2

Taking a Long View of Reading Development

Connie Juel

INFLUENCES THAT MOTIVATED THE DIRECTION OF MY WORK

Like Isabel Beck, my interest in reading research started in the elementary classroom. In the early 1970s I was a fourth-grade elementary school teacher. Two things puzzled me about my children's reading: (1) Why was there such a huge range in reading ability among these 9-year-olds? (2) Why were there a couple of children each year who were still learning to decode at a very basic level? I worked closely with my very poor readers and saw how hard they struggled to learn and how their poor reading was affecting almost all aspects of their schoolwork. I can still remember the face of a 9-year-old girl absolutely aglow when she sang during music time; yet that face turned downward and unsmiling as she stared at a printed page. This young girl determined my future.

After school one day I drove down to my alma mater and entered a building I had never been in as an undergraduate—the School of Education. I inquired in the dean's office about who could help me help children learn to read. There seemed some confusion and some asking around, which I thought was odd, as this seemed like a pretty reasonable question to pose in this building. Eventually I was directed to the office of Professor Robert Calfee. Fortunately, he was in, and even

more fortunately, as I now know as a university professor myself, he was willing to invite this stranger in and to spend some time talking. I have to say that at the time I did not understand much of what he was saying, and as his doctoral student a couple of years later I would still struggle—but I came a lot closer. For he began by talking about design—specifically, fractional factorial research designs that could be used to carefully unpack the reading process. What I did understand at the time, though, was that to help my students with reading difficulties, I needed to understand the reading process—to *really* understand it, as he put it. A passion entered my mind; I wanted to do just that. So I would take a year off to study this process, get a master's degree, and head back to the classroom better equipped to help my students. But the passion turned into doing one of those fractional factorial studies on the reading process for my doctorate.

RESEARCH AGENDA

During the 1970s exploration into the reading process was dominated by psychology and cognitive information processing models (e.g., Gough, 1971). To start to test these types of models, factorial designs and analysis of variance worked well. For each hypothesized component process (what the boxes in the model represented), you needed to find some unique factors that influenced that process and that did not interact with other hypothesized component processes. You might hypothesize that decoding involved going through orthography and not subverting it by using top-down processing—that is, using knowledge of the topic you were reading rather than looking closely at letters in words as you read. You then needed to demonstrate that decoding was not affected by top-down factors but rather by factors unique to it (such as the regularity of a word's spelling patterns). This methodology would serve me well for several years as a researcher; I published several studies examining the reading process using factorial designs and this form of analysis.

Of course, boxes in a design of reading processes assume input by the reader. For each component process, you can consider how the learner's current ability, knowledge, and emotions affect that process. And, given that each individual is a unique constellation, there are likely different patterns for how the component processes might develop and interact and how the whole enterprise of reading development might be on different trajectories for every individual. There is likely, however,

to be considerable overlap in the "component processes" developed on these trajectories. Stage models of reading, which I participated in, tried to capture those commonalities.

For those of us whose passion was education, the next step was to consider what and how school instruction contributes to the process of learning to read and write. Schools involve teachers, group settings, and textbooks, among other things. These various dimensions interact both with the component processes of reading and with the individual learner's contribution to those processes. Brian Byrne (2005) views what the schools need to do as a problem of subtraction: the components of reading processes minus what the learner contributes to each. Complexity, he would note, is added because every child does not bring the same ability to generalize and reason or the same knowledge to a component.

Consider that a component process of reading is word recognition (however you wish to subdivide it). Children bring to this process differential knowledge about letter sounds and words and differential skill in generalization and transfer. Clearly, teachers cannot teach every word that children will see in print or every spelling–sound mapping. Ability to generalize and transfer will vary between children, and teachers will need to teach accordingly (Byrne, 2005).

I think that the roles decodable text and phonics play, to a large extent, is to encourage children to generalize. Juel and Roper/Schneider (1985) found that children could induce untaught letter–sound correspondences better in decodable text than in less decodable text. The text itself seemed to foster generalizing, reasoning, and transfer.

In trying to understand reading—to *really* understand reading, as my advisor had put it—we must understand its component processes, what the individual contributes, and all that is involved in instruction. There are two other dimensions that I have been concerned with, one more than the other, in my own research. First, there is a social contribution to some, if not all, aspects of reading. There are home, community, and peer interactions and discourse in general that is permeated with social interactions that influence reading. Reading often seems to be a solitary activity, and I think that dimension should be valued, but, clearly, social constructivist thinking has highlighted the social aspects of learning and comprehending.

Second, a longitudinal view of learning has been a very important dimension of study for me. What contributes to learning to read in first

grade, for example, may be something that was learned more in kindergarten than in first grade. Or knowledge that facilitates a process such as word recognition may change with age and exposure to text. It is the longitudinal aspect of my research work that I highlight in this chapter.

I have always been interested in how we change over time. Just as we aren't the same people today that we were as babies or teenagers or that we will be at the end of our lives, the reading process might change over time. From this perspective, I was interested in how classroom instruction affected the learning-to-read process and how the influence of this instruction would vary across time. If you were instructing a "baby" versus a "teenage" reader, how might instruction differentially affect the learner? And would instruction you received as a "baby" reader reach up to influence you as an older reader? Would, say, what you learned in kindergarten affect or even control what you learn in first grade?

My research agenda these past three decades has been to understand the reading process: the component processes of reading, the learner's contribution to these processes, and the instructional, social, and longitudinal contributions. My earliest work focused more, I initially thought, on the cognitive processes in reading than on instruction, but the outcome of that work would focus me squarely on instruction.

RELATIONSHIP OF RESEARCH TO INSTRUCTION

Early Research

Individual growth modeling (hierarchical linear modeling; HLM) was not available to reading researchers in the 1980s, or I would have been a quick convert. As it was, I began to use path analysis and regression to try to explore development of the reading process over time. The model I began with was heavily influenced by what would come to be known as the "simple view of reading" (Gough & Tunmer, 1986; Hoover & Gough, 1990). The simple view of reading is that reading comprehension is the product of two fundamental processes, word recognition and listening comprehension. If either process is missing, then no matter how good the other process, reading comprehension will be nonexistent. Assuming perfect word recognition, then, the reader comprehends written text as well as he or she would if the same text were spoken.

I also assumed a simple view of writing. This assumption led to the model in Figure 2.1. In a longitudinal study of first- and second-grade children, we tried to model development, particularly of word recognition skill (Juel, Griffith, & Gough, 1986). We thought that the foundational sources of knowledge that readers use to identify (or spell) words were cipher knowledge (i.e., the knowledge underlying spelling–sound patterns that enable pronunciation of pseudowords such as *buf* or *zlip*) and word-specific lexical knowledge (e.g., knowing that the "long" *e* in green is spelled *ee* rather than *ea* or *ene*). For accomplished readers, cipher knowledge may become a closed set, but we are always adding to our word-specific lexical knowledge. How, for example, do you spell *iridescence*—one or two r's? We thought word-specific lexical knowledge would come about largely through exposure to print, though for adults it may well be through writing.

Today it may be hard to think of a time when the term *phonemic awareness* was relatively unknown. It was not common to put phonemic awareness on the table in those days, but in this study we did. We thought it a prominent contributor, along with exposure to printed words, to a child's ability to create cipher knowledge. We also thought that whereas the first-grade child might be more dependent on cipher knowledge, the balance would shift in favor of lexical knowledge over time.

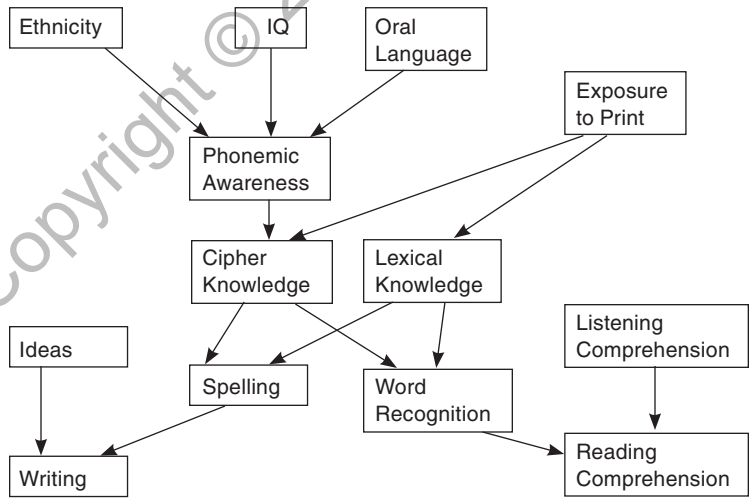

FIGURE 2.1. The simple model of literacy acquisition.

We were just starting to open the Pandora's box of phonemic awareness to look at what might predict it. We were studying children in Austin, Texas, who were about one-third Hispanic (I'm using labels from the locale and time period), one-third black, and one-third Anglo. We hypothesized that dialect and second language might influence development of phonemic awareness in English. Ethnicity did indeed influence phonemic awareness, but we could not verify our hypothesis in this study. Although we considered phonemic awareness as a precursor to developing cipher knowledge, I should have listened to my friend Isabel, who suggested that it developed in a reciprocal relation. That would be what other studies would find and what we would later jointly conclude (Beck & Juel, 1992). Like other researchers, we found generally low correlations among IQ, listening comprehension, and phonemic awareness. That is, phonemic awareness seemed to be a somewhat unique or modular understanding/skill in development: A child with either a high or a low IQ could have problems with phonemic awareness.

In terms of the simple view, we found that first-grade reading comprehension was largely determined by word recognition ability (defined, using a narrow view, as accurate pronunciation). Word recognition in first grade was largely determined by cipher knowledge in first grade, with a shift to lexical knowledge in second grade. These findings may hold as developmental trends. Two major views of word recognition that were emerging at this time, however, would indeed show this as too simple. First, Linnea Ehri's "amalgamation" theory and her studies suggest that these sources of knowledge are not so easily separated (Ehri, 1992; Ehri & Wilce, 1985). Second, "connectionist" computer-generated models of word recognition marked the beginning of research suggesting that almost every word is its own learning domain—that is, that it is words, rather than children, that go through, in a sense, "stages" of incremental knowledge additions (Plaut, 2005; Seidenberg & McClelland, 1989).

I wanted to know what happened to the children in our study after second grade, so I continued to follow their development through 2 more years, through fourth grade (Juel, 1988). I also learned that I wanted to talk to the participating children to see how things looked from their end. I wanted *their* words. This marked my departure into mixed methods. From here on, qualitative data would be a part of my quantitative studies. Even more, case studies became an interest. I took

eight children from this study and did expanded case studies, with extensive interviews and reading and writing samples (Juel, 1994).

Probably the two most quoted findings from my research work over the years came from a longitudinal study (Juel, 1988). One finding was that the probability of a child still being a poor reader at the end of fourth grade, given that the child was a poor reader at the end of first grade, was .88. This probability was cited as though it were gospel; I always found this a bit strange, because it was a relatively small sample, because the children were all from one school, and because there was no intervention in place. Still other researchers, across curricula and languages, have found a similar statistic, though some have not. I think our current emphasis on early intervention has helped lessen the probability.

The interview data from this study revealed a very consistent and troubling pattern. Those children who, early on, struggled with reading began to hate reading. In this study I thought I was cleverly ascertaining motivation for reading by asking children such questions as, "Which would you rather do: clean your room or read a book?" This question did elicit the most quoted statement from this study. In response to this question, Javier volunteered, "I'd rather scrub the mold around the bathtub than read." The children who had struggled for 4 years with reading, however, were just as likely to respond to a direct question about their feelings toward reading with "I hate it."

Later Research

The finding of a .88 probability of not overcoming a poor start in reading, as well as children's subsequent development of hatred toward reading, had a big impact on me. In all my work after the 4-year longitudinal study I would engage in research with a direct focus on either intervention or classroom instruction.

I have been involved in several studies of different forms of early reading intervention that employ one-on-one tutoring. This form of tutoring builds on a common early method of learning: apprenticeship. Although tutoring is a common staple in upper-class families, it was not commonly available to the middle class until commercial enterprises such as Sylvan Learning made it financially possible. For lower-income families, however, it has remained out of reach unless the school system has made it available, as with Reading Recovery. It has not been financially possible, however, for school systems to provide

enough one-on-one tutors for the number of children who could use a boost in developing reading and writing.

I had a unique opportunity when I was at the University of Texas at Austin to work with two ends of what Stanovich (1986) labeled the "Matthew effect"—adults who were poor readers and children on their way to becoming poor readers. Members of the men's athletic department were always concerned about students who were admitted to the university based on their athletic abilities rather than on their academic skills. Indeed, many star players, particularly in basketball, had come from some of our nation's poorest urban communities. In line with what we know about the relation between socioeconomic status (SES) and achievement, these now-adult students were often such poor readers and writers that college-level work proved extraordinarily difficult. The athletic department had tried to ameliorate this situation. They had required a year-long study-skills–reading–writing course for those students who scored poorly on entrance on reading and vocabulary assessments. The class was not very popular, however, so the department readily agreed to an experiment. Half their poor readers would enroll in the current year-long course; half would enroll in a new year-long course with me. In my group the students would tutor a first- or second-grade child twice a week for 45 minutes each time.

The tutored children attended an all-minority school (largely African American with a smaller Latino population) in one of the poorest areas in the city. In addition to tutoring, the university students attended a weekly night course for 2½ hours and did 4 hours of outside reading per week. They read in books selected by a committee of student athletes; these ranged from novels to books on sports to biographies of growing up poor and African American. We communicated about their reading in a written journal. Also, in class they shared their thoughts on the outside reading, we talked about tutoring, and we spent part of the class on our own version of "writing workshop." In the workshop the tutors wrote books for the children they tutored. These books often starred their children as the central characters and were very popular with the children. We "published" the books; hence the books needed to be edited before being bound, and this created a time to work on the mechanics of writing.

Embedded in the year-long course with tutoring were two experiments. First, pre–post comparisons were made in reading, vocabulary,

and study skills between the university students in the tutoring course and those in the regular course; these comparisons showed significance in favor of the tutors in each area. Second, we compared the reading growth of tutored versus nontutored first- and second-grade children. Our nontutored children were not left without attention: Each had weekly visits from a student athlete who served as a mentor but who otherwise did not tutor the children. This "control" group did not perform as well in reading development as did those children who received tutoring.

This was an early attempt to develop a tutoring program, and I did not know what the most effective activities for the children would be. The tutors chose among seven tutoring activities for each 45-minute session: (1) reading children's literature; (2) hearing word sounds (phonemic awareness); (3) learning to write and recognize the alphabet; (4) working on phonics; (5) writing favorite words in a journal, then dictating sentences about the words to the tutor and drawing pictures about the words; (6) writing stories, postcards, and texts; and (7) making buildup readers. Buildup readers (Guszak, 1985) slowly introduce words from the basal readers used in the classrooms and words that exemplify taught phonics patterns. Five buildup readers were created, corresponding to the five levels of first-grade basal readers used in the classrooms. The length of the buildup readers ranged from 25 to 100 pages. On the first page of the first buildup reader, the word *run* appeared 30 times, with a few blanks to write in what was running. The child and tutor could jointly decide what to write in the blank—for example, what they wanted to have "run," be it a Ninja Turtle, themselves, or a pet. Buildup readers allow slow and repeated introduction to words throughout the levels of readers (though page 1 was by far the most dramatic, with only one word). Buildup readers formed a solid link to the words the children were expected to read in their classroom basal textbooks and in the tutoring sessions.

Although overall the tutored children outperformed the control group, there was considerable variation in effectiveness among tutoring dyads. So, in addition to the two central experiments, I examined the effectiveness of both time spent in the seven tutoring activities and the quality of those activities on child learning.

Time spent on the journal was one of the activities that were least predictive of reading growth. In fact, it had a negative correlation with growth, probably because tutoring time became drawing time. The most

predictive activities were time spent on phonics and on the buildup readers. (In my current tutoring program at Stanford, the tutors use Beck, 2006, to guide the phonics portion of their lesson.) In the dyads that showed the most growth, tutors engaged in scaffolded instruction and modeling of what they were trying to teach much more than tutors in less successful dyads did.

In this study, there was no observed difference in the amount of bonding and affection shown between the most successful and least successful tutoring dyads. Simply put, they all exhibited these outcomes. There is no doubt a special circumstance involved in these dyads, with a largely African American group of males tutoring a largely African American group of children. Not only did almost all the tutors and children share a history of being financially poor, but also the tutors had all struggled with literacy, just as did those they tutored. The tutors frequently wrote in their journals about how they identified with the children and how motivated they were to help them. I believe these overlaps created an especially powerful bond between these adults and their tutees.

In the time since this initial tutoring study, however, I have been involved in other tutoring programs. I would say that, in general, tutors become quite attached and committed to the children being tutored—whether the tutors are community or university student volunteers or are tutoring as part of an official class. There is something special about the one-to-one experience itself that breeds a powerful bond.

Although tutoring is helpful, the heart of literacy instruction lies in the classroom. The two-pronged attempt to try to help children get off to a good start in reading led me from tutoring to studies of classroom instruction. I discuss two studies here because they build off each other. The first was a year-long study in which we closely examined four first-grade classrooms (Juel & Minden-Cupp, 2000). A central issue of this study was how to capture what was going on in classrooms in terms of word recognition instruction (e.g., phonics instruction done in small groups and focusing on phonograms). We were interested in delineating the instructional practices that seemed to best foster learning to read words for particular profiles of children (e.g., children with differential literacy skill). The second study was a larger longitudinal study of preschool through first grade, with 13 classrooms in kindergarten and 13 in first grade. This study built on the observational system developed in the first study.

A real obstacle, it seems to me, in doing research on classroom instruction is that we have no agreed-on observation instruments. This may be too much to ask, not only because every researcher wants to look at different things but also because we do not have agreed-on norms for what or how literacy instruction should go. So the first longitudinal study I describe was partly dedicated to developing a classroom observation system that I could use in the second study.

The grain size of the observation instrument, as well as the categories on it, determines what researchers find. In a large grain size, we might look at how often phonics is taught. In a smaller grain size, we might parse that out into types of phonics (e.g., onset–rime, letter-by-letter) and the instructional context (e.g., in small groups, using letter cards, reading decodable text, sounding out words as they are written), among other possibilities. A still finer grain analysis might capture what a specific child was told to do (e.g., "Sound it out, Nora") or whether a child was even directly spoken to during the instruction.

One problem with classroom research is deciding which grain size to capture. Sometimes you don't know where to aim. Because it is hard to recapture live instruction and because even video must be analyzed at some level, choosing what to look at is critical; the grain size of what is recorded will limit findings to that grain size. At first doing everything at a fine grain size might seem the most sure-footed, but such data can be unwieldy.

Another obstacle to classroom observation is that so much is going on at the same time in a classroom. Several reading groups or centers may be going at once; activities are frequently embedded in other activities (e.g., letter–sound instruction occurs in the context of writing a whole-class letter to another classroom). Of course, this is the reason that experimental work can be so helpful, as the researcher can control the factors of interest; but my interest lies in what goes on in real classrooms, and how to capture what goes on inside these rooms is very challenging.

We tried to create a flexible observation system that we could return to, as needed, to increase or decrease the grain size (Juel & Minden-Cupp, 2000). A classroom observer sat with a laptop creating a running narrative of everything possible, including children's names. We observed in each of four classrooms for a minimum of an hour every week throughout the school year—usually the entire 90-minute language arts block. These four classrooms each had three reading groups.

We observed each low group each week and the other two groups at least every 2 weeks. We also observed whole-class instruction when it occurred during language arts.

Two research assistants subsequently coded the laptop narratives until we achieved an interrater reliability over .95. We began by coding four things: (1) activities (e.g., a read-aloud by the teacher, phonics); (2) materials (e.g., trade books, picture sort); (3) strategies (e.g., analogy, sound and blend); and (4) units (e.g., phonogram, initial consonant, and word). We added distinctions under the headings as needed. We found that the focus of *activities* generally clumped into five major categories: reading, writing, oral language, letter sound, and whole word (Figure 2.2). We could move from these five broad-grained clumps (e.g., *writing*) to the finer grained analysis (e.g., the 37 listed categories, such as *morning message*), to an even finer grained analysis taken directly from the narratives (e.g., Nora was asked to write her cousin's name in the morning message).

In the four-classroom study we found the middle-grain analysis (i.e., the 37 categories) sufficient to characterize the four classrooms. In the next study I describe we had to go down to the fine-grained narrative analysis to understand what was happening. The point is that I don't think you always know ahead of time which grain size to use.

Let me start by discussing the first study of the four classrooms (Juel & Minden-Cupp, 2000). Here we examined the effects of different forms and contexts of classroom life on growth on various reading assessments given in September, December, and May. We also asked the children to read specific words introduced to their reading groups and to explain how they identified the words.

We found that instruction varied enormously in the four classrooms, even though they were within the same school. We also found a considerable interaction between the type of instruction and reading group. Children who entered first grade low in alphabet knowledge and were placed in the low reading group did exceedingly better if they were placed with the teacher who did the most phonics instruction. The phonics activities in this classroom were very hands-on, with 66% of them involving sorting word cards into categories based on orthographic patterns and 17% involving "writing for sound"—writing dictated words that contained target spelling patterns. After February, however, instruction in this classroom resembled that in the other two reading groups, with more of a focus on reading and little on phonics.

Instructional Activities						
	Reading	Writing	Oral Language	Letter–Sound	Word	Other
1.	Reading text	Writing text	Discussion on topic	Oral phonemic awareness	Sight-reading words	Conference with teacher
2.	Choral reading	Individual writing	Meaning of story	Writing for sound	Spelling words	Peer coaching
3.	Round robin reading	Morning message	Read-aloud by teacher	Letter–sound/ decoding phonics	Word wall	Text structures
4.	Pair reading	Grammar or punctuation	Vocabulary	Letter identification	Cutting sentences up into words	Book/print awareness
5.	Individual reading	Handwriting		Word families	Concept of word	Nonliteracy activity
6.	Rereading	Language Experience Activity writing		Rhyming		
7.	Free choice/ reading	Copying		Matching letter cards to words		
8.	Expressive reading					
9.	Learning/ reciting memorized poem					

FIGURE 2.2. Overview of coding. After Juel and Minden-Cupp (2000).

This classroom had the most differentiated initial instruction overall and the highest overall mean passage-reading score at the end of first grade, a whopping mean level of late second grade.

In comparison, in another classroom there was little phonics instruction for anyone; rather, there was a considerable amount of reading in Little Books and trade books, a fair amount of journal writing, and some word wall use. The low-reading-group children fared exceptionally poorly in this classroom. The children did not learn to read the words

they saw in the books, and they could not employ any useful strategies for decoding unknown words. On the other hand, those children who entered first grade at or above average in alphabet knowledge and letter–sound knowledge did exceptionally well in this classroom.

The need for differential instruction seemed clear. Yet the finding was based on only four classrooms. That led to the second study, following a larger group of children in different grades. I am just recently finishing the data analysis. True to form, it was a longitudinal study following the growth of literacy and language from the end of preschool through the end of first grade. In this study I wanted to examine the two main factors in determining children's reading achievement that have been investigated by researchers looking to improve children's long-term literacy outcomes: (1) the role of incoming characteristics and (2) the role of instruction. In particular, I was interested in how reading skill grew in response to different forms and amounts of instruction and different incoming literacy and language profiles of children.

Literacy and language profiles were created at three time points: preschool, kindergarten, and first grade. The idea behind these was that profiles of multiple skills and abilities would yield a richer glimpse of a child than would the typical research focus on one or two characteristics. The profiles were based on Konold, Juel, McKinnon, and Deffes (2003). I need to take a bit of time to discuss the development of these profiles.

Richard Woodcock graciously provided us with the cross-sectional norming sample data for the Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1997). These data came from 1,604 children ages 5–10 years. The WDRB includes six subtests of underlying types of knowledge thought to be involved in reading, as well as four subtests that assess reading achievement. The six underlying knowledge subtests include two measures of phonological skill: (1) Incomplete Words, in which a child hears a tape-recorded word that has one or more phonemes missing and has to identify the complete word; and (2) Sound Blending, in which an audiotape presents word parts (syllables and/or phonemes of words) and the child puts them together to form a word. Other underlying foundations for reading that are tested are: Oral Vocabulary, Listening Comprehension, Memory for Sentences, and Visual Matching. We used these six subtests to predict skill on the four reading achievement tests of Letter–Word Identification, Word Attack, Reading Vocabulary, and Passage Comprehension.

Cluster analysis yielded six profiles, three relatively flat: (1) 11% of 5- to 10-year-olds had flat profiles on all six subtest scores that hovered around a standard score of 80–85 with similarly below-average reading scores; (2) 25% of the children had flat profiles on all subtests, hovering in the mid- to low 90s, with slightly below-average reading scores; (3) 13% of children, with significantly more females, had subtest scores that were all high—above 110 and most about 120, with predictably high reading scores.

Three of the six profiles had distinctive spikes or dips in performance: (1) 17% of children had all standard scores around 100, except for a notable high Visual Matching score around 115; (2) 16% had all standard scores around 100, with particular strengths in phonological skills on Incomplete Words and Sound Blending; and (3) 15%, with a significantly larger number of boys, had average scores in Visual Matching and phonological skills but considerably higher scores in Memory for Sentences, Oral Vocabulary, and Listening Comprehension. It is perhaps telling that this last profile houses more males, for if an individual child had even more depressed scores in Visual Matching and in phonological skills, the profile would resemble those of children labeled dyslexic.

Not surprisingly, those children with flat profiles and high scores overall outperformed other clusters on all four reading measures, whereas children with flat but low scores scored below average on all four reading measures. The more interesting findings came by way of comparison between profiles defined by notable strengths. We found that at age 5 children with phonological strengths demonstrated statistically greater scores on Letter–Word Identification, Word Attack, and Reading Vocabulary than children with no secondary strengths or with secondary strengths in Visual Matching or Memory for Sentences. At age 6 children with both phonological processing strengths and visual matching strengths performed better on all four reading measures. Not until age 10 did those children with secondary strengths in Memory for Sentences, Oral Vocabulary, and Listening Comprehension show an advantage.

With that background, let me return to the second longitudinal study. We were interested in whether these clusters developed over time or in response to instruction. Our first mistake was to assume that we would find roughly these same percentages in the profiles in an overwhelmingly low-SES and minority population that did not equate

to the population that formed the WDRB norming sample. We followed children from preschool age through the end of first grade. (That is, we followed about half the children from preschool because about half our sample did not attend preschool.)

There were six preschools that fed into three elementary schools—schools that were close to one another and that drew from a similar population. About half our sample attended preschool ($n = 64$). The largest public preschool, as well as the community Head Start preschools, were restricted to children from the lowest SES strata. About half our sample did not attend preschool ($n = 78$). On entry into kindergarten, we found that our preschool and nonpreschool samples (children who, whether because of parental choice or SES restrictions, did not attend preschool) did not differ in performance on the Wide Range Achievement Test (WRAT; Wilkinson, 1993). Both groups were still learning the letters of the alphabet on entrance to kindergarten.

There was some difference between those who were and those who were not in preschool in terms of the profiles on the WDRB in kindergarten. But for both groups, the majority of children were in the flat profile, with the lowest scores across the board. The actual statistics are really daunting: 62% of our preschoolers were in the lowest profile toward the end of preschool compared with 11% of the WDRB norming sample. That percentage decreased to 56% in kindergarten and to 31% by first grade. This is positive movement by first grade, but it occurs mainly because the children move to the slightly below-average flat profile at which 47% of them now reside. (Of our sample that did not go to preschool, 41% are in the lowest profile in kindergarten, decreasing to 21% in first grade, and again moving mainly to the second below-average flat profile, at which we find 39% of them.)

Despite this ominous standing in the profiles, the 142 children who were in our sample in kindergarten and first grade did, on average, learn to decode, and their word recognition on the WDRB and the WRAT were average by the end of first grade. But their oral vocabulary and listening comprehension skills remained in the lowest profile range, which clearly does not bode well for them as they advance through the grades. The three schools worked hard to bring children to grade level in reading. Overall they were successful in word recognition. The mean WDRB Letter–Word Identification score in first grade was a standard score of 103, which is impressive given the underlying profiles. Overall, a lot was done in kindergarten and first grade to promote word

recognition, even though the amount of instruction varied considerably across classrooms. Oral Vocabulary, however, stayed at a standard score of 89 from preschool through first grade for those children who attended preschool (n = 64) and stayed at 91 from kindergarten through first grade for all others (n = 78). In other words, they needed attention to oral language, vocabulary, and knowledge along the lines identified by Beck, McKeown, and Kucan (2002).

Our second mistake was to assume that we would see profiles with secondary strengths emerge longitudinally and/or in response to specific instruction. We did see some growth in the profile with the rise in Visual Matching by first grade, approximating the population norm. The other secondary strength profiles failed to coalesce anywhere near the population norms. In other words, we found few children with a secondary strength in Memory for Sentences, Oral Vocabulary, and Listening Comprehension, nor children with secondary phonological strengths.

I want to go into more depth here describing the effects of instruction in kindergarten and first grade and why flexible grain size observation instruments are important. Again, we followed the kindergarten children through first grade, and they came from all the kindergarten and first-grade classrooms in three public elementary schools. The three schools drew from the same area of low-income and public housing projects, small duplexes, and apartments in a city in the Southeast. Of the 142 children, 68% qualified for free lunch, 2% for reduced fare; 68% were African American, 29% white, 51% male, and 49% female. Across the three schools there were 13 kindergarten and 13 first-grade classrooms. We administered the WDRB toward the end of kindergarten and first grade and the WRAT at the beginning and end of kindergarten and first grade.

We made observations in each of the 13 classrooms in kindergarten and first grade at least once a month, making sure that every child in the study was observed. Observers typed a running account of classroom instruction during language arts. These were later coded, with interrater reliability of .97, as per the method in Juel and Minden-Cupp (2000).

The most common instructional setting in kindergarten, used in eight classrooms, was to have the class divided into three heterogeneous groups for language arts. The same lessons, readings, and read-alouds were used for each group. The reasons given for having three

groups were that kindergarten children attended better in small groups and that the teacher could be more aware of each child in the group. Children not in the group were usually in centers or at special programs in the school (e.g., music, physical education). Five classrooms largely used a whole-class structure for instruction.

In the heterogeneous small-grouped kindergarten classrooms, the time spent on various activities in the groups was virtually the same, although considerable variation existed among classrooms (as much within a school as between schools) as to the amount of the instruction devoted to particular activities. Following are some ranges in the proportion of the language arts period devoted to different activities in these heterogeneous groups between classrooms: a 12–42% range in language arts activities devoted to phonics, a 0–26% range in attention to isolated words (e.g., on a word wall or spelling), a 14–46% range in activities with potential to foster oral language (e.g., read-alouds by teacher, explicit vocabulary work), a range of 4–38% in reading (i.e., in which text reading involved the children in reading, such as in choral reading), and a range of 0–26% in writing activities (i.e., of text longer than a single word).

In the five kindergartens in which most instruction was done with the whole class, a similar diversity of activities existed, ranging, for example, from a high of 36% to a low of 14% in letter–sound activities. Two of these five classrooms occasionally broke down from the whole-class structure into small ability-group instruction with low, middle, and high groups. The instruction between these two classrooms, however, was quite different. The low group in one classroom received 75% letter–sound activities, whereas the low group in the other classroom received only 26%. But these low groups reflected the overall proclivity of the teachers. The middle and high groups in the 75% phonics classroom received, respectively, 65 and 52% phonics activities, whereas the middle and high groups in the other classroom received, respectively, 26 and 0% phonics activities. These two classrooms were in the same school.

We fully expected that these differences in kindergarten experiences, particularly with phonics, would make a big difference in children's growth on, say, the WRAT from the beginning to the end of kindergarten. This was our third mistake. (Well, not entirely.) Letter–sound instruction had a positive impact on those children who entered kindergarten with little letter recognition, but it actually seemed to have a neg-

ative impact on those children who entered with some degree of facility with letter sounds. Perhaps this should not have been such a surprise, as it echoes the finding of our previous work (Juel & Minden-Cupp, 2000). An HLM analysis in the current study showed that kindergarten letter–sound instruction had a positive reach into first grade for those with the humblest beginnings in alphabetic knowledge on entrance to kindergarten: The more letter–sound instruction these children had in kindergarten, the faster their growth in word recognition in first grade was. For the more advanced kindergarten children, however, more kindergarten letter–sound instruction was associated with slower growth in kindergarten (compared with their comparably endowed peers who received little such instruction but instead engaged in more reading and similar activities).

That is the conundrum for early literacy: how to balance phonics with other activities and especially how to manage differential instruction. Looking even more carefully at the observational data, we first wanted to understand why there was one particularly successful kindergarten for all the children. This was a kindergarten with the common three-heterogeneous-groups structure. In terms of sheer division of the group instruction, the class was low average in terms of percentage letter–sound instruction (22%), relatively high in word instruction (26%), and one of the highest in involving children in reading (33%). We never observed text writing in this class, and read-alouds and oral language were low average (17%). Perhaps this balance worked for these children in terms of their word recognition growth. On average, children who came in below the mean on the WRAT grew 22 points in their WRAT standard scores by the end of the year, whereas children who entered kindergarten with above-average letter–sound knowledge grew 5 points. We suspected that the division of activities was not the only positive driving force. And here we were glad that we had the small-grained level of the observation narratives to give us a hypothesis of what might be behind at least some of the success.

As we reread the classroom narratives of this front-runner kindergarten class, what stood out was the sheer number of times that *every* child was called on in each activity. In the course of direct group instruction, which lasted about 30 minutes, each child was called on an average of 10 times. We could tell this because we had recorded every child's name and response when he or she was called on. We then went back and recorded the number of times every child in the study was

called on in language arts throughout kindergarten and first grade. It was clear that small-group instruction (whether heterogeneous or homogeneous in constituency) included calling both on more and on varied children than did whole-class instruction.

But how important was being called on? In a regression analysis predicting end-of-year kindergarten WRAT, two variables swamped the impact of all other skill and instructional variables (R^2 = .71): beta number of times called on = .542*** and beta raw score for WRAT at beginning of kindergarten = .976***.

We found a similar situation in first grade. Being called on mattered a lot, as did how the children entered first grade on the WRAT—this last one a predictable finding. In addition, development of Word Attack, as measured on the WDRB, and Sound Blending significantly influenced word growth. In two first-grade classrooms, the teacher had also been the children's kindergarten teacher. These two first grades demonstrated a high number of mean times called on, exceeding what the classrooms were like in kindergarten. Not surprisingly, these two classrooms, together with a first-grade class with a similarly high average number of "called on" children, topped the WRAT growth charts.

It is interesting that the teachers who remained with their students in first grade had not had distinctive classroom results in kindergarten nor more than average times-called-on scores. They distinguished themselves only in first grade. Certainly the two teachers knew the children well by first grade, and this may have contributed to the increased calling on *all* children. From the child's point of view, the consistency afforded by having the same adult as a teacher in two grades was no doubt important, too.

Overall, in 2 years of instruction—kindergarten through first grade—the factors that most predicted the WRAT at the end of first grade in regression (R^2 = .71) were raw score on WRAT at the beginning of kindergarten (beta = .31***), total times called on during the 2 years (beta = .21***), first-grade score on WDRB Word Attack (reading pseudowords; beta = .53***), and WDRB Sound Blending (beta = .17*).

I would be remiss and deny my roots if I didn't say that, at a macro level, in first grade the simple view of reading held sway. In a regression predicting the standard score on the Reading Comprehension cluster on the WDRB (R^2 = .73), WDRB letter–word recognition clocked in with significant beta at .76*** and the WDRB oral comprehension cluster at

.16**. However, at the top of my list of advice to teachers right now is simply to call on *every* child, and *often*.

CURRENT AND FUTURE DIRECTIONS

I doubt that the desire to follow people longitudinally will ever leave me, nor the tendency to collect almost too much data on them. (I am still analyzing data from the above-mentioned longitudinal study.) I am still passionate about observing interactions and learning in classrooms. I am still passionate about the promise of early intervention and tutoring and am currently conducting a study in that area. My new interest is a focus on teacher development. Together with my colleague Aki Murata I am studying teacher development from preservice to inservice years. We are involved in a longitudinal study, of course, comparing development in literacy and math. As I get older, there may not be any lengthy longitudinal studies ahead of me. But to *really* understand growth in reading skill, other researchers still have many more studies to do.

REFERENCES

Beck, I. L. (2006). *Making sense of phonics: The hows and whys.* New York: Guilford Press.

Beck, I., & Juel, C. (1992). The role of decoding in learning to read. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (2nd ed., pp. 101–123). Newark, DE: International Reading Association.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction.* New York: Guilford Press.

Byrne, B. (2005). Theories of learning to read. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 104–119). London: Blackwell.

Ehri, L. (1992). *Reconceptualizing the development of sight word reading and its relationship to recoding.* In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20,* 163–179.

Gough, P. B. (1971). One second of reading. In J. Kavanaugh & I. Mattingly (Eds.), *Language by ear and by eye* (pp. 331–358). Cambridge, MA: MIT Press.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Guszak, F. J. (1985). *Diagnostic reading instruction in the elementary school* (3rd ed.). New York: Harper & Row.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127–160.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Juel, C. (1994). *Learning to read and write in one elementary school.* New York: Springer-Verlag.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243–255.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35,* 458–492.

Juel, C., & Roper/Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly, 20,* 134–152.

Konold, T. R., Juel, C., McKinnon, M., & Deffes, R. (2003). A multivariate model of early reading acquisition. *Applied Psycholinguistics, 24,* 89–112.

Plaut, D. (2005). Connectionist models. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 24–38). London: Blackwell.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of reading. *Reading Research Quarterly, 21,* 360–407.

Wilkinson, G. S. (1993). *Wide Range Achievement Test—3.* San Antonio, TX: Harcourt.

Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery.* Itasca, IL: Riverside.