



Discussion-Based Approaches to Developing Understanding: Classroom Instruction and Student Performance in Middle and High School English

Arthur N. Applebee; Judith A. Langer; Martin Nystrand; Adam Gamoran

American Educational Research Journal, Vol. 40, No. 3. (Autumn, 2003), pp. 685-730.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8312%28200323%2940%3A3%3C685%3ADATDUC%3E2.0.CO%3B2-1>

American Educational Research Journal is currently published by American Educational Research Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aera.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

Discussion-Based Approaches to Developing Understanding: Classroom Instruction and Student Performance in Middle and High School English

Arthur N. Applebee and Judith A. Langer
University at Albany

Martin Nystrand and Adam Gamoran
University of Wisconsin, Madison

This study examines the relationships between student literacy performance and discussion-based approaches to the development of understanding in 64 middle and high school English classrooms. A series of hierarchical linear models indicated that discussion-based approaches were significantly related to spring performance, controlling for fall performance and other background variables. These approaches were effective across a range of situations and for low-achieving as well as high-achieving students, although interpretations are complicated because instruction is unequally distributed across tracks. Overall, the results suggest that students whose classroom literacy experiences emphasize discussion-based approaches in the context of high academic demands internalize the knowledge and skills necessary to engage in challenging literacy tasks on their own.

KEYWORDS: classroom discourse, English language arts, instruction, literacy, secondary education.

Since the early 1980s, many scholars who study the teaching of English and language arts have been engaged in an extended exploration of the nature of effective instruction. Often grounding their work in sociocognitive¹ theory about the nature of teaching and learning (Langer, 1985a), these researchers have drawn from disciplines ranging from anthropology (Heath, 1983) to psychology (Lee & Smagorinsky, 2000), to literary theory (Scholes, 1985), to linguistics (Cazden, 1979, 1988) in their efforts to understand the nature of classroom contexts that foster the development of higher levels of literacy among diverse groups of students. A variety of studies have examined what teachers know and how that relates to the instruction that they provide (e.g., Grossman, 1990); what students are actually asked to do in English language arts classes (e.g., Freedman, Simons, Kalnin, Casareno, & the M-Class

Teams, 1999); and how these activities influence what students learn (e.g., Hillocks, 1999). Our own research over the past 20 years has been part of this larger exploration of the nature of effective instruction. In separate studies, we have examined the teaching of writing (Applebee, 1981, 1984; Langer & Applebee, 1987; Nystrand, 1986; Nystrand, Gamoran, & Carbonaro, 2001); the relationship between writing and reading (Langer, 1986); the teaching of literature (Applebee, 1993; Langer, 1995; Nystrand, 1997; Nystrand & Gamoran, 1991); the structure of curriculum (Applebee, 1996); the effects of tracking on achievement and equality of opportunity (Gamoran, Nystrand, Berends, & LePore, 1995); and the roles of teachers and students in effective classroom interaction (Langer, 1995, 2001, 2002; Nystrand, 1997).

In the present study, we sought to take work in this area further by explicating commonalities among our previous findings and those of other researchers who share a focus on the importance of discussion-based approaches in teaching for in-depth understanding (Bransford, Brown, & Cocking, 1999). We sought to use those commonalities as a base from which to examine whether classrooms that emphasize such approaches are more successful than other classrooms in improving students' literacy performance. In a field where much of the research, including much of our own, has relied on case study methodologies and descriptive analyses, an additional goal was to examine how findings from previous work would hold up when examined in a relatively large-scale quantitative study, in a broad spectrum of middle and high school classrooms. In the sections that follow, we discuss the results of previous work and then examine literacy performance at the end of an academic year (controlling for initial performance) among middle and high school students whose classrooms varied in ways that earlier studies have suggested are related to performance differences.

ARTHUR N. APPLEBEE is a Professor of Education at the University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222; e-mail aapplebee@uamail.albany.edu. His area of specialization is the intersection of language and learning, particularly in the teaching of English language arts.

JUDITH A. LANGER is a Distinguished Professor at the University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222; e-mail jlanger@uamail.albany.edu. She specializes in research in literacy development and instruction.

MARTIN NYSTRAND is a Professor of English at the University of Wisconsin, Madison, 600 N. Park Street, Madison, WI 53706; e-mail nystrand@ssc.wisc.edu. His research examines the role of discourse, in both writing and classroom interaction, in English language arts instruction and learning.

ADAM GAMORAN is a Professor of Sociology and Educational Policy Studies at the University of Wisconsin, Madison, 1180 Observatory Drive, Madison, WI 53706; e-mail gamoran@ssc.wisc.edu. His research focuses on inequality in educational achievement and the organizational context of school reform.

Exploring the Nature of Effective Teaching and Learning in English

At times it may seem as if English language arts have always been at the center of a curriculum designed to develop a literate and cultured populace; that the abilities to read with ease, write fluently, think deeply, and communicate effectively were always its goals. However, the field of English language arts emerged as a separate—and major—school subject only in the 1890s, following the report of the Committee of Ten on Secondary School Studies (1894). It replaced a wide variety of specialized courses dealing with various aspects of English, including rhetoric, literary history, grammar, spelling, composition, and oratory, which had in turn evolved out of, or parallel with, previous courses in Latin and Greek (Applebee, 1974). The goals of English instruction have always been diverse, involving mastery of virtually every activity connected with the use of language, but there has been a consistent emphasis on the development of high-level literacy skills (reading and writing) in the context of the extended study of literature. Until the 1970s, however, these skills were treated as rather unproblematic: Students either did or did not comprehend a text and were either able or not able to transcribe their thoughts into effective writing. Thus early studies of instruction looked at relatively simple factors: how much time should be devoted to writing instruction, for example, or whether it was better to write independent “daily themes” or to engage in longer writing tasks that would take more than a day to complete (cf. Braddock, Lloyd-Jones, & Schoer, 1963).

A major shift in the focus of discussions about teaching and learning occurred in the last decades of the 20th century, reflecting a new interest in the cognitive and linguistic processes underlying performance. Within the field of English education, the shift is usually dated from (although not necessarily ascribed to) the publication of Emig’s (1971) case studies of the composing processes of 12th-grade students. The empirical character of Emig’s monograph, derived from her dissertation, was shaped by the emerging interest in cognitive studies in Cambridge, Mass., where she was a graduate student at Harvard from 1965 to 1969. Using think-aloud protocols, Emig traced the evolution of students’ ideas while they were writing and showed how that evolution illustrated the recursive, complex cognitive and linguistic processes that writers engage in as they struggle with both content and form.

Emig’s work was one of a number of studies that transformed the ways that literacy tasks were viewed, no longer as simple processes of transcription and decoding but, rather, as extended processes of composition and comprehension, during which the understandings of readers and writers develop and change. A groundswell of interest in cognitive research appeared at a variety of institutions, including the Center for Cognitive Studies at Harvard University (Bruner, Olver, & Greenfield, 1966); the Center for the Study of Reading at the University of Illinois, Urbana-Champaign (Anderson, 1977); Carnegie Mellon University (Gregg & Steinberg, 1980); and the University of California, Berkeley (Fillmore, 1982). At the same time, research in language and literacy, in parallel with changes in many other fields, moved gradually from an

examination of reading and writing processes (Adams & Collins, 1979; Flower & Hayes, 1981; Shaugnessy, 1977) to the study of literacy tasks embedded in classrooms (Applebee, 1984; Durkin, 1978–1979; Dyson, 1984; Graves, 1983; Tharp & Gallimore, 1988), to the study of how the activities of reading and writing are defined by the wide range of contexts in which people participate (Brandt, 1998, 2001; Heath, 1983; Street, 1995).

As research influenced by the cognitive revolution developed, so too did the basis for arguments about the nature of effective instruction. Early cognitive studies simply extrapolated instructional implications from studies of the performance of individuals or studies of contrasts between the performance of experts and that of novices (Flower & Hayes, 1980; Perl, 1979). Later studies looked directly at learning and instruction, examining, for example, how the social roles and expectations in individual classrooms shaped thinking as well as literacy learning (Langer, 1995, 2002; Lee, 2001; Rex, 2001; Wells & Chang-Wells, 1992). Emig (1971), at the beginning of this tradition, provided a good illustration of the assumptions in the early work: She criticized traditional approaches to writing instruction based on her analysis of her case study students' writing performance, but she did not examine classroom instruction. In contrast, more recent research has looked in great depth at individual classrooms and the differing kinds of learning that students develop depending on the emphases of their teachers (Athanases, 1998; Hillocks, 1999; Sperling & Woodlief, 1997).

As research in this area became more sophisticated, it began to move across traditional disciplinary boundaries, looking at how the social processes of the classroom create the context within which individuals develop the cognitive and linguistic processes—the tools for comprehension and understanding—associated with literacy. Noting the conjunction of issues related to social processes, learning, and development—issues usually studied in separate disciplines—Langer (1985a, 1987a) called for researchers to take a sociocognitive view of language and literacy learning. Such a view recognizes that the contexts within which literacy is used and learned lead to particular ways of thinking and doing—that culture (including the culture of the classroom), language, and cognition are inextricably intertwined. This recognition, in turn, implies that the social processes of the classroom and the individual development of students need to be examined simultaneously, with the ultimate goal of a better understanding of the nature of teaching and learning.

The starting point for sociocognitive studies of literacy instruction has been in the exchange of ideas—the discussion—that takes place in a classroom, including reading, writing, and the talk that surrounds them. A variety of investigators have argued that high-quality discussion and exploration of ideas—not just the presentation of high-quality content by the teacher or text—are central to the developing understandings of readers and writers (Alvermann et al., 1996; Eeds & Wells, 1989; Gambrell & Almasi, 1996; Guthrie, Schafer, Wang, & Afflerbach, 1995). Previous studies also suggest that non-mainstream students—low achievers, children of the poor, and second-language learners—fare poorly in classrooms with traditional instructional

approaches, which are structured in ways that fail to capitalize on these students' strengths and instead magnify their weaknesses (Gutierrez, 1994; Heath, 1983; Hynds, 1997; Marshall, Smagorinsky, & Smith, 1995). Such students typically do much better when instruction builds on previous knowledge and current ideas and experiences, permits students to voice their understandings and refine them through substantive discussion with others, and explicitly provides the new knowledge and strategies that students need to participate successfully in the continuing discussion (Langer, Bartolome, Vasquez, & Lucas, 1990; Langer, 1992, 2001, 2002; Lee, 1993, 1995, 2001; Ladson-Billings, 1995; Losey, 1995; Rex, 2001; Freedman et al., 1999).

A wide range of studies has documented the fact that the typical pattern of classroom discourse is one-sided, following a pattern of teacher question, student response, and teacher evaluation of the response (Cazden, 1988; Mehan, 1979). This I-R-E (initiation, response, evaluation) pattern is usually intermingled with lectures or presentations by the teacher and seat work, in which students work individually on study activities, often in workbooks or worksheets. Viewed from a sociocognitive perspective, such instruction places a premium on transmission of information, providing very little room for the exploration of ideas, which is necessary for the development of deeper understanding. The ideas that matter in this form of instruction are those of the textbook and teacher; the students' understanding of them is tested through the I-R-E pattern of recitation and through the completion of study materials.

During the 1960s and 1970s, a series of process-product studies attempted to understand effective instruction by finding behavioral correlates of desired outcomes (cf. Gage, 1963, 1994; Gage & Needels, 1989), including aspects of classroom interaction, such as calling on a wide range of students, responding positively to what students say, or asking higher-order questions. This program of research has been criticized for its behavioristic emphasis on low-inference observations of behaviors lacking broader theoretical significance, and for a corresponding lack of attention to classroom dynamics and the intentions of the participants (cf. Garrison & Macmillan, 1984, 1994). The results of such research often yielded findings that were seemingly obvious—for example, that more time devoted to productive instruction makes learning more likely (academic learning time). Moreover, this kind of research translated poorly into classroom practice (Garrison & Macmillan, 1994). To induce teachers to adopt certain low-inference classroom behaviors, for example, is relatively easy; but if they lack a broader framework for thinking about what the behaviors are intended to accomplish, they are likely to adopt them in idiosyncratic, formulaic, and ineffective ways (Applebee, 1976).

Other studies of classroom discourse have situated their observations of classroom interaction within more explicit theoretical and interpretive frameworks, often drawing on ethnographic or social-linguistic methodologies (cf. Cazden, 1988; Heath, 1983; Mehan, 1979). Nystrand (1997) drew on a wide variety of such studies to specify a number of features of classroom discourse that he argued were related to the overall epistemology of learning that would support the development of deep understanding in reading and writing. In

particular, he focused on Bakhtin's (1981) notion of dialogic interaction as essential to such discussion, contrasting it with the monologic interaction patterns of the typical recitation. Focusing on questions as sites of interaction, Nystrand examined the participants' understandings of their interactions as manifested in their discourse moves. In his research, for example, the significance and purpose of a question depended not only on how students responded to the question but also on how the teacher evaluated or followed up on students' responses. The focus was on the character of the discussion elicited and valorized by the questions that were asked. Studying more than a hundred eighth- and ninth-grade English classrooms in Wisconsin and Illinois, Nystrand (1997) used a variety of indexes of dialogic interaction to frame an investigation of the relationship between classroom language and literacy learning. As part of the study, Nystrand developed the Classroom Language Assessment System (CLASS), a real-time computer-based data collection system, to analyze the kinds and the patterning of questions teachers asked. Using pretests of literacy performance and an end-of-year measure that focused on the unique literature selections that each classroom had studied, Nystrand found that a variety of features of classroom discussion activities were significantly related to spring performance. The features that were associated with larger improvements in performance over the year included

- More use of authentic questions, which were used to explore differing understandings rather than to "test" what students might already know;
- More time for open discussion: whole-class discourse devoted to free exchange of ideas among students or between at least three participants; and
- More "uptake," in which a teacher's question "took up" and built on a student's previous comment, creating continuity in the discourse.

Treating these measures as indexes of an underlying classroom epistemology that stressed the value of exploration of ideas and the development of understanding through discussion, Nystrand argued that, for the most part, his results supported the importance of dialogic, as opposed to monologic, approaches to instruction.

Whereas Nystrand began his investigations of dialogic instruction with a fine-grained analysis of patterns of classroom interaction, Langer began by examining the cognitive and linguistic processes drawn upon by successful readers and writers as they developed their understandings in both literary and informational texts (1986, 1987b, 1990). Langer then looked at the broader features of classroom contexts that supported the development of such processes (1992, 1995). Arguing that much previous work had misrepresented the ever-changing nature of a reader's or writer's understanding of a text, Langer called this understanding an "envisionment" of an evolving text world (cf. Fillmore, 1982; Kay, 1987; Langer, 1985b) and used the term "envisionment building" to describe the cognitive and linguistic processes drawn upon by readers and writers. Rather than say simply that students "comprehended"

Discussion-Based Approaches to Developing Understanding

or “did not comprehend” what they were reading or writing about, Langer found that students’ envisionment of a text at any time was a mixture of understandings, questions, hypotheses, and connections to previous knowledge and experiences. She found that the envisionment changed and evolved with further reading, writing, discussion, or reflection. She also found that literature and exposition required very different sets of cognitive and linguistic strategies: Literary understanding generated an ever-moving horizon of possibilities, whereas understanding of exposition focused on a point of reference that was elaborated and clarified as the reading or writing continued. Differences in strategies for the two types of text led her to an extended investigation of the kinds of activities that teachers could use to help students become more effective envisionment builders, particularly when reading and writing about literature.

Like Nystrand, Langer stressed the importance of classroom discussion in the development of understanding. But rather than focus on overt behaviors during classroom interaction, she focused on the conventions established within the classroom as to what constituted effective and appropriate discussion. She found that envisionment-building classrooms—those that provided activities particularly effective in the development of students’ reading and writing skills—displayed a variety of discussion-based approaches to the development of understanding. Langer (1995) identified the following classroom practices as particularly important in supporting students’ meaning making:

- Teachers treat all students as capable envisionment builders with important understandings and potential contributions to classroom discussion;
- Teachers use instructional activities such as discussion to develop understandings rather than to test what students already know;
- Teachers assume that questions are a natural part of the process of coming to understand new material, rather than an indication of failure to learn, and that questions provide productive starting points for discussion; and
- Teachers help students learn to examine multiple perspectives (from students, texts, and other voices) to enrich understanding rather than focusing on consensus interpretations.

Langer found that the more than 50 teachers who participated in her study developed a wide range of discussion-based strategies (varying from whole group to small group, to workshop formats, for example) that reflected the teaching styles with which they were comfortable and the needs of their students. They all orchestrated these individual instructional approaches, however, to provide effective scaffolding for their students’ envisionment building as the students learned to deal with challenging texts or undertook challenging writing assignments.

In a later study, Langer (2000, 2001) examined English programs that regularly “beat the odds” on large-scale assessments of English achievement,

outperforming other programs that served demographically similar populations of students. She based her findings on case studies that followed 44 English teachers in 88 classrooms, in 25 schools in four states, for a 2-year period. Many of her findings focused on larger institutional issues, such as the ways in which a school and its district supported the professionalism of the teachers. But building on her work on envisionment-building classrooms, she also found a variety of characteristics across all facets of English and language arts instruction that related to discussion-based approaches, including an emphasis on depth rather than breadth of knowledge; the use of discussion to develop depth and complexity of understanding; overt teaching of knowledge and strategies needed for successful participation in reading, writing, and discussion activities; and connecting knowledge, skills, and ideas within and across lessons and grades.

Seeking a conception of curriculum that would complement socio-cognitive analyses of effective instruction, Applebee (1996) undertook a study of how successful teachers developed a sense of continuity and cohesiveness across time in their classrooms. Again, the nature of the discussion that was orchestrated proved to be a central feature. In case studies that traced the evolution of 19 English classes (totaling 32 semesters of instruction) in Grades 7–12, Applebee found that the most effective curricula were organized around specific topics that unified the reading, writing, and discussion that took place over a semester or a year. Applebee called these long-term explorations “curricular conversations” to distinguish them from the short-term focus of day-to-day classroom interactions. Applebee, Burroughs, and Stevens (2000) found that, when an entire course was integrated around one or more central topics of conversation, students’ knowledge and understanding developed cumulatively throughout the course as they revisited important issues and concepts from new perspectives, with gradually broadening frames of reference. These conversations were governed by a set of conventions, or “ground rules” (Durst, 1999), that embodied important features of the larger domain of English as an academic discipline. Ground rules for curricular conversations about literature, for example, typically required the use of appropriate literary terminology, as well as textual analysis to support arguments and opinions.

The studies reviewed so far have focused on generalized literacy instruction, usually in the context of English language arts classes. A related body of research has looked more specifically at the effects on reading comprehension of a variety of formats for discussion of text that students find difficult. These formats are designed to involve students in a joint construction of meaning within which a variety of comprehension strategies are introduced, scaffolded, and practiced—where teachers, students, and classmates “think aloud” about a text (Kucan & Beck, 1997). Related approaches include Langer’s (1981, 1984) Prereading Plan for developing the related knowledge base; Transactional Strategy Instruction (Brown, Pressley, Van Meter, & Schuder, 1996; Pressley, Wood, Woloshyun, Martin, King, & Menke, 1992); Questioning the Author (Beck, McKeown, Sandora, & Worthy, 1996; Sandora,

Beck, & McKeown, 1999); and Collaborative Reasoning (Anderson, Wilkinson, & Mason, 1991; Chinn, Anderson, & Waggoner, 2001). This line of research has evolved over the years, from studies that emphasized tightly controlled interactions that were focused on individual strategies, which had relatively modest benefits for comprehension, to approaches that integrate multiple strategies in ways that achieve larger effects. Although each research team has developed its own vocabulary to describe the discussion routines in which the strategies are embedded, these lines of research overlap significantly in both the form and the focus of the particular interventions advocated. The results converge to suggest that comprehension of difficult text can be significantly enhanced by replacing traditional I-R-E patterns of instruction with discussion-based activities in which students are invited to make predictions, summarize, link texts with one another and with background knowledge, generate and answer text-related questions, clarify understanding, muster relevant evidence to support an interpretation, and interrelate reading, writing, and discussion. In general, the cognitive strategies that are emphasized in multiple-strategy approaches are much wider than those studied earlier in single-strategy research. These approaches also share an emphasis on group discussion and problem solving in which multiple perspectives are proffered and examined, requiring evidence-based argument in support of individual points of view.

In the study presented here, we extend these previous lines of work by examining the relationships among various aspects of classroom discussion and then relating them to improvements in a generalized measure of literacy performance in a diverse set of classrooms. The approaches that we examine—in particular, emphases on dialogic interaction, support for environment building, and extended curricular conversations—reflect a common set of emphases on discussion-based approaches to teaching and learning but are designed to capture different aspects of the complex processes involved.

Tracking and Academic Demands

Our inquiry is complicated by the fact that differences among classrooms in the nature and quality of instruction are not random but may be related systematically to the position of each class in a school's hierarchy of ability groups and curricular tracks. In American secondary schools, the sorting of students into tracked English classes based on their purported interests and abilities is widely practiced (National Center for Education Statistics, 1994; Brewer, Rees, & Argys, 1995). These classes typically are distinguished by labels such as "honors" (or "advanced"), "regular," and "remedial" (or "basic"). Sometimes high school classes are distinguished by their subject matter, such as "Popular Fiction" or "Modern European Literature," but they are nonetheless intended for students at different levels of literacy performance. Student performance, which is unequally distributed across tracks to begin with, becomes even less equal over time (for reviews see Oakes, Gamoran, & Page, 1992; Gamoran, 2000).

A major reason for this growing inequality is that instruction is unequally distributed by track level. Teachers may compete for the opportunity to teach honors classes, those who have the best reputations being assigned to the higher-level classes (Finley, 1984). In English, observers report that higher-track teachers place more emphasis on classic literature, whereas lower-track teachers focus on juvenile fiction (Oakes, 1985; Gamoran, 1993). One study characterized low-track classes as “caricatures” of regular classes, bearing the outward appearance of regular classes while paying superficial attention to academic work (Page, 1991). Gamoran et al. (1995) reported that honors English classes spent more time in discussion than did other classes, and a recent national study found that tracking is the major axis of differentiation in high school English instruction and performance (Gamoran & Carbonaro, 2002–2003; Carbonaro & Gamoran, 2002).

At first glance, it may seem that more rigorous academic demands are appropriate in higher tracks. After all, the students assigned to high tracks have higher achievement at the beginning of the school year. A closer look, however, shows that the differentiation of academic demands contributes to an expanding achievement gap. Moreover, when low-track students have an opportunity to respond to more rigorous academic work, they make more progress than when they encounter a diluted academic curriculum (Valli, 1990; Gamoran, 1993). Gamoran found that teachers of relatively successful low-track classes held high expectations, which they manifested by refusing to relinquish the academic curriculum. As one determined low-track English teacher declared, “I know it’s not easy, you guys, I know it’s not easy, but we’re not going to have a *Weekly Reader* in this class. . . . So stick with it” (p. 15).

Although tracking and instruction are linked, the one does not *determine* the other; consequently, it is possible to disentangle their effects empirically. Past studies have shown that tracking, instruction, and achievement are related in two ways: First, the quality and rigor of instruction tends to be higher in high tracks, so that instruction accounts for (or mediates) the influence of tracking on achievement (e.g., Carbonaro & Gamoran, 2002). Second, the *impact* of instruction may differ across tracks; for example, if discussion in high-track classes focuses on literature and discussion in low-track classes focuses on students experiences, then discussion may have a stronger impact on literature achievement in high-track classes (Gamoran et al., 1995). Although tracking and instruction are linked, they exhibit enough independent variation to estimate their separate effects.

Because students in different tracks differ from one another in many ways, it is essential to take account of student characteristics in trying to estimate the effects of tracking and achievement. Most important, one must have an indicator of performance differences at the time that students are assigned to their classes. In addition, it is important to take account of student background characteristics such as gender, race/ethnicity, and social class, which are correlated with track assignment and with academic performance. Prior studies show that a rich array of prior achievement and background indicators is very helpful in distinguishing the effects of tracking from the effects of the correlates of track assignment (Gamoran & Mare, 1989).

The Nature of Literacy Performance

The nature and measurement of literacy performance is a contested area within the fields of English and literacy education research. The definitions that drive individual programs of research range from those that focus on the ability to decode or comprehend particular texts to those that include the ability to participate effectively in the specialized “discourses” (Gee, 1996) of the academic disciplines. Differences in definitions lead to differences in the extent to which reading and writing are conceptualized (and, by implication, taught and tested) as separate or interrelated, drawing on an underlying set of cognitive and linguistic knowledge and skills that are deployed in any literacy event. For the present study, the sociocognitive frame with which we began led us to focus on what Bereiter and Scardamalia (1987) call “high literacy”: the reading, writing, and discussion skills that allow students to participate effectively in the “disciplinary conversations” (Applebee, 1996) or “secondary discourses” (Gee, 1996) of English as a school subject—and, by extension, to do well in a wider array of tasks necessary for success in other school subjects, in life, and in work. Typically, such conversations involve a mixture of common readings or other experiences, accompanied by writing and discussion around common topics or themes, with successful participation reflected in the ability to talk and write effectively about what has been read or experienced, mustering arguments and appropriate evidence to support an individual point of view. At the same time, we acknowledge that acts of reading, writing, and speaking may lead to somewhat different patterns of deployment of underlying skills, as Langer (1986) has argued in an investigation of reading and writing differences within a sociocognitive frame.

This focus on contextually situated, higher literacy skills leads us to examine literacy through relatively complex tasks that reflect typical performances for students who are “doing English” in American middle and high schools. In this we are following the practice of the National Assessment of Educational Progress (NAEP; see Donahue, Voelkl, Campbell, & Mazzeo, 1999), whose reading assessments regularly ask students to write short-answer and extended responses to selections that they have read. Previous studies of English instruction in middle and high school make clear that the most typical assignment is a text-based essay, focusing on student responses and interpretations; this pattern is even more pronounced in high school than in middle school (Applebee, 1993, pp. 163, 167). Thus, for the present study, we focused on literacy performance as reflected in tasks that required students to write about fictional characters or their own experiences—relatively typical school tasks—although the students were asked to write from memory rather than about specific, available texts.

Just as literacy is a contested area, appropriate methods for modeling of gains in performance are also debated, although regression approaches generally are preferred over the use of simple gain scores (Willett, 1988–1989). The issue is further complicated in measures based on writing performance because absolute performance on such tasks is highly sensitive to seemingly minor shifts in topic. Therefore, in the present study we looked at end-of-year

performance on set writing tasks, controlling for the effects of beginning-of-year performance on a related task and on other measures expected to be related to prior performance.

The Study

For the present study, we drew on our previous research, itself situated within the larger body of research on literacy instruction, to derive a variety of variables related to discussion-based approaches and, in turn, to examine the relationships among those variables and between them and students' literacy performance. In particular, we examined (a) the interrelationships among variables reflecting dialogic approaches to instruction, an emphasis on envisionment building, extended curricular conversations, and high academic demands; (b) the relationships between the variables and spring literacy performance (controlling for initial performance and related background variables); and (c) the interactions between the variables and grade level, school context, level of performance, and race/ethnicity.

Sample

For our study we sought a sample of middle and high school classrooms with diverse approaches to literacy instruction and diverse student bodies. Sites were located in 5 states selected for variation in their approaches to high-stakes assessment: California, Florida, New York, Texas, and Wisconsin. Within each state, 1 city and 1 suburban district agreed to participate, and within each district, 1 middle school and 1 high school were selected, for a total of 20 schools: 5 urban high schools, 5 urban middle schools, 5 suburban high schools, and 5 suburban middle schools. Before the first round of data collection was completed, the Texas urban middle school withdrew from the study, reducing the total to 19 schools. Within each school, we sought the participation of 4 classes, selected in cooperation with the department chair or language arts specialist to represent the array of tracks in the school; generally, this included 1 honors class, 1 remedial class, and 2 regular classes. (Half of the middle schools did not use tracking; in those cases we asked simply for 4 classes reflecting the diversity of the grade level.) Stratifying our sample by track insured that it reflected the diversity of performance within the student population in each school. Selecting classrooms with input from school leaders probably biased our sample toward experienced teachers without serious control problems; but it also insured that, in studying instructional effects, we chose classrooms where instruction was actually taking place. (Teachers and school leaders were unaware of the particular features of instruction on which we focused.) Participating teachers were offered a modest stipend for the out-of-school time required by the study.

We studied Grades 7 or 8 in the middle schools and Grades 10 or 11 in the high schools (one 12th-grade class was also included). In one school, only 2 classes agreed to participate, and two other schools contributed only

3 classes each. Three additional classes were dropped from the study because of low rates of student participation, and in 5 classes, student participation rates in the fall data collection were too low to include in the current analysis. Thus our sample for analysis consists of 64 classes in 19 schools, with participation rates of 95% and 84% at the school level and the class level, respectively.

A total of 1,412 students attended the 64 classes for the entire school year, and 1,111 agreed to participate in our study—a response rate of 79%. (Most of the nonrespondents in these classes, as well as in the classes that we dropped from analysis, simply failed to turn in consent forms; very few actively refused.) Missing data for individual students on fall or spring tests reduced the sample for analysis to 974 students, or 88% of the study participants and 69% of all students who were in the class for the whole school year.

Procedures

Nominations of schools in the targeted states were collected in the fall of the year preceding the study; initial screening was based on research team members' knowledge of the schools and on information about the populations served that was available on the Web. After selecting an initial set of districts, we contacted school or district administrators, first by telephone and later through school visits, to solicit cooperation and determine suitability for participation in the study. Team members visited all sites to explain the study to relevant administrators, to discuss with potential collaborating teachers the commitment required (for classroom observations, testing of student performance, and completion of questionnaires), to solicit the department chairs' help in selecting appropriate teachers and classrooms, and to be sure that there was firm support for the study from teachers, department leaders, the principal, and district administrators.

Data were gathered during the following year by a team of 5 field researchers. An initial visit was made to each classroom to explain the study to students and to distribute consent forms; active consent for participation was required from each student and from a parent or guardian. An assessment of initial literacy performance was administered during a class period as early as possible in the school year, usually in October. We made follow-up assessments and distributed a student questionnaire in two class periods during the spring of the year, usually in May or June, depending on the school calendar. Teachers were asked to complete a questionnaire at the same time, focusing on the target classrooms.

Four lessons were observed in each classroom, two in the fall and two in the spring. Lessons were selected in collaboration with the teacher; teachers were asked to select a class that included discussion of a work of literature, in whatever form such discussions usually took in the particular class. This provided some necessary cross-class consistency because overall patterns of organization and interaction are likely to be quite different for different parts of the curriculum, even within the same classroom (e.g., lessons are

organized differently for literature discussions, writing workshops, and play productions). During lessons, the field researcher used the CLASS computer program (Nystrand, 1999) to record class activities and interactions in real time and, at the end of the class, to complete a variety of ratings about other aspects of curriculum and instruction. Each class was audiotaped to allow the field researcher to edit the CLASS data. The audiotapes were also used by a separate team of data editors, who reviewed all CLASS datasets for consistency across raters. Because of the distribution of sites across five states, fall and spring observations usually were scheduled in conjunction with week-long visits that also included administration of student assessments.

Measures

Measures were derived from the following sources:

- A teacher questionnaire asking about educational background and experience, classroom composition, and instructional emphases related to dialogic instruction, envisionment building, and extended curricular conversations;
- A student questionnaire asking about home background, school achievement, and the amount of work required for various classes;
- CLASS 3.0 (Nystrand, 1999), an expanded version of Nystrand's (1997) program for analyzing classroom discussion and related activities, focusing especially on the types of questions that teachers and students asked, materials that they used, and their interactions with each other (the program also prompted observers for overall ratings of a variety of features related to the degree of emphasis on envisionment building and extended curricular conversations); and
- Measures of student literacy performance.

Each class was observed four times, twice in the fall and twice in the spring, and the data were averaged across observations. Teachers and students completed questionnaires in the spring; student performance was assessed in the fall and spring.

To examine the relationships of interest to the present study, a variety of variables were derived to reflect the study's design, student background characteristics, dialogic instruction, emphasis on envisionment building, extended curricular conversations, high academic demands, and student performance over the course of the academic year.

Design Variables

Dummy variables coded 0 and 1 were used for grade level of the school (middle school = 0, high school = 1) and for urbanicity (suburban = 0, urban = 1). The extent and nature of tracking was captured in two variables: Untracked school (coded 0, 1) at the school level; and low (remedial), middle, or high (honors) track at the classroom level (coded -1, 0, 1). The coding of tracking

assumes that track effects are linear within the sample, a reasonable assumption given the results of previous analyses using this dataset (Gamoran & Kelly, in press).

Background Variables

A variety of background variables were taken from student questionnaires. Gender and race/ethnicity were coded using dummy variables, on the basis of student self-report. Just over half (53.5%) of the respondents were girls; 32.1% were Hispanic American, 17.6% African American, and 6.3% Asian American. To provide a proxy for socioeconomic status, students were asked about the availability of home resources such as a daily newspaper, a dishwasher, two or more cars, more than 50 books, a VCR, a computer, and so on. This scale has been used in similar form in several national surveys, including the National Educational Longitudinal Survey, which began with eighth graders in 1988 (Carroll, 2000). The sum of items reported was then converted to z scores for analysis. (Raw scores ranged from 1 to 11 [the scale maximum], with a mean of 8.7 items.) For the sample as a whole, schools reported that 31.4% of the students received free or reduced-price lunches, ranging from 2% to 79% across schools. Teachers' reports at the classroom level were similar; they estimated that an average of 29.6% of the students came from low-income families, with a range from 0 to 90%.

An additional measure of overall academic ability was calculated as each student's grade point average for all subjects other than English. This was calculated on a scale from 0 (failing) to 4 (A), from student reports of average grades in each subject that the students were taking at the time. The mean on this scale was 3.0, a B average, with scores ranging from 0 (failing all subjects) to 4 (straight A's).

Evidence of Dialogic Instruction

A variety of measures related to dialogic instruction were taken from the CLASS observation system (Nystrand, 1999), a program implemented on laptop computers for collecting and analyzing unfolding classroom discourse during observations. Using this system with audiotape backup, the observers listed all questions posed by teachers or students during instructional episodes (defined as any coherent classroom activity centering around a particular purpose or topic) occurring within the class observations. In our research, a new episode starts when the teacher addresses a new purpose or topic. Sometimes episodes consist of two or more activities. For example, in addressing a particular objective, the teacher may initiate a question-and-answer session, which is then interrupted by periodic, brief lectures and culminates with a homework assignment. We treated such episode parts as segments, defined as any coherent part of an episode that differs from other activities constituting the episode.

We counted as questions all queries for information, including mainly international questions and some tag questions, but we did not count (a) procedural questions (e.g., "How many pages do we need to read?"); (b) rhetorical

questions; or (c) discourse-management questions or repair initiations (e.g., “What?” “Excuse me?” “Did we talk about that?” “Where are we [in the text?]”), or questions that initiated discourse topics, such as, “Do you remember our discussion from yesterday?”

All CLASS observations collected in the field were rechecked by the observer after the class and then recoded for consistency across observers by a separate team of data editors, who based their work on audiotapes of the observed lessons and discussions with the original observers.

We chose the variables described in the following paragraphs as evidence of dialogic instruction:

Open discussion. Open discussion was defined as free exchange of information among students and/or between at least three participants that lasts longer than 30 seconds. The three participants may include the teacher, although the teacher may be deliberately silent during some discussions. To standardize across classrooms, minutes of open discussion were calculated per 60 minutes of class time.

When conditions are right, especially following student uptake of authentic questions and other “dialogic bids” (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003) offered by the teacher, the result is an open discussion in which teachers and their students work out understandings face-to-face—the quintessential form of dialogic interaction. When this happens, the teacher’s role is mainly one of starting and keeping the ball rolling. Discussion tends to be marked by the absence of questions, from both teacher and students, except for purposes of clarification. When open discussion occurs in the midst of question-and-answer, it interrupts the normal I-R-E sequence of recitation. A return to test questions typically marks the end of discussion.

Authentic teacher questions. We defined teacher questions as authentic if they did not have a prespecified answer that the teacher was seeking. This variable was calculated as a proportion of all teacher questions.

Questions with uptake. We defined questions with uptake as questions that incorporated what a previous speaker had said. The variable was calculated as a proportion of all questions by teacher and students.

Evidence of Envisionment Building

A variety of measures related to envisionment building were taken from Langer’s (1995, 2001, 2002) studies of instructional strategies in effective English language arts classrooms. At the end of each classroom observation, observers completed two sets of ratings that related to the degree of classroom emphasis on envisionment building. One set focused on the extent to which particular activities were observed during the lesson; the other asked about the observer’s overall perception of the ground rules that the teacher had established for classroom interaction. These were relatively high-inference ratings that focused on the teacher’s intentions and expectations as reflected in overall patterns of activity and discussion. Completion of both sets of ratings was prompted by the CLASS program at the end of each observation.

Discussion-Based Approaches to Developing Understanding

Emphasis on envisionment-building activities. This type of activity was calculated as the mean of 11 items reflecting observed activities that promote envisionment building. Each item was rated from 0 (never/none) to 3 (almost always/almost all students); labels varied slightly depending on the focus of the question. Items were standardized (z scores) before the computation of total scores. Alpha for the total scale was .93. The 11 items were as follows:

- Students were allowed room to develop their own understandings in reading and writing activities;
- Students spent class time in purposeful conversation with peers and teachers;
- Students were encouraged verbally or through modeling to take a position, express opinions, or explore personal reactions;
- Students asked questions that showed comprehension;
- Students asked questions that showed evaluation or analysis;
- Students were allowed to shift discussions in a new direction;
- The teacher encouraged students to use others' questions and comments to build discussion;
- Students actually did so;
- Students responded to other students or to the teacher with challenges, comments, opinions;
- Students challenged the text (e.g., by bringing in alternative points of view); and
- The teachers' questions required analysis.

Ratings of envisionment-building activities. At the end of the observation, each observer rated the observed lesson on whether the teacher

- Treated all students as having important contributions and understandings;
- Treated instructional activities as a time to develop understandings rather than to test what students already knew;
- Assumed that questions were a natural part of the process of understanding rather than an indication of failure to learn; and
- Used multiple perspectives (among students, texts, and other voices) to enrich understanding.

Each item was ranked on a 4-point scale (1 = classrooms with no evidence of support for envisionment building, 4 = classrooms fully oriented toward supporting the development of students' evolving envisionments). The ratings were standardized (z scores) and then averaged to yield a scale with an alpha of .82.

Evidence of Extended Curricular Conversations

Two variables captured aspects of the extended curricular conversations taking place in the participating classrooms, one derived from observers' reports and the other from the teacher questionnaire.

Observed curricular conversations. We calculated this variable as the mean of observer ratings of four items reflecting the extent to which discussion was part of a continuing topic (across days): (a) Comments explicitly made connections with prior topics; (b) comments explicitly made connections with future topics; (c) students used the text to support or refute positions; and (d) students used the technical vocabulary of literary analysis. Each item was rated from 0 (never/none) to 3 (almost always/almost all students). Items were standardized (z scores) before computing the total score. Alpha for the four-item scale was .82.

Connections among class activities. This variable was calculated as the mean of teacher ratings of seven items on the frequency of activities involving interrelationships among reading, writing, and discussion ($\alpha = .72$). Frequency was recorded on scales that ranged from 1 (never) to 7 (every day). Teachers answered the questions in this list:

About how often

- do students in your class write about (or in response to) things they have read?
- do you discuss writing topics with your students before asking them to write?
- do you and your class discuss the readings that you assign?
- do you ask students to relate what they have read to their other readings?
- does your class relate its discussion to previous discussions that you have had?
- do you and your class discuss things that students have written about?
- do you ask students to explain their answers, ideas, or comments?

Evidence of High Academic Demands

We derived several measures that reflected the amount of work that was expected from the students. In each case, student-level reports of work required or completed were aggregated to the classroom level to provide an estimate of the overall level of work required by the teacher (as opposed to the relative effort expended by individual students within each class).

Emphasis on revision. Students made individual reports of the kinds of revisions that they made in their writing. These were averaged to form two scales, one for content (ideas and information, organization, and development; $\alpha = .57$) and one for mechanics (spelling, punctuation, grammar, and usage; $\alpha = .80$). Each scale ranged from 0 (reported no revisions) to 1 (reported doing each of the types of revision). For the present analyses, classroom-level means were then calculated from these student-level scales to indicate emphasis on revision of content and on revision of mechanics.

Hours of English homework per week. Individual student reports of weekly hours of English homework (from 0 to 4 or more) were aggregated to the classroom level for the present analyses.

Discussion-Based Approaches to Developing Understanding

Students complete reading and writing assignments. All students were asked how often they completed their reading and writing assignments. Responses were recorded on scales ranging from 0 (never) to 6 (every time). The separate scales for reading and writing were averaged ($\alpha = .72$) at the student level and then aggregated to the classroom level for the present analyses.

Materials Used and Writing Assigned

To provide additional information, observers also completed a series of checklists to record the presence or absence of a variety of types of materials and writing activities in classroom lessons or homework. For materials, observers recorded traditional selections, nonfiction (essays, biography, autobiography), drama, poetry, and young adult literature. For writing activities, they recorded short-answer exercises, note taking, and a variety of types of writing ranged along the scale for levels of abstraction (see Appendix). Materials and writing activities were averaged across the four lessons to yield a series of measures of the percentage of lessons in which they were observed.

Student Performance

Three tasks were used to assess performance:

Task 1 (Fall):

Think about the fictional characters in a book you have read or a movie you have seen recently. Of all the characters in this piece, think of the one you admire and respect the most. First, using specific details, briefly describe him or her. Then fully explain: Why do you admire this character? Use this paper to do your brainstorming (figure out what you want to say). Then write an essay starting on the next page.

Task 2 (Spring):

Think about a fictional character from a novel, short story, or play that you've studied in this class since fall. Of all the characters in the works you read for this class, think of the one you admire and respect the most. First, using specific details, briefly describe him or her. Then fully explain: Why do you admire this character? Use this paper to do your brainstorming (figure out what you want to say). Then write an essay starting on the next page.

Task 3 (Spring):

Think about an experience you have had that taught you something important. Tell what happened and what you learned, and explain why it is important. Use this paper to do your brainstorming (figure out what you want to say). Then write an essay starting on the next page.

To evaluate student performance on these tasks, we used a scoring system that gives credit both for the overall level of difficulty of the task

attempted and for the student's success in carrying it out. For level of difficulty, we evaluated the level of abstraction of the task attempted on a scale of 0–5; for success in carrying out the task, we scored the level of elaboration in the students response on a scale of 1–4 (see Appendix).

Level of abstraction. This scale was developed in previous studies by Britton, Burgess, Martin, McLeod, and Rosen (1975); Applebee (1981); and Nystrand (1997; Nystrand, Cohen, & Dowling, 1993). At the low end of the scale are *records*, such as lab notebook entries, which keep track of events as they unfold; at the high end are *theoretical papers*, such as extended essays involving logical argumentation entertaining propositions about propositions and about relationships among propositions. Between records and theoretical papers are *reports* and *analyses* making generalizations about observations and experience. This scale is treated here as an initial rating of the difficulty of the task that the student has undertaken in the writing: Although all students were given the same tasks, they chose to address them in more or less abstract ways that simplified or complicated the task undertaken (cf. Durst, 1987).

Level of elaboration. Given a task of a particular level of difficulty, students exhibited a wide range of levels of performance. The range was captured in the second scale, drawn from an informational writing scale used by the NAEP (Applebee, Langer, Mullis, & Jenkins, 1990), which is treated here as a rating of level of elaboration. This scale ranged from *unsatisfactory* (the writer provides only the barest information) to *elaborated* (the text is highly wrought, well developed, and tightly organized).

Results of assessment of student performance. Each response was rated on each scale by two independent raters. Discrepancies larger than 1 were adjudicated by a third rater. The adjudicated scores were then averaged.

On the abstraction scale, mean scores by grade level and task ranged from 1.80 to 2.23 (Table 1), with a mode of 2.0 on all three tasks at both grade levels. This result indicates that the largest number of students responded to each task by writing a report, in which they provided an account of the particular admired character or series of events and their reactions to them. This

Table 1
M and SE for Performance Measures, by Task and Level

Measure and level	Task 1 (fall)	Task 2 (spring)	Task 3 (spring)
Level of abstraction			
Middle school	1.80 (.03)	1.95 (.03)	1.90 (.03)
High school	2.00 (.04)	2.23 (.04)	2.17 (.03)
Level of elaboration			
Middle school	1.62 (.02)	1.70 (.02)	1.75 (.02)
High school	1.88 (.02)	2.03 (.03)	2.00 (.02)

Note. Middle school $n = 470$, high school $n = 504$. Figures in parentheses represent *SE*.

was an appropriate, although simple, level of response to the three tasks, one that avoided, for the most part, the request for analysis. Students varied in the level of abstraction of the tasks that they attempted, however. In the fall, 16.8% of the responses across tasks and grades were rated as attempting a less abstract task, and 25.7% were rated as attempting a more abstract one; the comparable figures for the spring were an average of 10.8% and 31.3%, respectively.

For the elaboration score, means across levels and tasks ranged from 1.62 to 2.03; again, the most typical rating on all three tasks at both levels was 2. On this scale the rating of 2 reflects a minimally elaborated response without a strong organizing frame. Again, there was a range in levels of performance. In the fall, 44% of students fell below the mode and 14% were rated above it; in the spring, an average of 36% scored below the mode and 20% scored above it. These results are comparable to the results that NAEP found in national samples of middle and high school students whose performance was assessed by means of the same scale. On "Food on the Frontier," an informative task that similarly invited analysis at Grades 8 and 11, the modal response at both grades was 2.0, with only small percentages of students scoring at 3 or 4 on most of the tasks (Applebee et al., 1990, pp. 95–96). Mean scores for the NAEP task rose from 1.91 ($SE = .03$) at Grade 8 to 2.03 ($SE = .02$) at Grade 11, less than half of the mean differences between middle and high school levels on the elaboration score for the tasks in our sample (see Table 1). All of this suggests that the growth in literacy skills for the students in our sample, as reflected in differences between middle and high school, was at least as great as that for students in the nationally representative samples drawn by NAEP (as scored on the same scale).

Correlations between the abstraction and elaboration scales across the three assessments (one in fall and two in spring) ranged from .63 to .70. Each student's total score was calculated as level of abstraction of the task undertaken (0–5) plus level of elaboration in completing the task (1–4). The result was a total score scale of 1–9 for each of the three tasks, with inter-rater reliabilities of .77, .82, and .78 for Tasks 1 to 3, respectively, similar to reliabilities in other studies that have used this composite measure of overall performance (Nystrand, 1997).

Scores from the two spring measures were averaged to provide an end-of-year score ranging from 1 to 9, while the fall measure along with other variables was used to provide a control for initial differences in literacy performance. Scores on the fall measure averaged 3.66; those on the spring assessments averaged 3.94.

It is important to note that the three sets of ratings involved different writing tasks, benchmarked independently. Although there may have been some practice effect because of the similarity among the three tasks, such an effect is constant across classes and does not affect our estimates of relative performance in the spring. In fact, with the methodology used in the present study (with different tasks and benchmarking), the overall means and the difference in means between fall and spring do not provide estimates of

absolute performance, even in the absence of practice effects. They do, however, allow an examination of the relationship of spring performance to the classroom variables of interest, controlling for the effects of a variety of other variables, including initial literacy performance as measured by the fall task.

Tasks 1 and 2 measured students' performance in writing about fictional characters; Task 3 was a more general measure of writing performance. Analyses from earlier studies (Nystrand, 1997; Nystrand & Gamoran, 1991) show that performance on Task 2 was significantly related to other measures of literacy performance in which students answered a series of questions about five works of literature (stories, novels, dramas, and short plays) that they had read during the year. The correlations between the writing prompt and the other measures were as follows: Extent of recall, $r = .488$; depth of understanding, $r = .526$; ability to relate story endings to denouement, $r = .367$; ability to relate story conflict and/or ending to theme, $r = .441$; understanding the internal motivations of characters, $r = .509$; interpretive treatment of the major selection, $r = .422$; and the level of discourse used to discuss theme and conflict, $r = .401$. As an additional validity check, we obtained standardized reading and language test scores for a subsample of 107 students who took the Stanford 9 or the Wisconsin Survey of Academic Skills. The correlation between the average spring writing score (combining the two spring writing prompts, Tasks 2 and 3) and the standardized reading scores was $.547$; the correlation of average spring writing score with standardized language arts scores was $.487$.

Analyses

As a first step, we used principal components analyses with Varimax rotations to reduce related sets of data to a smaller number of scales based on internal consistency and interpretability. In the present report, these derived scales include the following measures (as defined earlier in the "Measures" section): envisionment-building activities, extended curricular conversations, and connections among class activities.

To assess the interrelationships among the remaining variables, after the main variables had been derived, another principal components analysis was carried out on the subset of independent measures representing the key sets of variables reflecting dialogic instruction, envisionment building, extended curricular conversations, and high academic demands. A two-factor solution accounted for 55.1% of the original variance. Two standardized factor scores were generated from this solution and used to examine the relationship of performance to instructional emphases.

Because the data in this study vary significantly at the student, classroom, and school levels, we used multilevel models of 974 students nested within 64 classes in 19 schools to decompose variance in spring literacy performance into student, class, and school levels, using the statistical package HLM 5 (Raudenbush, Bryk, Cheong, & Congdon, 2000). The hierarchical linear models analyzed by HLM 5 provide maximum likelihood estimates of

regression effects estimated simultaneously at different levels, accounting for the nested nature of the data and providing standard errors, degrees of freedom, and significance tests appropriate to the variables included at each level within the analysis. Because our primary focus is on classroom effects, a variety of variables were entered as controls at the school level (grade level, urbanicity) and the student level (fall performance, grade point average, socioeconomic status, race/ethnicity, and gender). With these controls in place, we then examined the classroom variables of interest, including instruction and track placement. We examined interactions to assess differences in the effectiveness of particular approaches for different groups of students.

Results

Instructional Emphases

Table 2 provides means and standard deviations on the key variables in this study, as well as their breakdown by school grade level and track. Minutes of open discussion—defined in the present study as more than 30 seconds of free exchange of ideas among students or between at least three participants—reflects class discussion that sustains itself in an uninterrupted exploration of ideas and understanding. Open discussion usually begins in response to an open-ended question about which students can legitimately disagree, and it ends with the introduction of a new activity or a return to an I-R-E pattern of interaction. During open discussion, all participants are partners in the development of understanding. Sustaining such exchanges is difficult and may require considerable scaffolding by the teacher and considerable previous experience in discussion by the students. In the present study, open discussion averaged 1.7 minutes per 60 minutes of class time. Although the figure seems low, this is in part because 30 seconds is a considerable amount of time in the ordinary pace of classroom discourse. (This measure is related to Boyd and Rubin's [2002] measure of "student critical turns," defined as 10 seconds of uninterrupted talk with coherence and substantive engagement; Boyd and Rubin found that even when the 10-second criterion was applied, student critical turns required considerable scaffolding and support from the teacher.) The amount of open discussion in the present study is twice as high as Nystrand (1997) found in his previous investigation of 8th-grade literature classes (p. 43), and 7 times higher than he found in 9th-grade classes (p. 45). (Nystrand labeled this variable "discussion," but we have modified the label to distinguish these exchanges from other kinds addressed in our more general emphasis on discussion-based approaches.) In addition to minutes of open discussion, approximately 19% of the teachers' questions were rated as "authentic," that is, as not seeking a prespecified answer. And 31% of all questions asked involved uptake, building on a previous comment rather than moving through a series of unrelated issues one at a time.

These interaction patterns are fairly consistent across levels and tracks. The major difference is that low-track students engaged in considerably less open discussion than did high-track students (an average of 42 seconds, as opposed

Table 2
M and SD for Classroom and Student Variables, by School Level and Track

Variable	Grade level of school		Track level				Total	(SD)
	Middle school	High school	Low	Middle	High	Mixed		
Classroom variables	(<i>n</i> = 31)	(<i>n</i> = 33)	(<i>n</i> = 10)	(<i>n</i> = 26)	(<i>n</i> = 13)	(<i>n</i> = 15)	(<i>n</i> = 64)	
<i>Evidence of dialogic instruction</i>								
Authentic teacher questions (mean proportion)	.19	.19	.22	.19	.21	.14	.19	(.13)
Open discussion (mean, minutes out of 60)	1.65	1.74	.70	1.44	3.30*	1.42	1.70	(3.02)
Questions with uptake (mean proportion)	.31	.31	.31	.31	.29	.31	.31	(.12)
<i>Evidence of envisionment building</i>								
Emphasis on envisionment-building activities (mean <i>z</i> score)	-.10	.09	-.52	-.06	.80*	-.24	0.00	(1.00)
Ratings of envisionment-building activities (mean <i>z</i> score)	-.12	.11	-.15	.13	.39	-.46	0.00	(1.00)

Evidence of extended curricular conversations

Observed curricular conversations (mean z score)	- .36	.34*	-.40	-.11	.66*	-.12	0.00	(1.00)
Coherence within lessons (teacher report) (mean, ratings of 1 to 7)	4.97	5.54*	5.04	5.04	5.77*	5.35	5.26	(.86)
<i>Evidence of high academic demands</i>								
Emphasis on revision of mechanics (student aggregate) (mean, 0 to 1)	.57	.58	.50	.58	.70*	.53	.58	(.16)
Emphasis on revision of content (student aggregate) (mean, 0 to 1)	.54	.62	.53	.60	.73*	.47	.58	(.19)
Hours of English homework per week (0 to 4)	.99	1.37	.88	.98	2.01*	1.01	1.18	(.67)
Students complete reading and writing assignments (mean, 0 to 6)	4.99	5.00	4.79	4.99	5.19*	4.98	5.00	(.45)
Student variables								
Fall performance (1 to 9)	3.42	3.89*	3.12	3.55	4.39*	3.32	3.66	(1.19)
Spring performance (mean, 1 to 9)	3.65	4.21*	3.48	3.89	4.63*	3.51	3.94	(1.04)
Grade point average excluding English (0 to 4)	3.11	2.94*	2.67	2.95	3.21*	3.09	3.02	(.78)
Home resources (z scores)	.04	-.04	-.21	-.02	.22*	-.11	.00	(1.00)

* Significant difference between middle and high school or between high and low tracks, $p < .05$.

to 3.3 minutes, per class). The amount of open discussion did not vary significantly between middle school and high school.

Other measures in Table 2 show some differences between middle and high school instruction. Both measures of engagement in extended curricular conversations are significantly higher for high school than for middle school. The middle school students also had slightly higher grade point averages and, as expected, lower scores on the performance measures in spring and fall. They also showed a trend toward fewer hours of homework per week than the high school students, $p < .10$, two-tailed.

There is considerable variation across tracks on these variables, suggesting that there are significant differences in what teachers expect of higher- and lower-track students. Lower-track students are taught with significantly less emphasis on envisionment-building activities; extended curricular conversations; connections among reading, writing, and discussion activities; revising activities; and homework. And they are in classrooms where the class as a whole is less likely to complete assigned reading and writing tasks. Considering, in addition, lack of open discussion time, we see that lower-track classes receive significantly less instruction of the kinds that previous studies suggest contribute to higher literacy performance.

The other result to note in Table 2 is the significant relationship between track and home resources (our proxy for socioeconomic status). Students with more home resources were concentrated in the high-track classrooms, and those with fewer home resources were concentrated in the low-track classrooms—results consistent with those of many other studies (e.g., Gamoran et al., 1995).

Taken together, our measures of instruction reflect broad and important differences in approaches to teaching and learning in the classrooms that we studied. As examples, consider two brief excerpts of high school literature instruction. In the first session, a high school class on the *Iliad*, the teacher reviews the main points of plot, setting, and narrative. Notice the tightly scripted nature of his questions, each of which leaves room for only a few words of response from the students:

Teacher: According to the poet, what is the subject of the *Iliad*?

Mary: Achilles' anger.

Teacher: [Looking for another answer] Where does the action of the first part of Book I take place when we enter the story?

Joshua: On the Achaean ship?

Teacher: Well, they're not on their ships. Let's see if we can give you a little diagram. . . .

Corrine: Was it on the shore?

Teacher: Yes, it's on the shore.

The teacher then proceeds to sketch a map of Troy on the board, pointing out key sites as he does so.

This exchange is followed by more review questions about plot and narrative. All told, 10 of 13 questions asked in the instructional episode from

which the preceding dialogue is taken are teacher test questions with no uptake. Only one teacher question exhibits uptake, and only one question (“Didn’t they put a wall up in Ireland?”) is authentic, asked by a student trying to make sense of the wall the teacher describes in connection with Troy; but the question is treated as a distraction by the teacher. (“In Ireland?” the teacher replies. “I’m not familiar with that.”) Moving the class back on track, he continues, “So, let’s take a look at some of the other questions. What’s the story behind the quarrel. . . ?” Students’ envisionments of the text are treated as right or wrong, complete or incomplete, rather than as dynamic constructions reflecting the students’ evolving understanding of this complex and potentially interesting work of literature.

Although the term “recitation” usually refers to *students’* oral presentations of previously learned material, the preceding excerpt demonstrates how completely the teacher can do the actual reciting. The students play a minor and supporting role in what is said, mainly by responding with an occasional word or two to the teacher’s periodic questions. We get the impression that the teacher is working from a list of topics and questions, covering predetermined points in a predetermined order; that he has done so in the past and will continue doing so in the future; and that the makeup of each class affects the script very little. Students are procedurally, not substantively, engaged in this lesson.

In contrast, we now examine a more openly discursive sequence from a middle school classroom that rated high on our measures of instruction. The students have just reviewed their homework on *To Kill a Mockingbird*, and after some initial questions about details of the story the teacher opens the floor to student reactions and questions. His initial question points students to the complexities at the end of the story, when Bob Ewell attacks the children, Jem and Scout, but ends up dead. Reviewing what happened, the sheriff decides that Ewell fell on his own knife, although Heck Tate knows that in fact it was the reclusive Boo Radley who saved the children. The teacher’s questions push students to articulate and expand on their initial reactions and to respond to the moral ambiguities that the incident generates:

Teacher: How does Bob Ewell get killed?

Student: Boo Radley [did it].

Teacher: How did you figure out that Boo killed him?

Student: . . . But I guess that I thought that the knife. . . . I really didn’t understand this [part]. I thought it was Boo at the beginning, but then I was not sure.

The teacher steps in with some additional scaffolding to help the students in developing their understanding of what actually happened in the story. The students then focus on the issue of the cover-up:

Student: It said that he doesn’t want to reveal it to the [sheriff] because . . . it would ruin, you know, Boo’s life.

Student: Right.

- Student:* Even if he totally [did it].
- Student:* He'd get all this attention and he couldn't, obviously . . .
- Student:* No, he wouldn't be able to [continue to live as before] if they all found out that he did it.
- Teacher:* Why not?
- Student:* Well, he's going to have to go to trial, and, uhh, . . . all this stuff, and everyone will know about that he has. . . .
- Student:* I think it's worth it. . . .
- Teacher:* So you think that Heck Tate was wrong in covering up?
- Student:* Yeah! Well, Heck Tate said that anyway, it's gonna be self-defense anyway, however it comes up on the trial, because you can really argue it that way. So you just have to go through the whole trial and then it would be up to the jury and stuff just to get to some answer that you already know about.

At this point, the teacher's role is mainly one of directing conversational traffic, focusing issues, and guiding students through the text to answer their own questions. The students respond to one another rather than just to the teacher, and their turns grow in length and detail as they become involved in the issues. Discourse quickly evolves into a discussion lasting more than 15 minutes. The teacher wraps things up by instructing students on their next tasks in small groups, where, as in the preceding whole-class discussion, they grapple with big issues, figuring out an interpretation of the novel that can form the basis for an essay assignment. Of the 23 teacher questions in the instructional episode as a whole, 26% are authentic, either posing questions on which there may be disagreement or tapping into students' reactions. And 57% of the questions involve uptake, reinforcing the flow of discussion and insuring that ideas build on one another. The focus of the exchange, starting with homework reading and extending through discussion and essay writing, is on the development of students' envisionments of the text. The teacher's prompts often lead students to find evidence to support their initial reactions.

Writing Activities and Reading Materials

The main variables analyzed in this study were related to processes of discussion and interaction and were intended to capture the degree of emphasis on the development and elaboration of ideas in each classroom, as reflected in classroom observations and reports from students and teachers. It would also be valuable to examine the quality of the content that formed the basis of discussion in the classes studied, but our focus on consistencies in the patterns of interactions across lessons that were sampled from different parts of the year made it difficult to derive useful measures of content quality. Observers did, however, record the kinds of materials that were used and the kinds of writing activities that were undertaken in association with the observed lessons, without any assumption that one kind was likely to be more valuable than another. These data are summarized in Table 3 to provide a fuller picture of the classroom activities observed.

Table 3
Materials Used and Writing Activities Observed (Mean % of Observations)

Type of material or activity	Grade level of school		Track level				Total (<i>n</i> = 64)	(<i>SD</i>)
	Middle school (<i>n</i> = 31)	High school (<i>n</i> = 33)	Low (<i>n</i> = 10)	Middle (<i>n</i> = 26)	High (<i>n</i> = 13)	Mixed (<i>n</i> = 15)		
<i>Materials used</i>								
Traditional selections	28.2	47.0*	30.0	40.4	42.3	33.2	37.9	(33.9)
Nonfiction (essays, biography, autobiography)	20.2	17.4	15.0	19.2	21.2	18.3	18.8	(23.6)
Drama	18.6	13.6	7.5	21.2	9.6	18.3	16.1	(24.9)
Young adult literature	21.7	2.2*	10.0	1.9	5.8	35.0	11.7	(18.3)
Poetry	7.3	12.4*	18.3	5.8	7.7	13.3	9.9	(18.8)
<i>Writing activities observed</i>								
Short-answer exercises	37.9	32.1	25.8	32.7	34.6	45.0	34.9	(25.8)
Note taking	32.2	45.7	40.8	40.4	42.3	33.3	39.2	(31.1)
Records or reports	43.6	54.0	53.3	59.6	38.5	36.7	49.0	(40.1)
Analysis (including persuasion)	41.9	43.7	39.2	47.1	42.3	38.3	42.8	(35.1)
Imaginative writing	25.0	7.6*	17.5	17.3	5.8	21.7	16.0	(25.3)
Personal experience	8.9	18.2	12.5	17.3	7.7	13.3	13.7	(21.3)

*Significant differences between middle and high school, or between high and low tracks, $p < .05$.

Table 3 averages the occurrences for the observed activities across the four observations of each classroom. Most of the classes that we observed included traditional literary selections, usually novels or short stories (these were not coded separately). Writing activities during a lesson or assigned as homework required approximately equal amounts of reporting and analyzing, although a considerable proportion of the classes also involved note taking or short-answer activities (typically, study questions). In comparison with high school classes, middle school classes were more likely to use young adult literature and less likely to use traditional selections and poetry. The amount of imaginative writing (usually stories) was also higher in the middle school classes. Average differences by track also appeared to be substantial: Upper-track classes read more traditional literature and essays; lower-track classes read more young adult literature and poetry. However, because we found a great deal of variation within tracks, none of the differences between high and low tracks were statistically significant.

Relationships Among Instructional Variables

To assess the relationships among measures associated with dialogic instruction, emphasis on envisionment building, extended curricular conversations, and high academic demands, we carried out a principal components analysis with Varimax rotation. (Since these are all classroom-level measures, the n for this analysis is 64.) A two-factor solution provided the best balance between interpretability and power to explain the original variation. The rotated solution, summarized in Table 4, includes two components. The first, which we interpret as a measure of emphasis on discussion-based approaches to the development of understanding, is defined by variables that reflect emphases on dialogic instruction, envisionment building, and extended curricular conversations. The second is defined by variables reflecting high academic demands for student work. The separation between the two sets of variables is quite clear, although two of the instructional variables (uptake and coherence) load less fully on their respective factors than do the remainder of the measures.

These results suggest that dialogic instruction, envisionment building, and emphasis on extended curricular conversations are in fact related aspects of a common emphasis on discussion-based instructional activities that support the development of understanding. The results also indicate that an emphasis on high academic demands (as reflected primarily in the amount of academic work that students are expected to do) is independent of the other aspects of instruction that we studied. Factor score estimates of both rotated components were generated directly from the principal components analysis for use in our remaining investigations of the relationships between instruction and performance.

Relationships Between Instruction and Performance

The key relationships between variations in instruction (as reflected in the degree of emphasis on discussion-based approaches and on high academic

Table 4
Rotated Component Matrix for Classroom-Level Variables

Variable	Component	
	1	2
<i>Evidence of dialogic instruction</i>		
Authentic teacher questions ^a	.802	-.101
Open discussion ^a	.830	.040
Questions with uptake	.450	.062
<i>Evidence of envisionment building</i>		
Emphasis on envisionment-building activities	.842	.184
Ratings of envisionment-building activities	.788	.049
<i>Evidence of extended curricular conversations</i>		
Observed curricular conversations	.658	.288
Coherence within lessons (teacher report)	.429	.265
<i>Evidence of high academic demands</i>		
Emphasis on revision of mechanics (student aggregate)	-.018	.864
Emphasis on revision of content (student aggregate)	.126	.818
Hours of English homework per week	.028	.735
Students complete reading and writing assignments	.351	.524

Note. Principal components analysis with rotation to the Varimax criterion. Components 1 and 2 together account for 55.1% of the original variation. $n = 64$ classrooms.

^aAnalyzed with a square root transformation to stabilize the variance.

demands) and students' literacy performance were investigated with a series of hierarchical linear models. With HLM, an initial analysis without any of the independent variables partitions the variation among levels: In our case, 10% of the variation in spring performance was accounted for at the school level, 29% at the classroom level, and 61% at the student level. Reliability estimates derived from the HLM analysis (Raudenbush & Bryk, 2002, p. 230) were .85 for class-level means and .47 for school-level means. The lower reliability at the school level reflects the relatively small number of cases at that level and suggests that tests of school-level effects will have limited power. As the next step in the analysis, we added a series of control variables at school and student levels, including urbanicity; grade level of the school (middle school was coded 0, high school was coded 1); fall performance; grade point average for subjects other than English; gender (female); home resources; and race/ethnicity (African American, Hispanic American, or Asian American). The results of these analyses (Table 5, Model 1) indicate that the control variables functioned as we expected, on the basis of the design of the study and the results of previous research (e.g., Donahue et al., 1999). At the school level, spring performance was lower in urban schools than in suburban schools and higher in high schools than in middle schools. At the student level, fall performance, overall grade point average, being female rather than male, not being Hispanic American, and having more resources in the home were all associated with better performance on the spring performance measure, $p < .05$,

Table 5
HLM Estimates of Relationships to Spring Performance

Variable	Control (Model 1)		Track (Model 2)		Instruction (Model 3)		Track and instruction (Model 4)	
	<i>b</i> (<i>SE</i>)	<i>p</i>	<i>b</i> (<i>SE</i>)	<i>p</i>	<i>b</i> (<i>SE</i>)	<i>p</i>	<i>b</i> (<i>SE</i>)	<i>p</i>
Intercept	−0.30 (.088)	.004	−0.28 (.122)	.037	−.29 (.083)	.004	−.28 (.107)	.021
<i>School effects (control variables)</i>								
High school (versus middle school) ^a	.44 (.100)	.001	.35 (.136)	.022	.29 (.096)	.008	.32 (.115)	.013
School is urban (versus suburban) ^a	−0.19 (.107)	.099	−0.22 (.117)	.083	.08 (.105)	.448	−.17 (.111)	.155
School is untracked ^a			−0.04 (.160)	.802			−.02 (.138)	.874
<i>Classroom effects</i>								
Track level ^a			.41 (.069)	.001			.27 (.083)	.002
Discussion-based approaches (Component 1)					.15 (.043)	.001	.11 (.043)	.013
High academic demands (Component 2)					.21 (.051)	.001	.11 (.059)	.067
<i>Student effects (control variables)</i>								
Fall performance	.31 (.028)	.001	.29 (.028)	.001	.30 (.028)	.001	.29 (.028)	.001
Grade point average (excluding English)	.13 (.027)	.001	.12 (.027)	.001	.11 (.027)	.001	.11 (.027)	.001
Female ^a	.29 (.049)	.001	.28 (.049)	.001	.28 (.049)	.001	.28 (.049)	.001
Home resources	.05 (.028)	.103	.04 (.028)	.204	.04 (.028)	.201	.03 (.028)	.235
African American ^a	−.07 (.075)	.344	−.07 (.074)	.364	.04 (.074)	.591	−.05 (.074)	.492
Hispanic American ^a	−.12 (.063)	.060	−.10 (.063)	.095	−.08 (.063)	.195	−.09 (.063)	.171
Asian American ^a	−.03 (.103)	.771	−.02 (.103)	.872	.00 (.103)	.978	.00 (.103)	.989

Note. From a series of hierarchical linear models with three levels (school, classroom, and student). $n = 19$ schools, 64 classes, and 974 students. Estimates were derived from the computer program HLM 5. b = regression weight. Figures in parentheses represent SE .

^aCoded 0, 1, and uncentered in HLM analyses, except for track level, which is coded −1, 0, 1. All other variables are standardized (z scores).

one-tailed. Together, the control variables explained 99.9% of the original variation at the school level, 61.8% at the classroom level, and 15.6% at the individual level. (Note that at the classroom level, with no separate control variables, this means that a large part of the between-classroom variance was accounted for by measured differences at the school and student levels.)

The investigation of instructional effects (measured in this study by high academic demands and emphasis on discussion-based approaches) is complicated by the links between tracking and instruction. As in previous work (Gamoran et al., 1995), track levels and instructional quality are positively correlated in the present data, and estimates of either can be inflated when the other is omitted. To provide as full a picture as possible, we estimated three models, also included in Table 5: one included track (Model 2), another included the instructional variables (Model 3), and the third included both track and instructional variables (Model 4). Looking at the coefficients for track and for instruction across the three models, it is clear that track level, an emphasis on discussion-based approaches, and high academic demands are all related to higher levels of performance (reflected in the p values for the regression coefficients). This is clear when tracking and instruction are considered in separate models (Models 2 and 3) and when they are included together in the same analysis (Model 4). The coefficients for all three effects, however, are noticeably reduced when they are included in the same model. The coefficient for track level drops by 34.1% ($100 \times [.41 - .27]/.41$) when the two instructional measures are added; that for academic demand drops by 47.6% when tracking is included (although it is still significant, $p < .03$, one-tailed). The coefficient for discussion-based approaches also falls when tracking is included, but somewhat more moderately (26.7%). The difference in impact on the coefficients for the two instructional measures suggests that the experiences of students in high and low tracks are differentiated by the degree of emphasis on discussion-based approaches, but even more so by the level of academic demands. At the same time, the nonsignificant regression coefficient for tracked as opposed to untracked schools indicates that, overall, students in schools without tracking performed as well as those in schools with tracking.

The results for the instructional variables indicate that students in classrooms with high academic demands and more emphasis on discussion-based approaches show higher end-of-year literacy performance across track levels. If we consider the coefficients in the combined model (Model 4), discussion-based approaches and high academic demands each have a classroom-level effect size of .53 (regression coefficient divided by a classroom-level standard deviation of .207), and track has a classroom-level effect size of 1.30.

Another way to look at these results is in terms of the percentage of between-classroom variance explained. For Models 1–4, the figures for between-classroom variance are .11478, .04869, .06144, and .04322, respectively. Tracking, when it alone is entered after the control variables in Model 1, thus accounts for 57.6% of the remaining variance ($100 \times [.11478 - .04869]/.11478$); instruction, when entered alone after the control variables, accounts

for 46.5% of the variance remaining; tracking and instruction, when entered together, account for 62.3% of the variance remaining after the control variables.

The design of the study, with planned contrasts between suburban and urban schools and between middle schools and high schools, also allows us to examine whether the approaches to instruction that we studied were differentially effective in any of these situations. The results for Model 4 (Table 5) indicated that spring performance (controlling for fall performance and the other control variables) tended to be higher in high schools than in middle schools, with a trend toward lower performance in urban as compared with suburban schools. Model 5 (Table 6, p. 720) adds cross-level interactions between the instructional variables (at the classroom level) and the planned contrasts (at the school level). None of these interactions are significant, suggesting that high academic demands and discussion-based approaches are effective in both middle and high schools and in both urban and suburban settings.

Given the overlap between tracking and instruction, it is important to consider whether various subgroups of students responded differently to the patterns of instruction captured in our two instructional variables. That is, did some students respond better than others to high academic demands and greater emphasis on discussion-based approaches? The first part of the analysis focused on how students of different ability responded to the instructional variables. Ability was captured in two ways: by track placement (at the classroom level) and by grade point average in subjects other than English (at the individual level). The interactions between these variables and the two measures of instruction are presented in Model 6 (Table 6). None of the interactions is significant, indicating that within this sample, better and poorer students benefited equally from discussion-based approaches and from high academic demands.

Another set of subgroups within our sample is defined by race/ethnicity. Again, to examine whether high academic demands and discussion-based approaches were beneficial for all students, we examined the interactions between these variables and race/ethnicity. The results are presented in Table 6, Model 7 (all interactions) and Model 8 (trimmed to exclude non-significant interactions).² The only significant interactions occurred among Asian American students; the pattern of regression coefficients indicates that although discussion-based approaches and high academic demands were beneficial to all students, Asian American students responded even more positively to such instruction than did their peers from other racial and ethnic groups. (The 61 Asian American students in the sample were distributed across 23 of the 64 participating classrooms; only 5 classes had 20% or more Asian Americans, and none had more than 27%. Thus this effect seems due to a difference in the responses of Asian Americans as compared with their peers in the same classes.)

Discussion

This study was undertaken to examine the relationship between (a) discussion-based approaches to challenging academic work in a diverse set of classrooms, and (b) the complex literacy skills that are reflected in tasks that require

students to write effectively about what has been read or experienced, mustering arguments and appropriate evidence to support an individual point of view. Discussion-based approaches have been widely advocated in recent years as beneficial to the learning of English language arts in general (e.g., Mayher, 1990), as well as to the learning of more specialized skills, such as reading comprehension strategies (e.g., Beck, McKeown, Hamilton, & Kucan, 1997). Previous work in this area has provided rich analyses of the unfolding of discussion over time, as well as experimental comparisons of the effects of discussion on writing performance (Sweigart, 1991) and on specific measures of reading comprehension and strategy use in small but well-controlled classroom settings (Langenberg, 2000). In the present study, we sought to provide evidence that an emphasis on discussion-based approaches, coupled with high academic demands, is positively related to literacy performance across a diverse set of classrooms at the middle and high school levels.

The overall results of the analyses reported here are strikingly consistent: We found that high academic demands and discussion-based approaches were significantly related to spring performance, with controls for initial literacy levels, gender, socioeconomic status, and race/ethnicity. Moreover, the lack of significant interactions between these measures and grade level of the school (middle school or high school), school location (urban or suburban), and academic ability (defined by track placement and grade point average in classes other than English) indicates that these approaches were effective across a range of situations, for students of varying levels of academic ability, whatever classrooms they were in. Analyses of interactions between the overall measures and race/ethnicity showed similar patterns of overall effectiveness, although students in one ethnic group—Asian Americans—responded even more positively than did their peers to both discussion-based approaches and high academic demands.

Another feature of the results is the extent to which effective instruction is correlated with tracking. There were significant differences between tracks on many measures, and regression coefficients were reduced considerably when tracking variables were included in the models. The impact of high academic demands, in particular, was greatly reduced when track level was included in the model, to the point where it shows only a trend toward significance in the final model (Model 8, $p < .07$, one-tailed). Taken together, the pattern of differences suggests that lower-track students have less engagement in all aspects of effective English instruction: dialogic instruction, environment-building activities, extended curricular conversations, and high academic demands. Time spent on open discussion was indicative of the general pattern: Not only did the means of high- and low-track classes differ significantly, but the low-track classes also showed a significant restriction in range. The observed maximum for average minutes of open discussion per hour in low-track classes was 3.7, as opposed to 14.5 minutes in high-track classes. In these circumstances, it becomes much harder to determine how well discussion-based approaches work for lower-track students; to some extent, they have not been tried.

Table 6
HLM Estimates of Interactions Between Classroom Variables and School Context, Track, GPA, and Race/Ethnicity

Classroom or student effect	Interaction with urbanicity and school grade level (Model 5)		Interaction with track and GPA (Model 6)		Interaction with race/ethnicity (Model 7)		Trimmed interaction (Model 8)	
	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>
Classroom effects^a								
Track level	.25 (.092)	.007	.25 (.085)	.005	.27 (.082)	.001	.27 (.082)	.001
<i>Discussion-based approaches (Component 1)</i>	.13 (.060)	.028	.07 (.049)	.139	.07 (.048)	.155	.10 (.043)	.023
Discussion × urban	-.03 (.099)	.729						
Discussion × high school	-.05 (.091)	.563						
Discussion × untracked school			.14 (.121)	.242				
Discussion × track			.03 (.072)	.645				
<i>High academic demands (Component 2)</i>	.10 (.090)	.261	.12 (.063)	.055	.06 (.067)	.338	.09 (.058)	.144
Academic demands × urban	-.10 (.093)	.272						
Academic demands × high school	.10 (.089)	.257						
Academic demands × untracked school			-.12 (.128)	.350				
Academic demands × track			.03 (.067)	.708				

Student effects^a

<i>GPA (excluding English)</i>	.11 (.027)	.001	.11 (.027)	.001	.11 (.027)	.001	.11 (.027)	.001
<i>GPA × discussion-based approaches</i>			.00 (.027)	.939				
<i>GPA × high academic demands</i>			-.04 (.030)	.153				
<i>African American</i>	-.06 (.074)	.444	-.06 (.074)	.402	-.06 (.076)	.417	-.06 (.068)	.408
<i>African American × discussion-based approaches</i>					.05 (.068)	.434		
<i>African American × high academic demands</i>					.02 (.089)	.785		
<i>Hispanic American</i>	-.09 (.062)	.164	-.09 (.063)	.148	-.09 (.064)	.159	-.09 (.041)	.029
<i>Hispanic American × discussion-based approaches</i>					.05 (.057)	.338		
<i>Hispanic American × high academic demands</i>					.05 (.070)	.500		
<i>Asian American</i>	.00 (.103)	.975	-.01 (.103)	.959	.05 (.112)	.681	.05 (.111)	.456
<i>Asian American × discussion-based approaches</i>					.36 (.112)	.002	.34 (.110)	.003
<i>Asian American × high academic demands</i>					.25 (.113)	.024	.24 (.110)	.030

Note. From a series of hierarchical linear models using HLM 5 with three levels (school, classroom, and student). In addition to the variables displayed, all models included an intercept, school grade level, urbanicity, tracked/untracked school, fall performance, gender, and home resources. $n = 19$ schools, 64 classes, and 974 students. b = regression weight. Figures in parentheses represent SE .

^aDummy variables for urbanicity, school grade level, untracked school, and race/ethnicity are coded 0, 1, and uncentered in HLM analyses. Track is coded -1, 0, 1. All other variables are standardized (z scores).

Classrooms are complex places in which instruction always involves an ongoing negotiation of roles and relationships among teachers, students, and subject matter, which plays out during a lesson and over time. Previous studies have provided a rich array of detailed analyses of how such negotiations evolve, both in individual lessons and over time, at various grade levels and in various institutional contexts (Athanases, 1998; Dyson, 1997; Gutierrez, 1994; Hillocks, 1999; Sperling & Woodlief, 1997). Yet even the brief excerpts from classroom activity presented in this article reflect some fundamental differences in approaches to teaching and learning that remain quite stable as other aspects of the interactions evolve. The discussion-based approaches that were investigated in the present study reflect different roles for teachers and students, different assumptions about what constitutes effective teaching, and different conceptualizations of what should count as doing English well. The approaches that contributed most to student performance on the complex literacy tasks that we administered were those that used discussion to develop comprehensive understanding, encouraging exploration and multiple perspectives rather than focusing on correct interpretations and predetermined conclusions.

Emphases on discussion-based approaches and high academic demands were important influences on literacy performance across the range of classrooms that we studied. Middle and high school classrooms, for example, differed somewhat in the materials studied (e.g., *To Kill a Mockingbird* in the middle school, the *Iliad* in the high school). But at both levels, high academic demands and discussion-based approaches were significantly related to literacy performance.

A major difference between the present investigation and much of the related research on discussion-based instruction in writing or comprehension is the generic nature of the activities that we investigated. We focused on dialogic discussion, envisionment-building activities, extended curricular conversations, and high academic demands that could be manifested in a variety of ways in individual classrooms, rather than focus on specific discussion strategies or specific scaffolding techniques. In contrast, research on comprehension strategy instruction, for example, typically has focused on an array of specific techniques for structuring discussion and embedding comprehension strategies. Our measures of discussion-based activities focused not on a technique or set of techniques but, more generally, on the presence and extent of discussion and related activities designed to involve students in the exploration of ideas. The positive results that we obtained suggest that the spontaneous scaffolding or support for developing ideas that are generated during open discussions is a powerful tool for learning. This conclusion parallels one from the National Reading Panel review of comprehension strategy instruction (Langenberg, 2000), which found particular strength in approaches that involved a variety of strategies embedded in the natural flow of classroom discussion of difficult texts, because skilled reading “involves an ongoing adaptation of multiple cognitive processes” (p. 4.47).

Some limitations in our study should be noted. One has to do with the measure of literacy performance and its relationship to classroom experiences. The character analysis task is designed to tie performance to the curriculum dur-

ing the year, but the link is not as tight as in Nystrand's previous study (1997), which used teachers' reports of works taught each week to fashion a literature test directly tied to the curriculum. Taken together, the two tasks that we used for the spring assessment captured aspects of comprehension, literary analysis, and general writing ability; but, clearly, there are many other components of performance in English that are not tapped by these measures and that might show different relationships to measures of curriculum and instruction.

Another measurement issue concerns the degree of emphasis on extended curricular conversations. In the trade-off between larger numbers and more intensive data gathering, one casualty was the ability to monitor curriculum as it evolved over time. The substitute in the present study consisted of more limited measures of connections among activities reported by the teacher and connections within and across lessons that emerged during the course of the daily interactions recorded by the observers. These are important components of extended curricular conversations, but future research needs to look in more depth at the topics in which students are engaged and how those topics develop over extended periods of time. Our earlier studies have shown, for example, that being engaged in dialogic interaction contributes to performance only if the discussion focuses on academic topics (Nystrand, 1997).

Another limitation has to do with the range of variation within the sample: Although schools and classrooms were selected purposively rather than randomly to increase variation in approaches to instruction, the majority of classrooms still looked quite traditional and teacher-dominated. This relative homogeneity limited the examination of the relationships between instruction and differences in student performance. In particular, it made it difficult to investigate the most appropriate balance among instructional approaches that emphasize discussion and approaches that emphasize other kinds of activities. We do not know how much is enough.

Conclusion

The results of the present study support emphasis on high academic demands and discussion-based approaches that involve dialogic instruction, envisionment-building activities, and extended curricular conversations about important academic topics. The discussion-based approaches required changes on many levels: in the structure of moment-to-moment interactions among students and their teachers; in the ways that activities were orchestrated to support students' developing understandings; and (to the extent that we could measure it) in the overall shape of the curriculum. All of these changes reflect an underlying emphasis on providing students with needed knowledge and strategies in the course of an active exploration of new ideas and experiences through challenging reading, writing, and discussion. From the sociocognitive perspective with which we began, the results suggest that when students' classroom experiences emphasize high academic demands and discussion-based approaches to the development of understanding, students internalize the knowledge and skills necessary to engage in challenging literacy tasks on their own.

APPENDIX

Rubrics for Scoring Literacy Assessments

Level of Abstraction

- 0 = Generalization without any development; or report or generalization with irrelevant information; or writing sample that does not answer the question.
- 1 = Record (as in a lab notebook): What's happening?
- 2 = Report: The writer gives an account of a particular series of events or thoughts or feelings (i.e., a narrative): What happened?
- 3 = Generalization, generalized narrative, or descriptive information (including descriptions of physical features, activities, and cognitive experiences). The writer is tied to particular events and places but detects a pattern of repetition in them and expresses the pattern in general form: What seems to be happening?
- 4 = Low-level analysis: The writer makes genuine analyses, but organization is loose and relationships between/among points are not perceived and/or not made explicit. Major points in the text might be shuffled without altering the meaning or effectiveness of the text.
- 5 = Analysis: The writer relates points hierarchically or logically by means of coherently presented classificatory utterances (e.g., thesis statements, topic sentences, transitional expressions), producing a highly wrought text.

Level of Elaboration

- 1 = Unsatisfactory: The writer provides only the barest information, misinformation, or disjointed details.
 - 2 = Minimal: The writer provides some details but in unrelated ways. No organizational framework is created for the reader to use in understanding how the various pieces of information in the sample relate to each other. Major points in the text might be shuffled without altering the meaning of the text. This category includes texts that are organized merely by indicators such as "First," "Second," "Third," and "Finally."
 - 3 = Adequate: The writer describes and interrelates most of the information and presents details within a clear, coherent (although perhaps implicit) organizing framework.
 - 4 = Elaborated: The writer composes an extended description within a cohesive and explicit organizing framework to provide a context for the reader. The meaning or effectiveness of the text would be altered if the order of points were disturbed. The text is highly wrought and tightly organized.
-

Notes

The research reported here was conducted at the National Research Center on English Learning and Achievement (CELA), an R&D center located at the University at Albany, State University of New York, in collaboration with the University of Wisconsin, Madison. CELA's work is sponsored by the Institute of Education Sciences (IES), formerly the Office of Educational Research and Improvement, U.S. Department of Education. Our research was supported in part under the Research and Development Centers Program (Award No. R305A960005), administered by IES. However, the contents do not necessarily represent the positions or policies of the Department of Education or IES.

We would like to thank the many individuals who contributed to this study. Field researchers included Samantha Caughlan, Yu-Min Chien, Elena Hernandez, Sheila Flihan, and Mickey Young. Additional data editors included Mary Juzwik, Kevin O'Connor, and Kathy Platz. Special thanks are also due to Sean Kelly, who developed and maintained the database on which our analyses are based. Last but not least, we owe a special debt of gratitude to the teachers and students who completed our instruments and shared their classrooms with us.

¹Other scholars use the terms *sociocultural* or *situated* to describe a similar focus on the ways in which language, learning, and cognition are inseparably intertwined with the social or cultural context within which they are embedded (cf. Brown, Collins, & Duguid, 1989; Lave, 1988; Lave & Wenger, 1991; Greeno & Middle School through Applications Project Group, 1998; Wells, 1999).

²In interpreting the size and significance of effects in the final model (Model 8), it is important to remember that we are using multilevel models with different degrees of freedom and different standard deviations of the intercept when applied to effects at the school level ($SD = .121$, $df = 15$), the classroom level ($SD = .203$, $df = 60$), and the individual level ($SD = .719$, $df = 958$). The values reported in the table can be interpreted as effect sizes based on the standard deviations of the original data because all variables, including spring performance, were entered into the analysis as contrasts, dummy variables, or z scores (calculated at the level at which the variable was measured). For any individual student, these values allow a direct calculation of a predicted score on spring performance. However, if effect sizes are calculated to reflect the nested structure of the data and the other variables included in the analysis, they increase at all levels (because all standard deviations are lower than 1.0); but the effect sizes at the classroom and school levels increase the most. For example, adjusting for the level of analysis in Model 8 yields an effect size of .47 at the individual level for the interaction of Asian background with high-quality instruction (as compared with .34 in the original metric); a similar adjustment for high-quality instruction yields an effect size of .49 at the classroom level (as compared with .10 in the original metric). These effect sizes are more comparable to those calculated in between-group analysis-of-variance or covariance designs.

References

- Adams, M. J., & Collins, A. M. (1979). A schema-theoretic view of reading. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 1–22). Norwood, NJ: Ablex.
- Alvermann, D. E., Young, J. P., Weaver, D., Hinchman, K. A., Moore, D. W., Phelps, S. F., et al. (1996). Middle and high school students' perceptions of how they experience text-based discussions: A multicase study. *Reading Research Quarterly*, *31*(3), 244–267.
- Anderson, R. C. (1977). The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 415–431). Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, R. C., Wilkinson, I. A. G., & Mason, J. M. (1991). A microanalysis of the small-group, guided reading lesson: Effects of an emphasis on global story meaning. *Reading Research Quarterly*, *26*(4), 417–441.
- Applebee, A. N. (1974). *Tradition and reform in the teaching of English: A history*. Urbana, IL: National Council of Teachers of English.
- Applebee, A. N. (1976). Microteaching, component skills, and the training of teachers: An evaluation of a research and development project. *British Journal of Educational Technology*, *2*(7), 35–44.
- Applebee, A. N. (1981). *Writing in the secondary school: English and the content areas* (Research Monograph 21). Urbana, IL: National Council of Teachers of English.
- Applebee, A. N. (1984). *Contexts for learning to write: Studies of secondary school instruction*. Norwood, NJ: Ablex.
- Applebee, A. N. (1993). *Literature in the secondary school: Studies of curriculum and instruction in the United States*. Urbana, IL: National Council of Teachers of English.
- Applebee, A. N. (1996). *Curriculum as conversation: Transforming traditions of teaching and learning*. Chicago: University of Chicago Press.
- Applebee, A. N., Burroughs, R., & Stevens, A. S. (2000). Creating continuity and coherence in high school literature curricula. *Research in the Teaching of English*, *34*, 396–429.

- Applebee, A. N., Langer, J. A., Mullis, I. V. S., & Jenkins, L. (1990). *The writing report card, 1984-88: Findings from the nation's report card*. Princeton, NJ: National Assessment of Educational Progress.
- Athanases, S. Z. (1998). Diverse learners, diverse texts: Exploring identity and difference through literary encounters. *Journal of Literacy Research, 30*(2), 273-296.
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin: University of Texas Press.
- Beck, I. L., McKeown, M. G., Sandora, C., & Worthy, J. (1996). Questioning the author: A year-long classroom implementation to engage students with text. *Elementary School Journal, 96*(4), 385-414.
- Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text*. Newark, DE: International Reading Association.
- Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum Inquiry, 17*(1), 9-30.
- Boyd, M. P., & Rubin, D. L. (2002). Elaborated student talk in an elementary ESoL classroom. *Research in the Teaching of English, 36*(4), 495-530.
- Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). *Research in written composition*. Urbana, IL: National Council of Teachers of English.
- Brandt, D. (1998). Sponsors of literacy. *College Composition and Communication, 49*(2), 165-185.
- Brandt, D. (2001). *Literacy in American lives*. Cambridge, UK: Cambridge University Press.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Research Council.
- Brewer, D. J., Rees, D. I., & Argys, L. M. (1995). Detracking America's schools: The reform without cost? *Phi Delta Kappan, 77*, 210-212.
- Britton, J. N., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities (11-18)*. London: Macmillan Educational for the Schools Council.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second grade readers. *Journal of Educational Psychology, 88*, 18-37.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth: A collaboration at the Center for Cognitive Studies*. New York: Wiley.
- Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal, 39*, 801-827.
- Carroll, D. (2000). *NELS 88/94 public use data files and electronic codebook: Base year through third follow-up* (Document No. NCES 2000238). Washington, DC: National Center for Education Statistics.
- Cazden, C. B. (1979). Peekaboo as an instructional model: Discourse development at home and at school. *Papers and Reports on Child Language Development, 17*, 1-19.
- Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.
- Chinn, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly, 36*(4), 378-411.
- Committee of Ten on Secondary School Studies. (1894). *Report of the Committee of Ten on Secondary School Studies, with the reports of the conferences arranged by the committee*. New York: American Book Co. for the National Education Association.
- Donahue, P. L., Voelkl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states*. Washington, DC: U.S. Department

Discussion-Based Approaches to Developing Understanding

- of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
- Durkin, D. (1978–1979). What classroom observations reveal about reading comprehension. *Reading Research Quarterly*, 14, 491–533.
- Durst, R. K. (1987). Cognitive and linguistic demands of analytic writing. *Research in the Teaching of English*, 21(4), 347–376.
- Durst, R. K. (1999). *Collision course: Conflict, negotiation, and learning in college composition*. Urbana, IL: National Council of Teachers of English.
- Dyson, A. H. (1984). Learning to write/learning to do school. *Research in the Teaching of English*, 18(3), 233–264.
- Dyson, A. H. (1997). *Writing superheroes: Contemporary childhood, popular culture, and classroom literacy*. New York: Teachers College Press.
- Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English*, 23(1), 4–29.
- Emig, J. (1971). *The composing processes of twelfth graders* (Research Monograph 13). Urbana, IL: National Council of Teachers of English.
- Fillmore, C. J. (1982). Ideal readers and real readers. In D. Tannen (Ed.), *Analyzing discourses: Text and talk* [Georgetown University Round Table on Languages and Linguistics, 1981] (pp. 248–270). Washington, DC: Georgetown University Press.
- Finley, M. K. (1984). Teachers and tracking in a comprehensive high school. *Sociology of Education*, 57, 233–243.
- Flower, L. S., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, 23, 2–32.
- Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365–387.
- Freedman, S. W., Simons, E. R., Kalnin, J. S., Casareno, A., & the M-Class Teams. (1999). *Inside city schools: Investigating literacy in multicultural classrooms*. New York: Teachers College Press; Urbana, IL: National Council of Teachers of English.
- Gage, N. L. (1963). Paradigms for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 94–141). Chicago: Rand-McNally.
- Gage, N. L. (1994). The scientific status of research on teaching. *Educational Theory*, 44(4), 371–383.
- Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *Elementary School Journal*, 89(3), 253–300.
- Gambrell, L. B., & Almasi, J. F. (Eds.). (1996). *Lively discussions! Fostering engaged reading*. Newark, DE: International Reading Association.
- Gamoran, A. (1993). Alternative uses of ability grouping in secondary schools: Can we bring high-quality instruction to low-ability classes? *American Journal of Education*, 101, 1–22.
- Gamoran, A. (2000). High standards: A strategy for equalizing opportunities to learn? In R. D. Kahlenberg (Ed.), *A notion at risk: Preserving public education as an engine for social responsibility* (pp. 93–126). New York: Century Foundation.
- Gamoran, A., & Carbonaro, W. J. (2002–2003). High school English: A national portrait. *High School Journal*, 86(2), 1–13.
- Gamoran, A., & Kelly, S. (in press). Tracking, instruction, and unequal literacy in secondary school English. In M. T. Hallinan, A. Gamoran, W. Kubitschek, & T. Loveless (Eds.), *Stability and change in American education: Structure, processes, and outcomes*. Clinton Corners, NY: Eliot Werner.
- Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology*, 94, 1146–1183.
- Gamoran, A., Nystrand, M., Berends, M., & LePore, P. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal*, 32, 687–716.

- Garrison, J. W., & Macmillan, C. J. B. (1984). A philosophical critique of process-product research on teaching. *Educational Theory*, 34(3), 255–274.
- Garrison, J. W., & Macmillan, C. J. B. (1994). Process-product research on teaching: Ten years later. *Educational Theory*, 44(4), 385–397.
- Gee, J. P. (1996). *Social linguistics and literacies: Ideology in discourses* (2nd ed.). Bristol, PA: Taylor & Francis.
- Graves, D. (1983). *Writing: Teachers and children at work*. Exeter, NH: Heinemann.
- Greeno, J. G., & Middle School through Applications Project Group. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53, 5–26.
- Gregg, L., & Steinberg, E. (Eds.). (1980). *Cognitive processes in writing*. Hillsdale, NJ: Erlbaum.
- Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Guthrie, J. T., Schafer, W., Wang, Y. Y., & Afflerbach, P. (1995). Relationships of instruction to amount of reading: An exploration of social, cognitive, and instructional connections. *Reading Research Quarterly*, 30(1), 8–25.
- Gutierrez, K. D. (1994). How talk, context, and script shape contexts for learning: A cross-case comparison of journal sharing. *Linguistics and Education*, 5, 335–365.
- Heath, S. B. (1983). *Ways with words*. New York: Cambridge University Press.
- Hillocks, G., Jr. (1999). *Ways of thinking, ways of teaching*. New York: Teachers College Press.
- Hynds, S. (1997). *On the brink: Negotiating literature and life with adolescents*. New York: Teachers College Press.
- Kay, P. (1987). Three properties of the ideal reader. In R. Freedle & R. Duran (Eds.), *Cognitive and linguistic analysis of test performance* (pp. 208–224). Norwood, NJ: Ablex.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67(3), 271–299.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491.
- Langenberg, D. N. (2000). *Report of the National Reading Panel: Teaching children to read: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development.
- Langer, J. A. (1981). From theory to practice: A pre-reading plan. *Journal of Reading*, 25, 152–156.
- Langer, J. A. (1984). Examining background knowledge and text comprehension. *Reading Research Quarterly*, 19(4), 468–481.
- Langer, J. A. (1985a). A sociocognitive view of literacy learning. *Research in the Teaching of English*, 19(4), 235–237.
- Langer, J. A. (1985b). Levels of questioning: An alternative view. *Reading Research Quarterly*, 21(5), 586–602.
- Langer, J. A. (1986). *Children reading and writing: Structures and strategies*. Norwood, NJ: Ablex.
- Langer, J. A. (1987a). A sociocognitive perspective on literacy. In J. A. Langer (Ed.), *Language, literacy, and culture: Issues of society and schooling* (pp. 1–20). Norwood, NJ: Ablex.
- Langer, J. A. (1987b). The construction of meaning and the assessment of comprehension. In R. Freedle & R. Duran (Eds.), *Cognitive and linguistic analyses of test performance* (pp. 225–244). Norwood, NJ: Ablex.
- Langer, J. A. (1990). The process of understanding: Reading for literary and informative purposes. *Research in the Teaching of English*, 24, 229–260.
- Langer, J. A. (1992). Discussion as exploration: Literature and the horizon of possibilities. In G. E. Newell & R. K. Durst (Eds.), *Exploring texts: The role of discus-*

Discussion-Based Approaches to Developing Understanding

- sion and writing in the teaching and learning of literature* (pp. 23–43). Norwood, MA: Christopher Gordon.
- Langer, J. A. (1995). *Envisioning literature: Literary understanding and literature instruction*. New York: Teachers College Press.
- Langer, J. A. (2000). Excellence in English in middle and high school: How teachers' professional lives support student achievement. *American Educational Research Journal*, 37(2), 397–439.
- Langer, J. A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal*, 38(4), 837–880.
- Langer, J. A. (2002). *Effective literacy instruction: Building successful reading and writing programs*. Urbana, IL: National Council of Teachers of English.
- Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning*. Urbana, IL: National Council of Teachers of English.
- Langer, J. A., Bartolome, L., Vasquez, O., & Lucas, T. (1990). Meaning construction in school literacy tasks: A study of bilingual students. *American Educational Research Journal*, 27, 427–471.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lee, C. D. (1993). *Signifying as a scaffold for literary interpretation: The pedagogical implications of an African American discourse genre*. Urbana, IL: National Council of Teachers of English.
- Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, 30(4), 608–630.
- Lee, C. D. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal*, 38(1), 97–141.
- Lee, C. D., & Smagorinsky, P. (Eds.). (2000). *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry*. Cambridge, UK: Cambridge University Press.
- Losey, K. M. (1995). Mexican American students and classroom interaction: An overview and critique. *Review of Educational Research*, 65(3), 283–318.
- Marshall, J. D., Smagorinsky, P., & Smith, M. (1995). *The language of interpretation: Patterns of discourse in discussions of literature*. Urbana, IL: National Council of Teachers of English.
- Mayher, J. S. (1990). *Uncommon sense: Theoretical practice in language education*. Portsmouth, NH: Boynton/Cook.
- Mehan, H. (1979). *Learning lessons*. Cambridge, MA: Harvard University Press.
- National Center for Education Statistics. (1994). *Curricular differentiation in public high schools* (Document No. NCES 95-360). Washington, DC: U.S. Department of Education.
- Nystrand, M. (1986). *The structure of written communication: Studies in reciprocity between writers and readers*. Orlando, FL: Academic Press.
- Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.
- Nystrand, M. (1999). *Classroom language assessment system (CLASS 3.0)*. Madison, WI: Center on English Learning and Achievement.
- Nystrand, M., Cohen, A., & Dowling, N. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment*, 1, 53–70.
- Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 25, 261–290.
- Nystrand, M., Gamoran, A., & Carbonaro, W. (2001). On the ecology of classroom instruction: The case of writing in high school English and social studies. In

- P. Tynjala, L. Mason, & K. Londa (Eds.), *Writing as a learning tool* (pp. 57–81). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Nystrand, M., Wu, L., Gamoran, A., Zeiser, S., & Long, D. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, 35, 135–196.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570–608). New York: Macmillan.
- Page, R. N. (1991). *Lower-track classrooms: A curricular and cultural perspective*. New York: Teachers College Press.
- Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English*, 13, 317–336.
- Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, 27, 91–109.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Rex, L. A. (2001). The remaking of a high school reader. *Reading Research Quarterly*, 36(3), 288–314.
- Sandora, C., Beck, I., & McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Journal of Reading Psychology*, 20, 177–212.
- Scholes, R. (1985). *Textual power: Literary theory and the teaching of English*. New Haven, CT: Yale University Press.
- Shaughnessy, M. (1977). *Errors and expectations*. New York: Oxford University Press.
- Sperling, M., & Woodlief, L. (1997). Two classrooms, two writing communities: Urban and suburban tenth-graders learning to write. *Research in the Teaching of English*, 31(2), 205–239.
- Street, B. (1995). *Social literacies*. London: Longman.
- Sweigart, William. (1991). Classroom talk, knowledge development, and writing. *Research in the Teaching of English*, 25, 469–496.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, UK: Cambridge University Press.
- Valli, L. (1990). A curriculum of effort: Tracking students in a Catholic high school. In R. N. Page & L. Valli (Eds.), *Curriculum differentiation: Interpretive studies in U.S. secondary schools* (pp. 45–65). Albany: State University of New York Press.
- Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press.
- Wells, G., & Chang-Wells, G. L. (1992). *Constructing knowledge together: Classrooms as centers of inquiry and literacy*. Portsmouth, NH: Heinemann.
- Willett, J. B. (1988–1989). Questions and answers in the measurement of change. *Review of Research in Education* 15, 345–422.

Manuscript received August 29, 2002

Revision received February 14, 2003

Accepted May 18, 2003