Brown (1990).

17. Some small-group work rated lowest on autonomy and student production of knowledge concerned grammar and sentence errors, not literature. All of the highest-rated small-group work concerned literature. We are unable to rule out, for the lowest-rated, small-group sessions, the possibility that their focus on grammar and sentence correction rather than literature, accounted for the negative impact of collaborative seatwork on literature achievement.

CHAPTER 3

We wish to acknowledge Sarah Bing-Prineas’s invaluable help in the early stages of this research.

1. These predictions were the result of a residual analysis based on instructional and background variables outlined in Chapter 2.

2. Students in classes whose format most closely approximated conversational tended to score higher on the literature achievement test, regardless of the actual number of student or teacher questions.

CHAPTER 4

Parts of the sections “Dialogism and Students at Risk” (pp. 104–105) and “Literature Instruction” (pp. 105–108) are reprinted with editorial changes from “High School English Students in Low-Ability Classes: What Helps?” by Martin Nystrand, *The Newsletter of the National Center on Effective Secondary Schools,* January 1990, pp. 7–8.

1. “The meaning of a word is determined entirely by its context. In fact, there are as many meanings of a word as there are contexts of its usage” (Volosinov, 1973, p. 79). Compare also Volosinov’s (1973) observation that “the forms of signs are conditioned above all by the social organization of the participants involved and also by the immediate conditions of their interaction. When these forms change, so does sign” (p. 21; emphasis in translation).

2. Volosinov (1973) argues that thought, or inner speech, “resemble[s] the alternating lines of a dialogue . . . joined with one another and alternating[ing] with one another not according to the laws of grammar or logic but according to the laws of evaluative (emotive) correspondence, dialogical deployment, etc., in close dependence on the historical conditions of the social situation and the whole pragmatic run of life” (p. 38; emphasis in translation).

3. When writers fail to elaborate potential trouble sources or do so inadequately, the result is “misconstraint,” that is, a mismatch between what the writer has to say and what the reader needs to find out. If the topic is inadequately elaborated, the reader will find the text *ambiguous* (What’s this about?). If the writer inadequately elaborates what is said about the topic, the reader will find the text *abstruse* (What’s the point!). For more, see Nystrand (1986, Ch. 3).