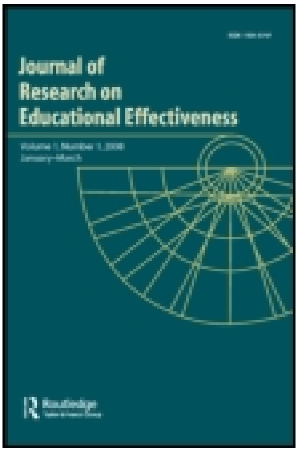
This article was downloaded by: [Joanne F. Carlisle]
On: 29 September 2011, At: 02:35
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK




Journal of Research on Educational Effectiveness
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uree20

# Teachers' Knowledge About Early Reading: Effects on Students' Gains in Reading Achievement

Joanne F. Carlisle [a], Ben Kelcey [b], Brian Rowan [a] & Geoffrey Phelps [c]

[a] University of Michigan, Ann Arbor, Michigan, USA
[b] Wayne State University, Detroit, Michigan, USA
[c] Educational Testing Service, Princeton, New Jersey, USA

Available online: 28 Sep 2011

To cite this article: Joanne F. Carlisle, Ben Kelcey, Brian Rowan & Geoffrey Phelps (2011): Teachers' Knowledge About Early Reading: Effects on Students' Gains in Reading Achievement, Journal of Research on Educational Effectiveness, 4:4, 289-321

To link to this article: http://dx.doi.org/10.1080/19345747.2010.539297

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

*Journal of Research on Educational Effectiveness*, 4: 289–321, 2011
Copyright © Taylor & Francis Group, LLC
ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2010.539297

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

# Teachers' Knowledge About Early Reading: Effects on Students' Gains in Reading Achievement

**Joanne F. Carlisle**
University of Michigan, Ann Arbor, Michigan, USA

**Ben Kelcey**
Wayne State University, Detroit, Michigan, USA

**Brian Rowan**
University of Michigan, Ann Arbor, Michigan, USA

**Geoffrey Phelps**
Educational Testing Service, Princeton, New Jersey, USA

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

**Abstract:** This study developed a new survey of teachers' knowledge about early reading and examined the effects of teachers' knowledge on students' reading achievement in Grades 1 to 3 in a large sample of Michigan schools. Using statistical models that controlled for teachers' personal and professional characteristics, students' prior reading achievement, and the clustering of high-knowledge teachers in schools and school districts with particular demographic composition, we found that the effects of teachers' knowledge about early reading on students' reading achievement were small. In 1st grade, students in classrooms headed by higher knowledge teachers performed better on year-end tests of reading comprehension but not word analysis. In 2nd and 3rd grades, the effects of teachers' knowledge on either measure of students' reading achievement were not statistically significant. Although the study suggests new forms of statistical analysis that might produce better estimates of the effects of teachers' knowledge on students' reading achievement, further research is needed to improve the conceptual and psychometric properties of measures of teachers' knowledge of reading and to investigate the relation of their knowledge and their instructional practices.

**Keywords:** Teachers' knowledge, early reading, student achievement

Research shows that after controlling for differences in students' previous learning and home background, student achievement varies widely from classroom to classroom at the same grade level within a school (e.g., Scheerens & Bosker, 1997). Mounting evidence suggests that at least part of this variation in student achievement results from stable "teacher effects," commonly defined as the fixed or random effects of specific teachers on their students' achievement gains across several years of observation (Nye, Konstantopoulos, &

Address correspondence to Joanne F. Carlisle, University of Michigan, School of Education, 610 East University Avenue, Ann Arbor, MI 48109-1259, USA. E-mail: jfcarl@umich.edu

Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti, & Miller, 2002; Sanders & Horn, 1994). Of current interest to education researchers is the extent to which these teacher effects on student achievement arise because of variation in teachers' pedagogical and content knowledge in the subject area they are teaching (e.g., Hill, Ball, & Schilling, 2008). This question is of particular importance because research shows that other indices of teachers' professional knowledge (e.g., degree attainment, certification status) are only weakly related to student achievement (e.g., Croninger, Rice, Rathbun, & Nishio, 2003; Wayne & Youngs, 2003).

The study reported here contributes to this line of work by discussing a new measure of teachers' knowledge about early reading and by reporting on an empirical study that used this measure to examine the effects of teachers' knowledge on students' reading achievement in about 900 first- through third-grade classrooms in Michigan. As we discuss in greater detail next, the study was designed to resolve a number of uncertainties arising from previous research on the effects of teachers' knowledge on students' early grades reading achievement. As we shall see, these uncertainties revolve around how to measure teachers' knowledge for teaching early grades reading and how to estimate the effects of such knowledge on students' reading achievement in light of various confounding factors in the matching of students to teachers within and between schools.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

## THE PROBLEM

For more than a decade, researchers have argued that to be effective, early grades reading teachers need a relatively high level of knowledge about "the linguistic foundations" of early reading (Moats, 2009b). Moats (1994, 1999) developed an early and influential approach to measuring teachers' knowledge in this domain known as the Informal Survey of Linguistic Knowledge. This survey included items designed to measure teachers' content knowledge about the relations between the spoken and written aspects of language; about the sound structure of words; and about related topics in grammar, morphology, and orthography. A decade after Moats (1994) introduced this measure, many of the items from her original survey continue to be used in studies of teachers' knowledge of early reading—and for good reason. As Piasta, Connor, Fishman, and Morrison (2009) noted in justifying their use of items from Moats's survey, "Current theories of reading emphasize the necessity of the alphabetic principle to link phonological, orthographic, and semantic knowledge, particularly in the beginning stages of literacy" (p. 225). Thus, they reasoned that teachers' knowledge of the alphabetic principle and of mappings between language and print was essential for effective early reading instruction.

Researchers have used items from Moats's (1994) survey to address several interrelated research questions about the teaching of early grades reading. Some studies have examined the extent to which teachers actually know about the linguistic foundations of early reading; others have investigated whether specific professional development programs can increase teachers' knowledge in this domain; still others have asked whether increasing teachers' linguistic knowledge leads to more emphasis on explicit instruction in phonemic awareness, phonics, or other code-related aspects of reading; and a few studies have examined whether teachers with greater knowledge in this area have a more positive impact on students' reading achievement than do teachers with less knowledge in this area.

The findings from this body of research address (but leave open) a number of important questions about the nature of teachers' knowledge about teaching early grades reading, about how to measure this construct, and about whether teachers' knowledge in this domain is related to teaching effectiveness, as measured by gains in students' reading achievement. For example, several studies have shown that the average teacher of early-grades reading lacks strong knowledge about the linguistic foundations of reading (e.g., Bos, Mather, Dickson, Podhajski, & Chard, 2001; McCutchen, Abbott, et al., 2002; Moats, 1994). In addition, a growing body of evidence has shown that teachers can be taught linguistic knowledge through programs of professional development (e.g., Bos, Mather, Narr, & Babur, 1999; Foorman & Moats, 2004; Garet et al., 2008; McCutchen, Abbott, et al., 2002; Spear-Swerling & Brucker, 2003, 2004). Additional evidence suggests that professional development can affect teaching practice, with research tending to show that teachers who participate in professional development aimed at increasing knowledge about the linguistic foundations of reading also provide students with more explicit instruction in phonemic awareness, phonics, and other code-related areas of reading (Bos et al., 1999; McCutchen, Abbott, et al., 2002; Garet et al., 2008).

What is less clear from research is the extent to which teachers' knowledge about the linguistic foundations of reading has an effect on students' reading achievement in the early grades. In fact, the accumulated evidence, across different studies, suggests that the effects on students' reading achievement of teachers' knowledge in this area might be limited to certain domains of reading performance (e.g., Moats, 2009b). Positive evidence of the effects of teachers' knowledge on students' reading achievement can be inferred from a study by Bos et al. (1999), which found that students of teachers who received professional development aimed at increasing their knowledge of the linguistic foundations of reading showed greater achievement gains in some of these areas (e.g., phonemic awareness) than did students whose teachers did not receive this same professional development. It is important to recognize, however, that this study estimated the effect of teachers' participation in a professional development program on students' reading achievement, not the effect of their knowledge about reading. The same is true of other studies. For example, Garet et al. (2008) conducted a large randomized field trial of professional development program emphasizing (in part) the linguistic foundations of reading. These researchers found that teachers who participated in the program scored higher on a test of their "code-related" knowledge of reading than did teachers who did not participate in the program, although the students of participating teachers did *not* show statistically greater gains in reading achievement compared to students of teachers who did not attend the program. This study also did not examine the effects of teachers' knowledge about reading on their students' outcomes.

In addition, studies of the relationship between teachers' knowledge of the linguistic foundations of reading and students' achievement have shown inconsistent results. For example, in contrast to Bos et al. (1999), Spear-Swerling and Brucker (2004) found that students who were tutored by teachers with higher knowledge of the linguistic foundations of reading achieved higher word reading scores than did students tutored by teachers with lower scores, but this effect did not occur on students' test scores in the areas of letter-sound correspondence, reading of irregular words, or spelling. In another study, McCutchen, Harry, and colleagues (2002) reported positive correlations between measures of teachers' linguistic knowledge and kindergarteners' word-reading achievement, but these researchers did *not* find a relationship between teachers' linguistic knowledge and first and second graders' achievement in the domains of vocabulary, reading comprehension, spelling, or writing fluency.

In summary, the results of these various studies present a quandary. Research suggests that early-grades reading teachers have limited knowledge of the linguistic foundations of reading and that professional development can increase teachers' knowledge in this domain.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

The open question, however, is whether increasing teachers' knowledge in this domain will improve students' reading achievement. The pattern of uneven (and modest) effects previously described led Foorman and Moats (2004) to suggest the need for further research into the measurement and effects of teachers' knowledge for early grades reading—an area of research that we also see as important. The study reported here was designed to develop a new measure of teachers' knowledge of early-grades reading and to develop an empirical approach to estimating the causal effects of teachers' knowledge on students' reading achievement using nonexperimental data. In the sections that follow, we explain our approach to investigating these issues.

## APPROACH TO MEASUREMENT ISSUES

We concluded from our review of research on teachers' knowledge about early-grades reading that a somewhat different approach to measuring teachers' knowledge was warranted. As discussed next, this entailed addressing three interrelated problems: (a) the domains of knowledge to be assessed in measures of teachers' knowledge of early-grades reading, (b) the types of knowledge to be assessed within these domains, and (c) the resulting psychometric properties of any measures we developed.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

### Domains and Types of Teacher Knowledge

We begin by discussing the domains and types of knowledge to be assessed in our study. Our review of the literature suggested that existing measures of teachers' knowledge about early reading had two main properties. First, most reported measures focused on just one of several domains of specialized knowledge that teachers might need in order to teach early-grades reading effectively, namely, teachers' knowledge of the linguistic foundations of early reading. In our view, it makes sense to assume linguistic knowledge is an important component of teachers' knowledge for teaching reading in the early grades. However, knowledge in this particular domain would seem to be relevant mainly to code-related instruction. There is good reason to focus on teachers' knowledge beyond this limited domain (Snow, Burns, & Griffin, 1998). For example, most balanced reading programs in the early grades recognize the need to build students' oral language, not only to develop phonemic awareness and decoding skills but also to promote vocabulary, fluency in word recognition and text processing, and reading comprehension (Pressley et al., 2001; Snow, Griffin, & Burns, 2005). For this reason, we would argue that reading researchers need to expand their measures of teachers' knowledge for reading instruction beyond an exclusive focus on linguistic foundations.[1]

A second problem with most current approaches to measuring reading teachers' knowledge is the focus on a particular type of knowledge: teachers' *content* knowledge, defined here as knowledge of a particular *academic* body of work—in this case, linguistics.

[1]Not all measures of teachers' knowledge for teaching early-grades reading reviewed in this article have focused *only* on the linguistic foundations of reading, although this is true of most of the measures used in the publications cited in our literature review. In particular, it is worth noting that the measure of teachers' knowledge developed by Garet et al. (2008) was carefully (and more or less evenly) balanced across the five areas of reading discussed in Snow et al. (1998). The measure used by Foorman and Moats (2004) was also reported to focus on all five of these areas.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

McCutchen, Harry, et al. (2002), for example, described the academic nature of items in Moats's (1994) original Informal Survey of Linguistic Knowledge when they noted that "a perfect score on the [Moats] survey is difficult to achieve without considerable linguistic training" (p. 214). An important question for reading research, however, is whether *academic* knowledge of this sort is the only form of knowledge needed to teach early-grades reading effectively. To be sure, teachers need content knowledge to teach effectively (Shulman, 1986), but the possession of academic knowledge does not assure that teachers will be effective in teaching their assigned subjects. In reading, for example, teachers might be able answer a number of difficult questions about English phonology correctly and still not know how to effectively teach children who are having real problems grasping the concept of phonemes in words. As Shulman (1986, 1987) pointed out in his seminal discussions of pedagogical content knowledge, more than content knowledge is needed to teach effectively. Snow et al. (2005) referred to this knowledge as "usable" knowledge, knowledge that is "embedded in practice" (p. 11). Hill, Rowan, and Ball (2005) referred to such knowledge as a "specialized" form of content knowledge—that is, a deep understanding of both disciplinary knowledge and ways that such knowledge can be represented to foster student learning. From this perspective, we argue that measures of teachers' knowledge in any academic domain should assess not only teachers' academic knowledge but also their understanding of how that knowledge might be used effectively in practice.

## An Alternative Approach to Measurement

The work reported in this article builds on two additional insights from research conducted by others. One comes from prior research showing that the knowledge of early-grades reading teachers can be measured along two primary dimensions—knowledge relevant to the teaching of word reading and knowledge relevant to the teaching of reading comprehension. In particular, research conducted by Phelps and Schilling (2004) and by Garet et al. (2008) provided evidence that knowledge in the domains of word reading and reading comprehension define two measurable domains of teachers' knowledge about reading. Based on this insight, the work reported here aimed at developing a measure of reading teachers' knowledge that included questions focused on both word reading *and* reading comprehension.[2]

Another key insight comes from efforts to measure teachers' knowledge in fields of research other than reading. In particular, the line of work on mathematics teachers' knowledge conducted by Heather Hill and colleagues (e.g., Hill & Ball, 2004; Hill, Schilling, & Ball, 2004) demonstrates two important points that we attempted to build on in the work reported here. Their work shows that in addition to measuring teachers' academic content knowledge, it is possible to measure teachers' knowledge of pedagogy and student learning in specific areas of the school curriculum. Further, it is possible to develop assessment items that situate teachers' knowledge for teaching in instructional contexts. Thus, in our study, rather than asking teachers how many phonemes are in certain words (as is done in many studies of teachers' reading knowledge), we developed items that asked teachers

[2]Other researchers have proposed measuring additional aspects of teachers' knowledge. For example, Palincsar and Duke (2004) argued that knowledge of texts and genres is essential for teachers of reading, and Cunningham, Perry, Stanovich, and Stanovich (2004) included not only a measure of phonology and phonics but also a measure of teachers' knowledge of children's literature, although the latter did not account for students' reading performance.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

to determine whether a student's spelling errors indicate difficulty identifying sounds in words. Items of this sort situate teachers' content knowledge in instructional contexts.

Finally, our current approach grew out of our previous research on reading teachers' knowledge and reflects an evolution in our measurement efforts. In an initial study of the connection between reading teachers' knowledge for teaching and student achievement (Carlisle et al., 2009), we developed a measure to assess teacher's knowledge of the linguistic foundations of reading as disseminated at professional development seminars on Moats's (2003) *Language Essentials for Teachers of Reading and Spelling* that were attended by teachers in Reading First schools in Michigan. That measure, which we called Language and Reading Concepts, included twice as many items focused on phonology, phonics, and grammar as it did items focused on reading comprehension and vocabulary. Moreover, like many teacher assessments in the field of reading, the items in this initial measure assessed academic content knowledge (e.g., "Which of the following words has a prefix?") without situating that knowledge in instructional contexts. In this study, we estimated the effect of teachers' knowledge of reading on students' reading achievement, controlling for the sociodemographic characteristics of students in a classroom and for several characteristics of teachers' professional preparation for teaching (e.g., certification status, educational attainment). The student outcomes were the performances of first, second, and third graders on two subtests of the Iowa Tests of Basic Skills (ITBS): word analysis and reading comprehension. The results showed *no* statistically significant effects of teachers' knowledge measured in this way on students' covariate adjusted achievement in word analysis or reading comprehension.

In interpreting the results of this first study, we hypothesized that the lack of statistically significant effects of our measure on students' reading achievement might be attributable (in part) to the approach we had taken to measuring teachers' knowledge, in particular, the focus in our measure on assessing teachers' decontextualized, academic content knowledge and the overrepresentation of items assessing teachers' knowledge of the alphabetic code and aspects of teaching word reading. As a result, in a subsequent study, we developed a measure of teachers' knowledge that differed from this initial measure in three ways (Carlisle et al., 2008). First, we included items designed to situate teachers' knowledge in classroom practices. Second, the new measure had a better balance of items focused on word reading and comprehension. Finally, the measure was based not on the contents of a particular program of professional development but rather on experts' judgments of the knowledge that teachers needed to teach beginning reading effectively. Using propensity score matching (Rosenbaum & Rubin, 1983) to identify and contrast comparable classrooms and teachers, we estimated the effects of this new measure on classroom-to-classroom variation in students' reading achievement. The results of this second study indicated the presence of a small, positive effect on students' ITBS reading achievement scores in Grades 1 and 2 (with standardized regression coefficients of $b = .05$), but no effect of teachers' knowledge on this measure of students' reading achievement in Grade 3. This result suggested that the measure of teachers' knowledge that we developed emphasized knowledge likely to have relevance for teaching reading effectively in first and second grades but not in third grade.

## THE CURRENT STUDY

The present study was designed to address what we saw as two shortcomings in the studies that we (and others) have carried out. First, we again revised our measure of teachers' knowledge for early-grades reading so that items focused less on measuring teachers'

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

academic content knowledge and more on teachers' use of their content knowledge to make decisions about instruction or to analyze students' performance on reading/writing tasks. Second, because teachers with extensive knowledge about reading might not be distributed equally across schools, we also adjusted the propensity score methods we used to statistically control for the clustering of high-knowledge teachers in certain schools and to control for the potential influence that schools might have on students' academic achievement. We then used this analytic strategy to estimate the effect of teachers' reading knowledge on students' reading achievement.

The current study has two research questions: What is the reliability and dimensionality of the measure of teachers' knowledge that we developed (Teachers' Knowledge of Reading and Reading Practices, or TKRRP)? To what extent does teachers' knowledge about reading, as demonstrated on this measure, affect students' gains in reading achievement over a school year? Our first question focused on the psychometric characteristics of our newly developed test of teachers' knowledge. We saw this as a critical first step in our research, especially because so few prior studies examined the psychometric properties of the measures they used, a problem that could affect study outcomes. The second step involved developing an empirical strategy to estimate the effects of teachers' knowledge on students' reading achievement more directly than has been done in experimental studies of professional development programs while addressing the complex issues of causal inference that arise in nonexperimental studies.

With respect to this second problem, several issues are critical. The first is that in American schools, students who face the greatest challenges in learning to read (i.e., poor and minority students with lower levels of entry-level achievement) are also taught by the least qualified teachers (Darling-Hammond, 2004). In this situation, we can expect to find that teachers' knowledge is related to students' achievement simply because more knowledgeable teachers are clustered within schools that serve students who generally make larger gains in reading achievement. This clustering of particular types of teachers and students in particular types of schools motivated our use of a multilevel approach to propensity score matching. This approach is discussed in more detail next.

A second issue concerns how to estimate the effect of teachers' knowledge in light of findings that teachers' knowledge affects their instructional practices in ways that could improve students' reading achievement (e.g., Bos et al., 2001). Our approach to estimating a teacher knowledge effect in light of this endogenous process is to carefully match teachers on a large number of student, classroom, and school covariates known from previous research to affect students' reading achievement but *not* to match teachers in terms of their instruction.

A final issue arises because our analytic methods rely on observational (i.e., nonexperimental) data. In this situation, it can always be argued that our estimated effect of teachers' knowledge on students' achievement is subject to omitted variables bias. For this reason, we examined the robustness of our causal inferences about the effects of teachers' knowledge on students' achievement by conducting a sensitivity analysis. This analysis addressed the question of the extent to which our estimates of the effects of teachers' knowledge on students' achievement might be altered in light of any failure to include particular kinds of unmeasured variables in our statistical model.

## SAMPLE, DATA SOURCES, AND MEASURES

The current study examines these issues by studying a sample of teachers who worked in Reading First schools in Michigan during the 2006–2007 school year. Reading First

(Part B of No Child Left Behind Act of 2001) was specifically designed to improve the reading achievement of kindergarten through third-grade students in high-poverty schools with chronic underachievement in reading (U.S. Department of Education, 2002).

## Research Sample

The participants in this study were volunteers from the population of teachers working in Michigan's Reading First schools during the 2006–2007 school year.[3] About 72% of the Grade 1 through Grade 3 teachers in these schools volunteered to allow researchers to use survey results in research studies. Collectively, the 1,101 volunteer teachers taught in 138 schools and instructed 16,439 students. Of the teachers who agreed to participate, 297 first-grade, 275 second-grade, and 292 third-grade teachers had sufficient student achievement data to be included in the study. Although we were unable to conduct this study with the full population of Michigan Reading First teachers, we did have available data from both the population of Reading First teachers and the research sample. This allowed us to compare the characteristics of the two groups and determine the extent to which the volunteer sample differed from the larger population of teachers. On nearly all measures we used to assess differences across the groups, the two groups showed no statistically significant differences (see Tables 1 & 2). An exception, however, was that the research sample had a higher average score (of +.25 *SD*) on our measure of teachers' knowledge. Table 1 presents demographic information on the students taught by teachers included in the research sample and students taught by teachers in the larger population. Table 2 presents information about the teachers in the research sample and in this larger population.

## Sources of Data

Two types of student achievement data were used in this study: a classroom reading assessment that was used as a pretest measure of students' achievement and a standardized achievement test that was used as both a pre- and a posttest measure of achievement. We also included measures of students' sociodemographic characteristics in our statistical models. Data on students came from Michigan's Single Record Student Database (http://www.michigan.gov/cepi). Data on teachers included teacher scores on our measure of teachers' knowledge and data on teachers' professional and personal background. These data were collected from a survey instrument called "Teacher's Quest" administered three times a year to Reading First teachers in Michigan. Finally, school and district demographic and organizational data were gathered from the Michigan Department of Education website (http://www.michigan.gov/mde). These sources of data are described next.

[3]To qualify for Reading First funding in Michigan, districts had to meet eligibility requirements of low reading achievement (i.e., 40% or more of fourth-grade students scoring below the proficiency cut point on the state assessment, Michigan Evaluation of Academic Performance, Reading) for 2 of the preceding 3 years and low income (e.g., 1,000 or more students from families below the poverty line).

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

**Table 1.** Characteristics of students in the research sample and population

| | Grade 1 | | Grade 2 | | Grade 3 | |
|---|---|---|---|---|---|---|
| | Population[a] | Research Sample[b] | Population[c] | Research Sample[d] | Population[e] | Research Sample[f] |
| Disability | 0.09 | 0.10 | 0.11 | 0.11 | 0.13 | 0.13 |
| Limited English proficiency | 0.12 | 0.11 | 0.13 | 0.14 | 0.12 | 0.14 |
| Special education | 0.04 | 0.04 | 0.05 | 0.05 | 0.06 | 0.07 |
| Free or reduced lunch | 0.73 | 0.71 | 0.75 | 0.72 | 0.74 | 0.73 |
| Hispanic | 0.11 | 0.12 | 0.11 | 0.11 | 0.12 | 0.14 |
| White | 0.36 | 0.40 | 0.36 | 0.42 | 0.36 | 0.39 |
| Hawaiian | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| African American | 0.42 | 0.39 | 0.43 | 0.38 | 0.42 | 0.38 |
| Asian | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| American Indian | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 |
| DIBELS fall grade 1/2/3 NWF/ORF score | 22.56 | 23.73 | 36.90 | 38.66 | 62.16 | 63.13 |
| Male | 0.51 | 0.50 | 0.51 | 0.52 | 0.51 | 0.51 |
| Average age (in months) | 84.56 | 84.67 | 97.04 | 96.94 | 109.12 | 109.07 |
| ITBS–Reading comprehension | 149.58 | 149.85 | 166.59 | 167.81 | 179.08 | 179.79 |
| ITBS–Word analysis | 148.72 | 149.14 | 162.86 | 163.95 | 177.50 | 177.59 |

*Note*. Characteristics represent proportions with the exception of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores, age, and Iowa Tests of Basic Skills (ITBS) scores. NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency.

[a]$N = 9{,}187$. [b]$N = 5{,}720$. [c]$N = 8{,}904$. [d]$N = 5{,}231$. [e]$N = 8{,}719$. [f]$N = 5{,}488$.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011



**Table 2.** Selected characteristics of teachers in the research sample and population

| | Grade 1 | | Grade 2 | | Grade 3 | |
|---|---|---|---|---|---|---|
| Characteristic | Population[a] | Research Sample[b] | Population[c] | Research Sample[d] | Population[e] | Research Sample[f] |
| Teacher knowledge | 0.02 | 0.27 | 0.05 | 0.10 | 0.06 | 0.06 |
| White | 0.76 | 0.83 | 0.75 | 0.78 | 0.79 | 0.86 |
| African American | 0.16 | 0.10 | 0.17 | 0.15 | 0.16 | 0.11 |
| Hispanic | 0.04 | 0.07 | 0.04 | 0.04 | 0.02 | 0.01 |
| Bachelor's in elementary education | 0.69 | 0.69 | 0.68 | 0.71 | 0.65 | 0.68 |
| Bachelor's in early childhood education | 0.07 | 0.09 | 0.05 | 0.12 | 0.04 | 0.03 |
| Bachelor's in literacy education | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 |
| Bachelor's in special education | 0.15 | 0.15 | 0.16 | 0.14 | 0.20 | 0.11 |
| Master's in elementary education | 0.33 | 0.28 | 0.34 | 0.25 | 0.37 | 0.30 |
| Master's in early childhood education | 0.14 | 0.08 | 0.08 | 0.07 | 0.06 | 0.05 |
| Master's in literacy education | 0.16 | 0.14 | 0.15 | 0.12 | 0.14 | 0.11 |
| Master's in special education | 0.16 | 0.18 | 0.19 | 0.15 | 0.20 | 0.18 |
| Post-master's degree | 0.06 | 0.05 | 0.06 | 0.04 | 0.03 | 0.03 |
| Standard certification | 0.66 | 0.68 | 0.62 | 0.60 | 0.63 | 0.63 |
| Provisional/Temporary certification | 0.20 | 0.21 | 0.24 | 0.29 | 0.18 | 0.22 |
| Reading certification | 0.04 | 0.06 | 0.03 | 0.05 | 0.03 | 0.04 |
| Special education certification | 0.19 | 0.10 | 0.20 | 0.11 | 0.22 | 0.12 |
| High years teaching | 0.49 | 0.49 | 0.47 | 0.48 | 0.43 | 0.41 |
| Number of professional trainings | 3.55 | 3.74 | 3.44 | 3.62 | 3.23 | 3.28 |
| Teacher new to reading first in 2006–07 school year | 0.15 | 0.16 | 0.18 | 0.15 | 0.21 | 0.23 |
| Average fall NWF/ORF of classroom | 23.05 | 23.05 | 36.42 | 36.47 | 58.13 | 58.14 |

*Note*. Based on proportions except for trainings and knowledge. NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency.
[a]$n = 524$. [b]$n = 297$. [c]$n = 507$. [d]$n = 275$. [e]$n = 532$. [f]$n = 292$.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

### Measures of Students' Reading Achievement and Sociodemographic Characteristics

The student outcome measures for this study were developmental scale scores taken from students' spring 2007 performance on two subtests of the ITBS—the word analysis subtest and the reading comprehension subtest. The word analysis subtest asked students to identify and match sounds and spelling elements of words. The reading comprehension subtest asked students to select responses to basic reading comprehension questions that followed short passages. Test reliabilities (computed with Kuder-Richardson Formula 20) for each subtest for Grades 1, 2, and 3 were, respectively, word analysis: 0.85, 0.85, 0.85; reading comprehension: 0.91, 0.90, 0.91 (Hoover et al. 2003).

For each grade level, our statistical analysis adjusted these scores for prior reading achievement, using students' 2006 fall performance on one subtest of the Dynamic Indicators of Basic Early Literacy Skills, hereafter called DIBELS (http://dibels.uoregon.edu). For first grade, the pretest used for adjustment was student scores on fall performance on Nonsense Word Fluency, a subtest that asked students to decode two- or three-letter nonsense words on a printed page; credit was given for the number of letters correctly decoded in 1 min. For second- and third-grade students, the subtest used for adjustment was Oral Reading Fluency, which asked students to read aloud three passages; all three passages were scored for the number of words correctly read in 1 min, and the student's score was based on his or her median passage performance. In addition, for the second and third grades, we also included as a pretest measure of student's prior ITBS reading scores from the spring of 2006. ITBS could not be used as an adjustment in first-grade analyses because it was not administered to kindergartners.

Alternate form reliability for the DIBELS measures was reported in a document on the DIBELS website (Assessment Committee, 2002). For Nonsense Word Fluency, the median was 0.83 for first graders. For Oral Reading Fluency, the report gave a range of 0.91 to 0.96 for second graders reading a variety of different passages. Teachers were trained to administer DIBELS by the Reading First literacy coaches and the state's Reading First facilitators. Literacy coaches assisted in the administration of DIBELS and entered the DIBELS results into the web-based database maintained by the University of Oregon.

Prior to conducting the data analysis, sociodemographic information about students was linked to data on students' performance on the DIBELS and ITBS subtests. Demographic information on students included measures of age, gender, ethnic and racial background, status with regard to English language proficiency and disabilities, and eligibility for free or reduced-price lunch. Descriptive statistics on the test scores and demographic characteristics of students in the research sample are shown in Table 1.

### Measure of Teachers' Knowledge and Professional Background

To measure teachers' knowledge about early reading, we had teachers complete the TKRRP survey in the winter of 2007. The TKRRP was specifically designed to measure the knowledge about early reading that early elementary teachers in Grades 1 to 3 use as they teach children to read words and comprehend texts. The content of this test was developed in consultation with experts in the field of early-reading instruction who provided their views of the types and domains of knowledge teachers need in order to teach early reading well. Based on their advice, we selected for inclusion on the TKRRP items that focused on activities in oral language, reading, and writing that occur in teaching word reading (e.g., phonemic awareness, letter-sound relationships) and comprehension (e.g., morphology, text analysis, fluency). We also engaged in pilot testing the TKRRP to eliminate items that did not discriminate among teachers or show strong fit statistics. Questions on the TKRRP assessment were designed around scenarios that teachers might encounter when teaching reading in the early elementary years. It included 13 situations or scenarios, a number of which have multiple items for a total of 22 items. Appendix A shows the items on TKRRP.

The Teacher Information section of the survey provided information about participating teachers' personal and professional characteristics. These included race/ethnicity, undergraduate major, graduate major, attainment of a master's degree, type of certification currently held, and type and amount of professional trainings attended. Response options (e.g., master's degree) involved a simple yes or no response (represented through indicator variables) with the exception of professional trainings. The measure of professional trainings was a sum of the number of trainings the teacher indicated that he or she had completed; options included programs such as Reading Recovery and Orton Gillingham.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

### Measures of School and District Characteristics

School and district characteristics were constructed by aggregating student and teacher-level data. In addition, several measures were drawn from the Michigan Department of Education website (http://www.michigan.gov/mde), including the percentage of students that were male/female, an index based on the percentage of students in each racial/ethnic group (White/non-Hispanic, African American, Hispanic, Asian, American Indian, Hawaiian, other), and a proxy measure for the socioeconomic status of students at a school (i.e., the percentage of students eligible for free or reduced-price lunch).

### Missing Data

As a result of teacher and student mobility and/or absenteeism on the day an instrument was administered, approximately 10% of teachers and students at each grade level had missing data on one or more variables included in our statistical models. Rather than remove students or teachers with incomplete data, we used the computer program IVEWARE (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001) to multiply impute values, based on every measured variable. This produced five separate student- and classroom-level data sets, where each data set contained a different plausible value for any missing value for a particular case. We then conducted statistical analyses using each of these five data sets and averaged parameter estimates from these multiple analyses to arrive at our final estimates of the effects of teachers' knowledge on students' achievement. In the tables reported below, we adjusted all estimated effects for the increased uncertainty resulting from multiple imputation (for a discussion of the advantages and use of multiple imputation in data analysis in educational settings, see Peugh & Enders, 2004).

## STATISTICAL MODELS

We conducted three independent but parallel lines of data analysis for the study. Each analysis focused on estimating the effects of teachers' knowledge (as measured by TKRRP) on students' reading achievement at a single grade level. Moreover, for each grade-specific analysis, we estimated the effect of teachers' reading knowledge on students' achievement in two domains: reading comprehension and word analysis. These analyses also were conducted separately. In each of these latter analyses, we derived the estimands in two stages. In the first stage, we used various extensions of propensity score stratification (Rosenbaum & Rubin, 1983) to approximate an experiment. The goal of the propensity score stratification was to assure that classrooms whose teachers had varying levels of teacher knowledge were otherwise closely matched on a wide range of observed covariates. In the second stage of the analysis, we then controlled for the propensity strata in which classrooms were located and used hierarchical linear models to estimate the effects of teachers' knowledge on students' reading achievement. Hierarchical linear models were used in the analyses in order to adjust standard errors for estimates in light of the nesting of students within classrooms and the resulting lack of statistical independence among student test scores that potentially results from this nesting process.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

## Propensity Score Analysis

The propensity score analysis just discussed was intended to address an important problem in research on teachers' knowledge—the possible confounding of teacher knowledge with other related school, teacher, and student characteristics. If this confounding is not taken into account in our statistical models, estimates of the effects of teachers' knowledge on students' achievement could be biased. Because our study could not employ random assignment to address the possibility of confounding, we followed the common practice of attempting to remove confounding through development of a propensity score and through stratifying cases on this score. In essence, this propensity score stratification works to assure that we are estimating the effects of teachers' knowledge on students' achievement only within groups that are closely matched on a wide range of observed covariates.

In the current case, we conducted three separate, grade-specific multilevel propensity score analyses in which every available classroom, teacher, school, and district variable in our data set was used to model the propensity (or likelihood) that different types of teachers, working in different types of classrooms, located in different types of schools and districts, would have higher levels of teacher knowledge (as assessed by our TKRRP measure). The variables are listed in Appendix B. In each of these analyses, our statistical model for deriving a propensity score for a given classroom was a three-level, hierarchical linear model with random intercepts and slopes. In this approach, the propensity function for teacher $j$ in school $k$ in district $l$ was modeled as

$$\text{Level 1 (Teacher):} \quad TK_{jkl} = \beta_{0kl} + \sum_{p=1}^{P} \beta_{pkl} X_{pjkl} + \varepsilon_{jkl} \tag{1}$$

$$\text{Level 2 (School):} \quad \beta_{0kl} = \gamma_{00l} + \sum_{q=1}^{Q} \gamma_{0ql} W_{qkl} + r_{0kl}$$

$$\beta_{pkl} = \gamma_{p0l} + \sum_{q=1}^{Q} \gamma_{pql} W_{qkl} + r_{pkl} \tag{2}$$

$$\text{Level 3 (District):} \quad \gamma_{00l} = \pi_{000} + \sum_{n=1}^{N} \pi_n D_{nl} + u_{00l} \tag{3}$$

where $X$, $W$, $D$, $\beta$, $\gamma$, and $\pi$ are respective teacher, school, and district variables and coefficients. Further, $\varepsilon$, $r$, and $u$ are the appropriate random effects assumed to be from a multivariate normal distribution (Hong & Raudenbush, 2006).

These multilevel regressions produced a propensity function (or score) for each classroom in the study. This score is simply a scalar from which we can identify similar teachers on the basis of their likelihood to be a high-knowledge teacher (Imai & Van Dyk, 2004). Using the values of classrooms on this scalar, we stratified classrooms into five groups, based on quintiles of the resulting distribution, and then checked to see that, within strata, there was no association between teachers' TKRRP scores and each of the classroom, teacher, school, and district variables used to generate the score. Many of these variables, it should be noted, have been shown in previous research to be associated with student achievement and to influence the sorting of teachers into classroom assignments (e.g., Darling-Hammond, 2004; Kainz & Vernon-Feagans, 2007; Kieffer, 2010).

An important finding of the propensity score analysis was that, within all strata, there were no significant correlations among the variables used to generate the propensity score and the TKRRP scores of teachers heading a classroom, suggesting that simply by entering the strata location of a classroom as an independent variable in our analysis of the effects of teachers' knowledge on students' achievement, we can substantially reduce omitted variables bias. However, we cannot rule out the possibility of at least some omitted variables bias in our analysis because there could be one or more "unobservables" not included in our propensity analysis that are correlated either to treatment assignment and/or potential outcomes. As a result, after estimating the effects of teachers' knowledge on students' achievement controlling for propensity strata, we conducted a sensitivity analysis to provide information about how the results of our study might be biased, given a range of plausible assumptions about omitted variables bias.[4] This analysis is described next.

## Hierarchical Linear Model for Estimating Teacher Knowledge Effects

After grouping teachers into five strata based on their propensity score at each grade level, we proceeded to estimate the effects of teachers' knowledge on students' achievement using three-level, hierarchical linear regression models (Raudenbush & Bryk, 2002). In all, we estimated six separate regression models, one for each outcome variable (students' ITBS scale scores for word analysis and for reading comprehension) at each grade level. In each regression model, we tested the null hypothesis that the effect of teachers' TKRRP scores on the respective student outcome was zero. In testing this hypothesis, we estimated each of the models using five multiply imputed data sets. Accordingly, the point estimates, standard errors, variance components, and degrees of freedom in these analyses were based on all five data sets and were adjusted for the variance in parameter estimates within data sets and the variance in the parameter estimates between data sets (Peugh & Enders, 2004; Raudenbush & Bryk, 2002).

[4]Readers interested in the specific propensity score models estimated here can consult the technical report by Carlisle et al. (2008). The propensity score models are quite complex, involving several extensions to such models used in studies where treatments are categorical and treated subjects are not nested within higher level units. In particular, in developing a propensity score model to predict the likelihood that classrooms were headed by more and less knowledgeable teachers, we built on the work of Imai and Van Dyk (2004) on propensity score modeling for continuous treatment variables and on the work of Hong and Raudenbush (2006) on multilevel propensity score stratification.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

All independent variables at Levels 1 and 2 of the hierarchical linear models were centered around their respective grand means, and the outcome variables and the measure of teachers' knowledge were standardized to have a mean of zero and a standard deviation of 1. At Level 1 of the model, we included seven student covariates that have been found in previous research to be related to students' reading achievement: *male* (an indicator of whether a student is male), *age* (a continuous measure of each student's age in months), *disabled* (an indicator that specifies whether or not a student was coded as having a disability and received special education services), *LEP* (an indicator for whether or not a student has limited English proficiency), *FRL* (an indicator of eligibility for free or reduced-price lunch), *White* (an indicator of whether or not a student is White), and a student's *DIBELS score* for Fall 2006 and—for Grades 2 and 3 only—a student's *ITBS reading comprehension and word analysis scores* from the spring of the previous year (2006). The general form of the Level 1 model, then, was

$$Achievement_{ijk} = \pi_{0jk} + \sum_{p=1}^{n=7} \pi_p X_{pijk} + \varepsilon_{ijk} \tag{4}$$

where $Achievement_{\mathrm{ijk}}$ represents a reading achievement outcome for student $i$ in classroom $j$ in school $k$, which is seen as varying around $\pi_{0jk}$ (the average student score adjusted for the student-level independent variables (X) in the model); $\pi_{\mathrm{p}}$ are the seven regression coefficients for each of these independent variables, and $\varepsilon_{ijk}$ is a random effect for each student in the data set, where these random effects have a normal distribution with mean zero and variance $\sigma^2$.

At Level 2 of the model, the adjusted average achievement of students, $\pi_{0jk}$, was modeled as a function of indicators for the teacher subclasses estimated by the propensity score, $S_1$, $S_2$, $S_4$, $S_5$ (subgroup 3 is the reference group), a random effect, $r_{0jk}$, and a measurement of each teacher's reading knowledge, *TK*. The form of the model at Level 2 was

$$\pi_{0jk} = \beta_{00k} + \beta_{01} TK_{jk} + \sum_{q=2}^{5} \beta_{0q} S_{qjk} + r_{0jk} \tag{5}$$

where $\beta_{00k}$ is the average adjusted student achievement for a teacher's class, $\beta_{01}$ is average effect of teacher knowledge (*TK*) on adjusted achievement, and $S_{qjk}$ are the strata indicators with corresponding coefficients, $\beta_{0q}$. Moreover, $r_{0jk}$ is the random effect of teacher $j$ in school $k$ and has a normal distribution with mean zero and variance $\tau_\pi$.

Finally, we specified the school level of our model to contain only a random intercept, such that

$$\beta_{00k} = \gamma_{000} + u_{00k} \tag{6}$$

where $\gamma_{000}$ is the school average achievement and $u_{00k}$ is the random effect associated with each school and is distributed as normal with mean zero and variance as $\tau_\beta$. In constructing the Level 2 and 3 portions of the model, we note that teacher and school characteristics were adjusted through the propensity score strata rather than through covariance adjustment at the respective level. In addition, interactions and higher order terms (e.g., squared and cubic terms) were considered, and we selected models based on chi-squared tests of deviance statistics.

### Sensitivity Analysis

In these models, our causal estimates will be unbiased only if we are comparing teachers within schools and districts with similar characteristics. If we have not constructed comparable groups (i.e., through omission of key variables that are related both to the assignment of teachers with more or less knowledge and to students' reading achievement outcomes), our causal estimates will be biased. To assess the robustness of our inferences about the effects of teachers' knowledge on students' achievement, we conducted a sensitivity analysis (e.g., Rosenbaum, 1995). This analysis described the magnitude of relationships between an unobserved variable treatment assignment (in this case, the classroom teachers' TKRRP score) and student outcomes that would be needed to alter the original inference about the effects of teacher knowledge on students' achievement. In particular, the impact of an omitted variable on the estimated effect of teacher knowledge is dependent on the omitted variable's relationship with teacher knowledge, the ITBS outcome, and its relationship with measured covariates. To empirically characterize the potential impact of an omitted variable, we followed Hong and Raudenbush (2006) by assuming that the omitted variable had relationships with teacher knowledge and achievement similar in magnitude to one of the measured covariates. Further, to allow maximal impact, we conservatively assumed the omitted variable had no relationship with other measured variables. By estimating each measured variable's unconditional relationship with teacher knowledge and the outcome, we constructed several scenarios in which hypothetically omitted variables might work to alter our inferences. We then re-estimated the effects of teachers' knowledge on our outcome variable, accounting for each hypothetical omitted variable one at a time. If the inclusion of this hypothetical variable altered the statistical significance of our estimated teacher knowledge effect, we concluded that our results were sensitive to omitted variables bias. The results of this analysis are discussed next.

## RESULTS

We present our results in four sections. The first section gives the results of psychometric analyses of the TKRRP measure to address our first research question about the measurement properties of our newly constructed measure of teachers' knowledge. The next three sections provide analyses that address our second research question, which concerns the effects of teachers' knowledge on students' reading achievement. The second section, for example, describes our analyses of the distribution of teacher knowledge across teachers, schools, and districts. These results set the stage for the third section, which presents our model-based estimates of the effects of teachers' knowledge on students' reading achievement. The fourth section describes our analysis of the sensitivity of these estimates to omitted variables bias.

### Psychometric Analyses of TKRRP

In conducting a psychometric analysis of teacher responses to the TKRRP, we began with a binary exploratory factor analysis using marginal maximum likelihood to assess the number of dimensions of teachers' knowledge measured by the TKRRP scale. Results suggested that, in contrast to at least some prior research (Garet et al., 2008), the TKRRP item pool was best fit using a single underlying dimension of teachers' knowledge (Carlisle et al., 2008). On the basis of this analysis, we used a one-parameter Item Response Theory (IRT)

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

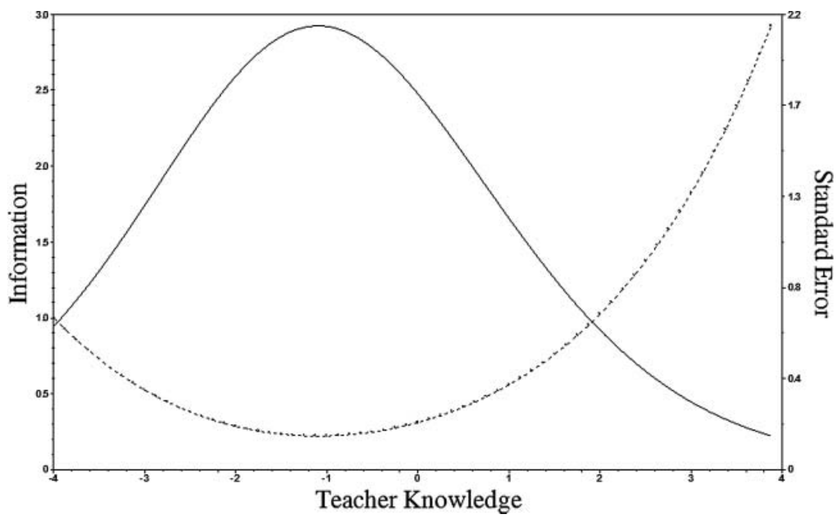



***Figure 1.*** Teachers' knowledge test information curve and standard error (Rasch).

model (Hambleton, Swaminathan, & Rogers, 1991) to combine the items into a single measure of teachers' knowledge about early reading.[5]

The TKRRP measure of teachers' knowledge for reading had a coefficient alpha of .756 and a one-parameter IRT reliability of .762. However, inspection of the test information curve (shown in Figure 1) shows that our measure was most reliable for teachers whose TKRRP scores were 1.25 *SD below* the mean score on the measure. Because test information is directly related to reliability, and because IRT-based test scores vary in reliability across persons, this statistic provides a useful diagnostic tool for determining the extent to which the TKRRP assessment is more or less reliable at particular points in the distribution of teacher scores. In our case, the test information curves show that the knowledge of teachers with scores well below the mean on the TKRRP was measured with greater reliability than was the knowledge of teachers with scores at or above the mean. In essence, the measure we developed can accurately assess whether teachers know relatively little about the content included on the TKRRP.

## Propensity Score Analysis

Once we developed the TKRRP measure of teachers' knowledge, we estimated the propensity score model discussed earlier. In essence, the estimation of this model allowed us to test the hypothesis that teachers' knowledge (as assessed by TKRRP) might be distributed unevenly across teachers, schools, and districts. At Grade 1, the propensity score analysis showed that approximately 11% of the variance in our TKRRP measure was among schools, 7% was among districts, and the remaining 82% was among teachers. Similar results were obtained using the data from Grades 2 and 3. In general, these analyses support the hypothesis that teachers who scored higher and lower on our TKRRP measure were

[5]We additionally assessed a two-parameter IRT model; however, Akaike information criterion and Bayesian information criterion indices indicated that the one parameter was sufficient. Further, the one- and two-parameter scores correlated around 0.99.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

unevenly distributed among schools and districts in the sample, although the analysis also found substantial variation in TKRRP scores among teachers located in the same schools and districts.

The propensity score analysis also allowed us to investigate the specific characteristics of teachers, classrooms, schools, and school districts that accounted for the observed variance in teachers' TKKRP scores. Here, we found that first-grade teachers scored higher on the TKRRP than did teachers at other grades (see the first row of Table 2); within grades, higher scoring teachers tended to be White women who had majored in early childhood education and were more experienced in teaching. Schools where teachers had higher average TKRRP scores tended to have a higher percentage of White teachers and a lower percentage of African American students; they tended to have teachers who had more professional trainings. Districts whose teachers had higher average knowledge scores tended to enroll a lower percentage of African American students and have higher percentages of White teachers and teachers with a master's degree.[6]

## Effects of Teachers' Knowledge on Students' Achievement

As discussed, the propensity score analysis was used to form five "propensity strata" within which teachers tended to have similar scores across a wide range of covariates. To control for differences in teacher background, we then added indicators variables for the propensity strata in which a given teacher was located (as shown in Equation 2), which allowed us to estimate the effect of teachers' TKRRP scores on students' reading achievement across comparable groups of teachers. The reader will recall that our analyses of the effects of teachers' knowledge on students' reading achievement were conducted separately for each grade and that for each grade, separate analyses were conducted for two ITBS subtests—word analysis and reading comprehension.

As a first step in the outcomes analysis, we partitioned the variance in students' reading achievement score into three components in a fully unconditional model (i.e., model with no covariates). As shown in Tables 3, 4, and 5, in these unconditional models, the majority of variance in students' achievement was among students within classrooms (between 81% and 87%, depending on the grade level and achievement domain). This variation, it should be noted, represents variance attributable both to errors in the measurement of student achievement and to variation among student outcomes, due to such factors as natural aptitude, motivation, and family support. A smaller yet statistically significant amount of variance in student achievement outcomes was found among classrooms (teachers) and schools: between 8% and 15% of the variance for classrooms (teachers) and between 3% and 8% for schools.

At the next step in the analysis, we estimated the multilevel statistical model described in Equations 1 to 3 for each grade level/achievement outcome of interest. The results of these analyses are presented in Tables 3, 4, and 5. In these tables, the dependent variables are listed in the columns (ITBS word analysis and reading comprehension scale scores), and the independent variables are listed in the rows.

[6]As noted earlier, a technical report by Kelcey et al. (2008) contains further details about the propensity score analyses just described. This report contains tables showing the effects of a wide range of covariates on teachers' TKRRP scores, describes how the five propensity strata used in the HLM analyses were constructed and presents the tests of covariate balance that we conducted in order to assess whether teachers within propensity score strata were balanced in terms of a wide range of covariates.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

**Table 3.** Effects of teachers' knowledge and other covariates on students' reading (Grade 1)

| | Estimate (SE) | |
|---|---|---|
| Effect | Word Analysis | Reading Comprehension |
| Intercept ($\pi_0$) | 0.40* (0.18) | 0.34 (0.18) |
| Teacher knowledge ($\beta_{01}$) | 0.02 (0.03) | 0.08* (0.03) |
| Strata 1 ($\beta_{02}$) | −0.02 (0.08) | 0.06 (0.07) |
| Strata 2 ($\beta_{03}$) | −0.02 (0.07) | 0.05 (0.06) |
| Strata 4 ($\beta_{04}$) | −0.07 (0.07) | −0.02 (0.06) |
| Strata 5 ($\beta_{05}$) | 0.01 (0.08) | −0.05 (0.07) |
| Male student ($\pi_1$) | −0.07* (0.02) | −0.13* (0.02) |
| Age in months ($\pi_2$) | 0.00 (0.00) | 0.00 (0.00) |
| Disabled student ($\pi_3$) | −0.31* (0.04) | −0.21* (0.04) |
| Limited English proficiency Student ($\pi_4$) | −0.08* (0.04) | −0.10* (0.04) |
| Student eligible for free or reduced lunch ($\pi_5$) | −0.14* (0.02) | −0.20* (0.03) |
| White student ($\pi_6$) | 0.23* (0.03) | 0.22* (0.03) |
| Fall DIBELS score ($\pi_7$) | 0.49* (0.01) | 0.51* (0.01) |

| Random Effects | Unconditional Model | Final Model | Unconditional Model | Final Model |
|---|---|---|---|---|
| Teachers ($r$) | 0.11 | 0.08 | 0.09 | 0.06 |
| Schools ($u$) | 0.08 | 0.04 | 0.07 | 0.02 |
| Residual ($e$) | 0.81 | 0.57 | 0.84 | 0.60 |

*Note*. DIBELS = Dynamic Indicators of Basic Early Literacy Skills.
$^*p < .05$.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011



**Table 4.** Effects of teachers' knowledge and other covariates on students' reading (Grade 2)

| Effect | Estimate (SE) Word Analysis | Estimate (SE) Reading Comprehension |
|---|---|---|
| Intercept ($\pi_0$) | 1.32* (0.18) | 1.05* (0.19) |
| Teacher knowledge ($\beta_{01}$) | 0.02 (0.04) | 0.02 (0.03) |
| Strata 1 ($\beta_{02}$) | 0.03 (0.09) | −0.03 (0.07) |
| Strata 2 ($\beta_{03}$) | −0.06 (0.07) | −0.10 (0.06) |
| Strata 4 ($\beta_{04}$) | −0.00 (0.07) | 0.01 (0.06) |
| Strata 5 ($\beta_{05}$) | 0.11 (0.08) | 0.02 (0.07) |
| Male student ($\pi_1$) | −0.01 (0.02) | −0.03 (0.02) |
| Age in months ($\pi_2$) | −0.01* (0.00) | −0.01* (0.00) |
| Disabled student ($\pi_3$) | −0.18* (0.03) | −0.21* (0.03) |
| Limited English proficiency student ($\pi_4$) | −0.07 (0.04) | −0.02 (0.04) |
| Student eligible for free or reduced lunch ($\pi_5$) | −0.11* (0.02) | −0.16* (0.02) |
| White student ($\pi_6$) | 0.08* (0.02) | 0.10* (0.03) |
| Fall DIBELS score ($\pi_7$) | 0.34* (0.01) | 0.41* (0.12) |
| ITBS–Reading comprehension spring 06 ($\pi_8$) | 0.15* (0.01) | 0.17* (0.01) |
| ITBS–Word analysis spring 06 ($\pi_9$) | 0.25* (0.01) | 0.15* (0.01) |

| Random Effects | Word Analysis Unconditional Model | Word Analysis Final Model | Reading Comprehension Unconditional Model | Reading Comprehension Final Model |
|---|---|---|---|---|
| Teachers ($r$) | 0.14 | 0.09 | 0.08 | 0.05 |
| Schools ($u$) | 0.07 | 0.01 | 0.08 | 0.02 |
| Residual ($e$) | 0.79 | 0.45 | 0.84 | 0.47 |

*Note*. DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ITBS = Iowa Tests of Basic Skills.
$^*p < .05$.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

**Table 5.** Effects of teachers' knowledge and other covariates on students' reading (Grade 3)

| Effect | Estimate (SE) Word Analysis | Estimate (SE) Reading Comprehension |
|---|---|---|
| Intercept ($\pi_0$) | 0.83* (0.20) | 0.55 (0.20) |
| Teacher knowledge ($\beta_{01}$) | 0.01 (0.02) | 0.01 (0.02) |
| Strata 1 ($\beta_{02}$) | −0.04 (0.06) | −0.03 (0.06) |
| Strata 2 ($\beta_{03}$) | 0.05 (0.05) | 0.01 (0.05) |
| Strata 4 ($\beta_{04}$) | −0.00 (0.07) | 0.02 (0.05) |
| Strata 5 ($\beta_{05}$) | −0.03 (0.06) | 0.01 (0.05) |
| Male student ($\pi_1$) | −0.04* (0.02) | −0.08* (0.02) |
| Age in months ($\pi_2$) | −0.01* (0.00) | −0.01* (0.00) |
| Disabled student ($\pi_3$) | −0.24* (0.03) | −0.14* (0.03) |
| Limited English proficiency student ($\pi_4$) | −0.16* (0.03) | −0.11* (0.04) |
| Student eligible for free or reduced lunch ($\pi_5$) | −0.14* (0.02) | −0.23* (0.02) |
| White student ($\pi_6$) | 0.12* (0.02) | 0.15* (0.03) |
| Fall DIBELS score ($\pi_7$) | 0.32* (0.01) | 0.31* (0.12) |
| ITBS–Reading comprehension spring 06 ($\pi_8$) | 0.17* (0.01) | 0.26* (0.01) |
| ITBS–Word analysis spring 06 ($\pi_9$) | 0.83* (0.20) | 0.55 (0.20) |

| Random Effects | Word Analysis Unconditional Model | Word Analysis Final Model | Reading Comprehension Unconditional Model | Reading Comprehension Final Model |
|---|---|---|---|---|
| Teachers ($r$) | 0.10 | 0.03 | 0.09 | 0.02 |
| Schools ($u$) | 0.04 | 0.01 | 0.06 | 0.01 |
| Residual ($e$) | 0.86 | 0.50 | 0.85 | 0.50 |

*Note*. DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ITBS = Iowa Tests of Basic Skills.
$^*p < .05$.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Looking across all three tables, we found that student characteristics explained statistically significant amounts of the variance in both the word analysis and reading comprehension achievement scores. In particular, at each grade level, students who scored higher on prior achievement tests tended to score higher on end-of-year reading achievement outcomes. Moreover, a student's race/ethnicity, eligibility for free or reduced-price lunch, and disability status also had statistically significant effects on achievement outcomes at all three grade levels. In the column labeled "Final Model," we can see that inclusion of student covariates, propensity strata, and a teacher's TKRRP score accounted for 20 to 40% of the between-classrooms variance in students' achievement (depending on the grade and achievement outcome under analysis) and about 75 to 80% of the variance in student achievement scores among schools (again depending on the grade and achievement outcome under analysis).

Although the results shown in Tables 3, 4, and 5 suggest that our statistical model accounts for substantial portions of variance in students' reading achievement among classrooms and schools, the results only partially support the hypothesis that a teacher's score on the TKRRP measure affects students' reading achievement. For example, Table 3 shows that first-grade teachers' knowledge, as measured by the TKRRP, had a statistically significant, positive effect on students' reading achievement—but *only* for students' ITBS reading comprehension scale score. As the table shows, a 1 *SD* increase in a teacher's knowledge score led to a 0.08 *SD* increase in first graders' reading comprehension achievement. However, this is the only significant effect of teachers' knowledge on students' achievement across all of the outcomes analyzed. That is, teachers' performance on the TKRRP measure did *not* have statistically significant effects on first-grade students' word analysis scale score, nor did the teachers' TKRRP score have statistically significant effects on second- or third-grade students' word analysis or reading comprehension test scores.

It is important to note that the effects just discussed are the *average* effects of teachers' knowledge on students' reading achievement. However, it is possible that the effect of teachers' knowledge on students' achievement varied across levels of teacher knowledge on the TKRRP measure. As a post hoc analysis, we examined this issue, using locally weighted polynomial regression analysis (e.g., Cleveland & Devlin, 1988). This analysis showed some evidence that the effects of teacher knowledge on students' achievement were stronger at the lower (as opposed to upper) end of TKRRP score distribution.

Figure 2 visually depicts this relationship in the first-grade data. The figure is a scatterplot, showing the fitted relationship between teachers' TKRRP scores (on the horizontal axis) and students' reading comprehension achievement scores (on the vertical axis), where the data points within the scatterplot are denoted by different symbols that stand for the propensity strata in which teachers were grouped. As Figure 2 shows, the slope of the line relating teachers' TKRRP score to students' achievement is much steeper at lower levels of TKRRP scores than at higher levels. This finding is important because, as noted earlier, the TKRRP measure used in the analyses reported here provided maximum information (i.e., was most reliable) for teachers who scored well below the mean on this teacher knowledge measure.

## Sensitivity Analysis

As a final step, we conducted a sensitivity analysis to assess how the effect of TKRRP scores on Grade 1 students' reading comprehension might change in the presence of omitted variables bias. This analysis indicated that our HLM estimate of teachers' knowledge on

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

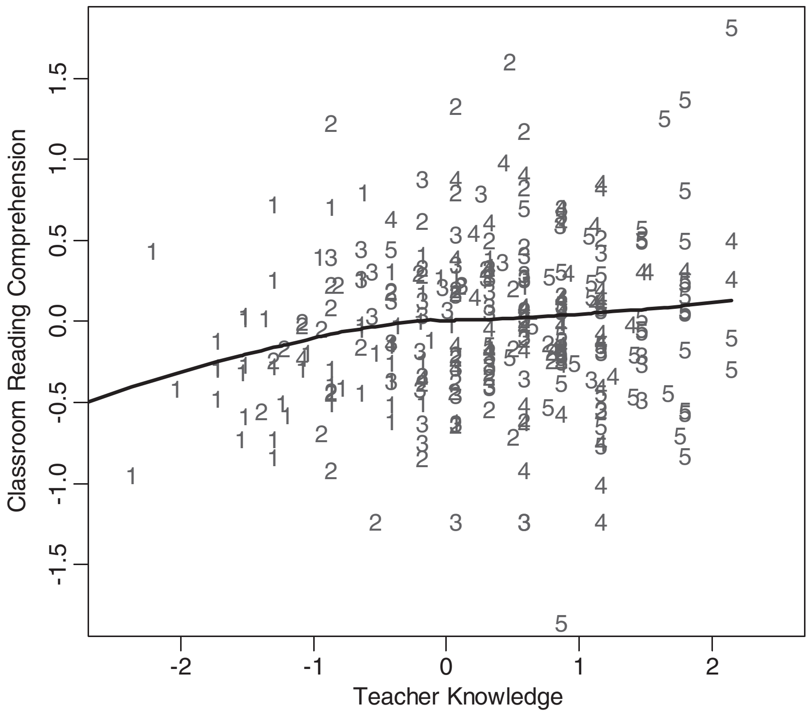



***Figure 2.*** Reading comprehension as a function of teacher knowledge by stratum.

first-grade students' reading comprehension achievement was robust to a wide range of assumptions about possible omitted variables but that the effect estimate would become statistically *in*significant if an unmeasured variable with effects on teacher knowledge and student achievement similar in magnitude to class average prior achievement was omitted from our regression analysis. That finding is particularly important because in the first-grade data set, we were able to use only one prior achievement measure (DIBELS), whereas at later grades, we included three prior achievement variables (DIBELS and two ITBS achievement scores). As a result, we cannot say with confidence that our estimated first-grade effect is robust to omitted variables bias.

## DISCUSSION

In this study we set out to address two problems confronting research on teachers' knowledge about early-grades reading. The first problem was to develop a new survey measure of teachers' knowledge in this domain, one that assessed not only teachers' knowledge about the linguistic foundations of reading but also knowledge about reading comprehension. Moreover, in assessing teachers' knowledge in these domains, we included survey items that assessed not just "academic" knowledge but also use of content knowledge in typical classroom situations. The second problem involved the use of this new measure in carrying out an empirical analysis of the effects of teachers' knowledge on students' reading achievement. Here, we developed a statistical model that addressed various problems that have plagued previous, nonexperimental studies of the relation of teachers' knowledge

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

and student outcomes—in particular, issues related to the clustering of higher knowledge teachers in particular kinds of schools and school districts and issues related to potential "omitted variables bias" in the estimation of teacher knowledge effects on students' reading achievement. We summarize and discuss our work on each of these problems in turn.

**Findings on the Psychometric Properties of the TKRRP Measure**

The measure of teachers' knowledge that we developed differed in certain respects from the measures used in previous research. First, it included items that reflected not only teachers' knowledge of the linguistic foundations of early reading but also their knowledge in the domain of reading comprehension. An important finding of our psychometric analysis was that items from both these content domains formed a *unidimensional* scale with a reasonable (albeit not strong) internal consistency of $\alpha = .756$. This finding contrasts with those of two other studies (Garet et al., 2008; Phelps & Schilling, 2004) that found that items in these domains formed two separable dimensions of teachers' knowledge. We are not sure why our results differ from those of these other studies. We can say that dimensionality in IRT measures arises for a number of reasons, including the relative mix of items in a measure that are drawn from particular content domains. Thus, although we make no strong claims about the dimensionality of teachers' knowledge of early reading, we strongly recommend additional measurement studies, using items that tap knowledge of different areas of reading.

Our psychometric analysis also revealed an important characteristic the TKRRP measure. As Figure 1 shows, the test information curve for our one-parameter IRT measure of teachers' knowledge indicated that the measure had much higher reliability at points where teachers' overall scores were well below the mean and that reliability dropped off sharply, once scores reached or exceeded the mean. We are uncertain as to why the TKRRP measure we developed had these properties, although one strong possibility is that this pattern is attributable to the unique characteristics of our study sample. A very large percentage of the teachers in the study sample had been working for more than a year in schools that participated in Michigan's Reading First program, and the extensive professional development that these teachers received as a result of their participation in this program could have affected our ability to reliably discriminate among teachers who scored at higher levels of the TKRRP measure. That is, the professional development associated with Reading First might have had strong effects on teachers' knowledge, which in turn affected the ability of our item pool to discriminate among teachers' with higher levels of knowledge about early-grades reading.

There is some evidence in our data of a professional development effect. In a subsidiary analysis, we found that 16% of the first-grade teachers, 15% of the second-grade teachers, and 23% of the third-grade teachers were new to Reading First in the year of our current study. Moreover, as Table 6 shows, at the first- and second-grade levels, these new teachers scored significantly lower on the TKRRP than did teachers who had been exposed to Reading First for more than 1 year. These results are consistent with research showing that participation in focused professional development can improve teachers' knowledge about how to teach reading effectively (e.g., Garet et al., 2008). These results suggest that there might be some value to conducting a psychometric analysis of the TKRRP measure in a sample of teachers who were not so uniformly exposed to professional development in the area of early-grades reading.

**Table 6.** Comparison of Teachers' Knowledge of Reading and Reading Practices scores of teachers new and old in reading first schools

| Teachers | Old | New | $p$ |
|---|---|---|---|
| Grade 1 | 0.27 | −0.04 | .019* |
| Grade 2 | 0.13 | −.28 | .0018 |
| Grade 3 | 0.10 | −.10 | .09 |

Finally, it is possible that the relatively weak effects of our TKRRP measure on students' achievement stem in part from the failure of the TKRRP to capture the kinds of knowledge that matter for teachers of second- and third-grade students for whom basic decoding is less an issue than comprehension of texts. Thus, we see a need for further exploration of ways to assess critical aspects of knowledge about reading. For example, one group of researchers has been investigating teachers' specialized knowledge for assisting students in understanding written texts, using a "video-viewing" task (Kucan, Palincsar, Khasnabis, & Chang, 2009).

## The Effects of Teachers' Knowledge on Students' Reading Achievement

Our second question focused on the effects of teachers' knowledge on students' reading achievement. Contrary to our expectations, this portion of the study did not provide strong support for the hypothesis that teachers' performance on TKRRP would significantly affect first through third graders' achievement gains. Indeed, across the six statistical models that we estimated, the standardized effects of TKRRP scores on students' achievement were .02 for word analysis and .08 for reading comprehension in first grade, .02 for both word analysis and reading comprehension in second grade, and .01 for both word analysis and reading comprehension in third grade. Although we expected the effects of teachers' knowledge on student outcomes to be relatively small, given the findings of others (e.g., Hill, Rowan, et al., 2005), we were surprised that the only statistically significant effect of teachers' TKRRP scores on students' reading achievement occurred in the first-grade sample, and then only for students' reading comprehension achievement. That effect would add about 3 to 4 weeks of additional learning to a student's reading comprehension score in first grade, moving the average first-grade student's test score from the 50th to the 53rd percentile.

## Contributions, Limitations, and Future Research

We see the study as having produced advances over previous research in various ways. For example, we have shown that it is possible to measure teachers' knowledge in domains other than the linguistic foundations of reading and that survey items can be developed to measure this knowledge as used in classroom situations. Nevertheless, additional research is needed, both to identify and measure other kinds of knowledge about early reading that might distinguish more and less effective teachers of reading in the early grades and to explore methods for measuring the enactment of knowledge in practice—the specialized knowledge needed to teach reading well.

Another contribution of our study is the development of a strategy for estimating the effects of teachers' knowledge on students' reading achievement, using nonexperimental

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

data. Because exploratory studies in education often use such data, we think the strategy we developed can be applied in future studies. In particular, our study suggests some strategies for taking into account the nesting of high-knowledge teachers in particular school and district settings through the use of multilevel propensity score stratification and for assessing the sensitivity of nonexperimental estimates of teacher knowledge effects on students' achievement to possible "omitted variables bias." We strongly encourage future exploratory studies that take advantage of these analytic approaches.

Several notes of caution need to be raised about our analyses of the effects of teachers' knowledge on students' reading achievement scores. First, as Figure 2 suggests, it is quite possible that unreliability in our measurement of teachers' knowledge produced significant *under*estimates of the effects of teachers' knowledge on students' reading achievement. Put differently, the measurement unreliability at higher levels of the TKRRP scale could easily have underestimated the average teacher knowledge effect, especially because measurement error in the independent variable tends to bias the relationship between that variable and the outcome measure toward zero.

At the same time, it is possible that the one statistically significant effect that we did find in our analysis was actually an *over*estimate of the true effect of teachers' knowledge on students' reading achievement. This is because the one analysis where we did find a statistically significant effect of teachers' knowledge on students' achievement (i.e., first graders' reading comprehension achievement) did not contain as full a set of pretest achievement measures as the analyses we conducted at other grade levels. Moreover, our sensitivity analysis showed that the results of this analysis could have been sensitive to this omission. Given these issues, we suggest that future research on the effects of teachers' knowledge on students' achievement might follow Sanders's (2006) advice to include as many measures of prior student achievement as possible in statistical models of teacher effects on students' achievement.

To some, the lack of inclusion of a measure of teachers' instruction in the present study might be seen as a major limitation. However, to others (e.g., Cochran-Smith & Zeichner, 2005; Moats, 2009a, 2009b), there are reasons to seek further understanding of the relation of teachers' content knowledge and students' reading outcomes. For example, Moats suggested that it is critical to understand the threshold of teachers' content knowledge that indicates that they have sufficient knowledge to teach reading effectively (Moats, 2009a). She expressed the hope that future studies might provide an assessment of teachers' knowledge capable of distinguishing teachers who are and are not adequately prepared to address the instructional needs of children who struggle in learning to read. As Moats (2009a) commented, "Teachers cannot teach what they do not understand themselves" (p. 387).

The study we conducted should not be considered to have produced definitive answers to the questions we raised at the outset of our study, but we believe that empirical research on the nature of teachers' knowledge and its effects on students' reading achievement cannot advance without additional theory development. We agree with Snow et al. (2005) and Foorman and Moats (2004) about the need for additional research that seeks to clarify the domains of knowledge required to teach early reading effectively, the ways in which such knowledge can be measured, and the processes by which teachers' knowledge works through instructional practice to affect student learning. Our results should not be interpreted as suggesting that teachers' knowledge is unrelated to the quality and effectiveness of their reading instruction. Rather, they illustrate the complexity of issues that need to be addressed to understand the extent to which teachers' knowledge about early-grades reading contributes to their students' achievement in reading over time.

## ACKNOWLEDGMENTS

This study was made possible by a Teacher Quality grant from the Institute for Education Sciences (IES) to the first author (Award #R305M050087); however, IES is not responsible for decisions about the method or interpretations of results expressed in the article. We are also grateful for the support of Michigan's Reading First program directors and elementary teachers who welcomed us into their classrooms. Finally, we particularly want to acknowledge the help and quiet wisdom of our colleague and friend, David Johnson, who passed away this winter.

## REFERENCES

Assessment Committee. (2002, February). *Analysis of reading assessment measures, Coding form for Dynamic Indicators of Basic Early Literacy Skills*. Retrieved from http:DIBELS.uoregon.edu
Authors (2009).
Bos, C., Mather, N., Dickson, S., Blanche, P., & Charde, D. (2001). Perceptions and knowledge of preservice and inservice educators about early reading instruction. *Annals of Dyslexia*, *51*, 97–120.
Bos, C. S., Mather, N., Narr, R. F., & Babur, N. (1999). Interactive, collaborative professional development in early literacy instruction: Supporting the balancing act. *Learning Disabilities Research & Practice*, *14*, 227–238.
Carlisle, J. F., Correnti, R., Phelps, G., & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing: An Interdisciplinary Journal*, *22*, 459–486.
Carlisle, J. F., Johnson, D., Phelps, G., & Rowan, B. (June 2008). Teacher knowledge of reading and reading practices. Technical Report #08-02. Assessment of Teachers' Knowledge About Reading. Ann Arbor, MI: University of Michigan.
Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596–610.
Cochran-Smith, M., & Zeichner, K. M. (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.
Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2003). *Teacher qualifications and first grade achievement*. College Park, MD: Center for Education Policy and Leadership.
Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Stanovich, P. J. (2004). Disciplinary knowledge of K–3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia*, *54*, 139–167
Darling-Hammond, L. (2004). Inequality and the right to learn: Access to qualified teachers in California's public schools. *Teachers College Record*, *106*, 1936–1966.
Foorman, B. R., & Moats, L. C. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remedial and Special Education*, *25*, 51–60.
Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, DC: U.S. Department of Education, National Center for Educational Evaluation and Regional Assistance.
Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.
Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results of California's mathematics professional development institutes. *Journal of Research in Mathematics Education*, *35*, 330–351.
Hill, H. C., Ball, D. L., & Schilling, S. (2005). Unpacking "pedagogical content knowledge": Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*, 372–400.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *Elementary School Journal*, *105*, 11–30.

Hong, G., & Raudenbush, S.,(2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*, 901–910.

Hoover, H. D., Dunbar, S. B., Frisbee, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., et al. (2003). *Iowa Test of Basic Skills: Guide to research and development*. Ithaca, IL: Riverside.

Imai, K., & Van Dyk, D. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistician*, *99*(467), 854–866.

Kainz, K., & Vernon-Feagans, L. (2007). The ecology of reading development for children in poverty. *Elementary School Journal*, *107*, 407–427.

Kelcey, B., Rowan, B., Carlisle, J., & Phelps, G. (September 2008). Studying the effects of teachers' knowledge about reading on students' reading achievement: A methodological report. Technical Report #08-03. Assessment of Teachers' Reading Knowledge. Ann Arbor: University of Michigan.

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, *39*, 484–486.

Kucan, L., Palincsar, A. S., Khasnabis, D., & Chang, C.-I. (2009). The Video Viewing Task: A source of information about assessing and addressing teacher understanding of text-based discussion. *Teaching and Teacher Education*, *25*, 415–423.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., et al. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, *35*, 69–86.

McCutchen, D., Harry, D. R., Cunningham, A. E., Cox, S., Sidman, S., & Covill, A. E. (2002). Reading teachers' knowledge of children's literature and English phonology. *Annals of Dyslexia*, *52*, 207–228.

Moats, L. C. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia*, *44*, 81–102.

Moats, L. C. (1999). *Teaching reading IS rocket science: What expert teachers of reading should know and be able to do* (No. 39-0372). Washington, DC: American Federation of Teachers.

Moats, L. C. (2003). *Language essentials for teachers of reading and spelling*. Longmont, CO: Sopris West.

Moats, L. (2009a). Knowledge foundations for teaching reading and spelling. *Reading and Writing: An Interdisciplinary Journal*, *22*, 379–399.

Moats, L. (2009b). Still wanted: Teachers with knowledge about language. *Journal of Learning Disabilities*, *42*, 387–391.

No Child Left Behind Act of 2001. (2002). Public Law No. 107-110, 115 Stat. 1425.

Nye, B., Konstantopoulous, S., & Hedges, L. V. (2004). How large are teacher effects? *Education Evaluation and Policy Analysis*, *26*, 237–257.

Palincsar, A. S., & Duke, N. (2004). The role of text and text-reader interactions in young children's reading development. *Elementary School Journal*, *105*, 183–197.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teachers of reading. *Elementary School Journal*, *105*, 31–48.

Piasta, S. B., Connor, C. M., Fishman, B. J., & Morrison, F. J. (2009). Teachers' knowledge of literacy concepts, classroom practices, and students' reading growth. *Scientific Studies of Reading*, *13*, 224–248.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C.C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, *5*, 35–58.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–96.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–458.

Rosenbaum, P. (1995). *Observational studies*. New York: Springer-Verlag.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for casual effects. *Biometrika*, *70*, 41–55.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers' College Record*, *104*, 1525–1567.

Sanders, W. L. (2006). Comparisons among various educational assessment value-added models. Presentation at the Power of Two—National Value-Added Conference. Columbus, OH: Battelle for Kids.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*, 299–311.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectivenss*. New York: Pergamon/Oxford.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L.S. (1987). Knowledge and teaching: Foundations for the new reform. *Harvard Educational Review*, *57*, 1–22.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snow, C. E., Griffin, P., & Burns, M. S. (Eds.). (2005). *Knowledge to support the teaching of reading*. San Francisco: Jossey-Bass.

Spear-Swerling, L., & Brucker, P. O. (2003). Teachers' acquisition of knowledge about English word structure. *Annals of Dyslexia*, *53*, 72–103.

Spear-Swerling, L., & Brucker, P. O. (2004). Preparing novice teachers to develop basic reading and spelling skills in children. *Annals of Dyslexia*, *54*, 332–364.

U.S. Department of Education. (2002, April). *Guidance for the Reading First Program*. Retrieved from http://www.ed.gov

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89–122.

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

## APPENDIX A

### Teachers' Knowledge about Reading and Reading Practices

### Part 2: Knowledge about Reading and Reading Practices
**Mark the best response to each question.**

31. Mr. Burnett noticed that some of his second graders are having difficulty reading common irregular words. To address this problem, Mr. Burnett created sets of words for students to practice. Which set is most suitable for this purpose? (Mark (X) one)

☐ a. when, until, which, after
☐ b. sweet, sugar, milk, banana

☐ c. because, does, again, their
☐ d. light, house, my, they

32. In her kindergarten class, Ms. Frank uses several different tasks to help her students identify sounds in words. Which directions indicate the use of a blending task? (Mark (X) one)

☐ a. “Put the sounds together to say the word. /t/a//p/.”
☐ b. “Tell me the first sound of ‘tap’.”
☐ c. “Say tap’. Now say it again but don’t say /t/.”
☐ d. “Say each sound in ‘tap’.”

33. Mr. Rink asked an aide to present each of the following words orally to a group of children and to have the children tell the aide how many phonemes (speech sounds) are in each word. Help create an answer key that Mr. Rink’s aide could use by marking (X) the number of phonemes contained in <u>each</u> word.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| a. freight | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ship | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. nation | ☐ | ☐ | ☐ | ☐ | ☐ |

34. A parent asks you what to do to help Juan, her second-grade son, become a more fluent reader. Which of the following the recommendation is most likely to help Juan develop reading fluency? (Mark (X) one)

☐ a. Have Juan read each book several times.
☐ b. Have him listen to books on tape.
☐ c. Have him read on his own for 20 minutes every evening.
☐ d. Read books to him every day.

35. A new third-grade teacher is having trouble picking books that are at the right reading level for his students. He asks you how he can help a student figure out whether a book is too hard. You suggest that he tell the student (Mark (X) one)

☐ a. to pick books on topics he/she knows something about.
☐ b. to avoid books with small print and few pictures or illustrations.
☐ c. not to pick books with more than five hard words on a page.
☐ d. not to select books written by unfamiliar authors.

36. During reading, analysis of word structure would be a useful strategy for understanding which of the following words? (Mark (X) one)

☐ a. discrimnate
☐ b. inalterable
☐ c. perspective
☐ d. institution

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

37. Mr. Danks, a kindergarten teacher, has students learn to recite nursery rhymes (such as Little Miss Muffet) and to sing songs (such as Twinkle, Twinkle Little Star). In what way are these activities most likely to support children's early reading development? Through fostering their (Mark (X) one)

☐ a. understanding of story structure.
☐ b. enjoyment of literature.
☐ c. development of vocabulary.
☐ d. development of phonological awareness.

38. The following are common words that children are usually taught to read in grades one through three. Some are phonetically regular (i.e., they conform to frequently-taught phonic rules in English), whereas others are phonetically irregular (i.e., they are exceptions to phonic rules). Please mark (X) whether each of the following words is phonetically regular or irregular.

| | Regular | Irregular |
|---|---|---|
| a. snowy | ☐ | ☐ |
| b. was | ☐ | ☐ |
| c. chunk | ☐ | ☐ |
| d. done | ☐ | ☐ |
| e. give | ☐ | ☐ |
| f. peach | ☐ | ☐ |

39. Mr. Lewis' class has been learning spelling rules for adding "ing" to base words. He is looking for groups of words that illustrate the various rules to give his students a complex challenge. Which of the following groups of words would be best for this purpose? (Mark (X) one)

☐ a. hopping, running, sending, getting
☐ b. hoping, buying, caring, baking
☐ c. seeing, letting, liking, carrying
☐ d. All of the word sets are useful for this purpose.

40. Mr. Hamilton, a first-grade teacher, notices that Rafael spends much of his free time writing. He notes that Rafael misspells many words but that his misspellings suggest knowledge of some letter-sound relations. For instance, he spelled *zipper* as *zipr* and *elephant* as *elifint.* To promote Rafael's spelling development, which would be the best step for Mr. Hamilton to take? (Mark (X) one)

☐ a. Engage Rafael in activities that promote phonological awareness.
☐ b. Teach him standard spelling patterns before he spends more time writing.
☐ c. Teach him standard spelling patterns within the context of his compositions.
☐ d. Encourage him to continue to write a lot.

41. Ms. Rico dictated the following story to her class:

I have a black and white dog.

Her name is Skipper.
One day she went to my school.
She liked playing with the kids.

She looked at her students' papers. Jesse's paper looked like this:

I have a blk and wit bog.
Hra name is skpr.
Wone bay she wat to mui skul.
She likt playg wethe the kibs.

Which of the following words in Jesse's writing provide evidence that Jesse can identify the correct number of speech sounds in words? (Mark "Yes" or "No" for each word.)

| | Yes | No |
|---|---|---|
| a. "blk" for black | ☐ | ☐ |
| b. "wit" for white | ☐ | ☐ |
| c. "skpr" for skipper | ☐ | ☐ |

42. Ms. Stanley, a kindergarten teacher, is preparing activities to teach phonological awareness in a developmentally appropriate sequence. Which of the following should she teach first? (Mark (X) one)

☐ a. Matching word sounds and letters.
☐ b. Identifying words that rhyme.
☐ c. identifying vowels that say their own name.
☐ d. Counting the number of speech sounds in words.

43. A first-grade teacher is preparing a read-aloud lesson for her class. She is thinking about selecting four or five words from the story to discuss with the students. Which category of words below, if selected by the teacher, will most affect whether students will understand the story? (Mark (X) one)

☐ a. names of characters
☐ b. the words that are hardest to pronounce
☐ c. words that students will encounter in other texts
☐ d. specialized words in the story

## APPENDIX B

### Variables Included in the Propensity Score Analysis

*Teacher-level pretreatment covariates:*
Gender of teacher, white teacher, black teacher, Hispanic teacher, Asian teacher, bachelors degree in early childhood education, bachelors degree in elementary education, bachelors

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011

degree in special education, bachelors degree in literacy education, masters degree, masters degree in elementary education, masters degree in early childhood education, masters degree in literacy education, masters degree in special education, post masters degree, possess a standard teaching certification, possess provisional certification, possesses a reading certification, possesses a special education certification, number of approved reading trainings/professional development seminars, number of years teaching, high number of years teaching, Reading First veteran status, average and standard deviation of class DIBELS nonsense word fluency in the fall, average and standard deviation of ITBS subtest scores (grades 2 and 3), proportion of class that is male, average age of class, proportion of class identified as special education, proportion of class eligible for free/reduced lunch, proportion of class identified as having a disability, proportion of class identified as having limited English proficiency, proportion of class that is black, proportion of class that is Hispanic and proportion of class that is white.

*School-level pretreatment covariates:*
School aggregates of all teacher and student characteristics as well as school wide measures of free/reduced lunch eligibility, proportion male and racial makeup.

*Cross-level interactions:*
Teacher's race (black), undergraduate certification (early childhood education), and reading certification were interacted with eligibility for free/reduced lunch, proportion students and teachers, proportion of teachers with post masters degree, average number of approved trainings, average prior abilities, proportion of teachers with high years experience, as well as separate random school effects (*r*).

Downloaded by [Joanne F. Carlisle] at 02:35 29 September 2011