CENTER FOR EDUCATION POLICY ANALYSIS at STANFORD UNIVERSITY
520 Galvez Mall, CERAS Building, 5th Floor | Stanford CA 94305-3084
Phone: 650.736.1258 | Fax: 650.723.9931

CEPA WORKING PAPER

Know the Child: The Importance of Teacher Knowledge of Individual Students' Skills (KISS)

Benjamin N. York

Working Paper

Center for Education Policy Analysis (CEPA)
Stanford University
520 Galvez Mall, CERAS Building, 5th Floor
Stanford, CA 94305

I give special thanks to Susanna Loeb, Deborah Stipek, Tom Dee, Julie Cohen, and Jane Rochmes for their thoughtful feedback on this paper. This research was supported by the Institute of Education Sciences (IES), United States (U.S.) Department of Education, through Grant R305B090016 to Stanford University. The opinions expressed herein are those of the author and do not reflect the views of IES or the U.S. Department of Education.

Know the Child: The Importance of Teacher Knowledge of Individual Students' Skills (KISS)
Benjamin N. York
CEPA Working Paper
March 2014

ABSTRACT

Teachers require knowledge of the unique skills that each child brings to the classroom in order to effectively target instruction towards students' learning needs. Despite substantial investments in programs aimed at enhancing teacher knowledge of individual students' skills (KISS), we know surprisingly little about how KISS is distributed or how teachers develop KISS, let alone the role that KISS plays in instruction and learning. In this study, I employ nationally-representative data to create KISS measures for kindergarten and first grade teachers. I use these measures to examine the distribution of KISS across schools, within schools, and within classrooms, as well as to investigate potential KISS development mechanisms and instructional uses. To estimate the effect of KISS on student learning, I use a set of student and subject fixed effects models that control for the non-random sorting of students into classrooms, average differences in how well teachers know particular children, and baseline achievement. I find that a standard deviation increase in KISS positively impacts kindergartners' and first graders' achievement by about 0.08-0.09 standard deviations. This result is highly robust to a number of different modeling choices and alternative explanations.

Benjamin N. York
Center for Education Policy Analysis (CEPA)
Stanford University
520 Galvez Mall, CERAS Building, 5th Floor, Room 511
Stanford, CA 94305
byork@stanford.edu

Introduction

Teachers require knowledge of the unique skills that every child brings to the classroom in order to target instruction towards students' needs – a pedagogical approach with strong empirical support (Connor, Morrison, Fishman, Schatschneider & Underwood, 2007; Connor et al., 2011). While school districts' use of programs designed to enhance teacher knowledge of individual students' skills (KISS) has rapidly grown over the past decade (Clune & White, 2008; Olson, 2005; Shepard, 2010), we know surprisingly little about KISS itself. Rigorous evaluations of programs aimed at improving KISS are sparse, mixed, and generally exclude analyses of program mechanisms, thus shedding little light on the importance of KISS (Carlson, Borman & Robinson, 2011; Henderson, Petrosino, Guckenburg & Hamilton, 2007, 2008; Quint, Sepanik & Smith, 2008). Moreover, the teacher knowledge literature largely ignores KISS, concentrating instead on other types of knowledge (Ball, Thames & Phelps, 2008; Hill, Ball & Schilling, 2008; Shulman, 1986). Currently, we know very little about how KISS is distributed, how teachers develop KISS, or the role of KISS in instruction and student learning.

In the present study, I review prior research on teacher knowledge, propose an initial definition for KISS, and use data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) to create KISS measures for kindergarten and first grade teachers. The ECLS-K data are uniquely well-suited for creating KISS measures because they include teacher ratings of students' proficiency in highly-specific topics, as well as information on students' actual proficiency in the same topics. KISS measures are based on how well teachers' ratings align with students' actual proficiencies. I use these measures to examine the distribution of KISS across schools, within schools, and within classrooms. I also investigate a number of potential mechanisms through which teachers might develop KISS, along with some of the possible instructional uses of KISS.

This chief innovation of this study is the empirical strategy that I employ for identifying the effect of KISS on student learning. Since the ECLS-K data has teacher ratings of students' proficiency and information of students' actual proficiency in both reading and math, I can create measures of teachers' KISS for every student in the sample in both subjects. The ECLS-K data also include students' spring direct cognitive assessment scores in reading and math. The combination of these features enables me to employ a series of student fixed effects models to estimate the impact of KISS on student learning. For example, in my preferred specification, I compare a teacher's KISS in reading for a particular child and the child's spring reading test score with the teacher's KISS in math for the same child and the child's math score (and repeat the process for all children in the sample). I also include subject fixed effects and baseline test scores in this specification. Estimates from this model are free of potential bias due to the non-random sorting of students into classrooms and average differences in how well teachers know particular children.

While the identification strategy described above goes a long way toward purging KISS effect estimates of bias, establishing a causal warrant for the estimates relies on two additional assumptions. First, to give a causal interpretation to student and subject fixed effects model estimates, it must not be the case that it is easier for teachers to gain KISS in the subjects that students are stronger in at the time of fall assessment. For example, if it is easier for a teacher to develop KISS for a child in math because the child is much stronger in math than reading, then it is likely the case that the students' relative subject strength, not KISS, is driving results. Nor can it be the case that teachers who are more skilled at teaching a particular subject gain more KISS and improve student achievement more in their strong subject as a strict function of their subject-specific teaching strength. For instance, if a teacher is much stronger at teaching reading than math and rates her students' proficiency in reading higher than her students' proficiency in math because she is more confident in her reading teaching abilities (and her students score higher on average in reading than math), then the teachers' relative strength in teaching reading is likely the true driver of results. To examine these alternative explanations, I conduct a series of statistical tests. I also examine the extent to which study results are robust to particularities of the ECLS-K data and alternative measurement and modeling choices.

I find that KISS is unevenly distributed by race and achievement between schools and within classrooms. While the sources of KISS are unclear, teachers with stronger knowledge in one subject provide more differentiated instruction in that subject. More importantly, I find that a standard deviation increase in KISS positively impacts kindergartners' and first graders' achievement by roughly 0.08-0.09 standard deviations. This result is highly robust to alternative explanations and different measurement and modeling choices.

The remainder of this paper proceeds as follows. In the next section, I review prior research on teacher knowledge and propose an initial definition for KISS. The subsequent section describes the study's data, measures, and methods. In the section after that, I present study results. I conclude this paper with a discussion of the findings.

Background

Prior Research on Teacher Knowledge

The benefits of instruction that is targeted towards each students' unique learning needs are becoming increasingly clear. For example, there is accumulating correlational evidence that the effects of different teaching strategies in elementary reading vary based on children's skills at the beginning of the year (Connor, Frederick & Katch, 2004; Connor, Morrison & Petrella, 2004; Connor & Slominski, 2006). One study shows that preschoolers with low initial letter-word recognition assessment scores differentially benefit from code-focused activities, whereas children with higher initial achievement benefit more than other students from meaning-focused activities (Connor & Slominski, 2006). A series of experiments based on this research provides even stronger evidence on the importance of targeted instruction (Connor, Morrison, Fishman, Schatschneider & Underwood, 2007; Connor et al., 2011).

To effectively target instruction towards students' unique learning needs, teachers require knowledge of individual students' skills. Perhaps not surprisingly, a number of programs aim to enhance teachers' KISS as a means to providing more targeted instruction. Probably the most common of these programs are interim assessments (Clune & White, 2008; Olson, 2005; Shepard, 2010). For example, a 2005 *Education Week* survey of superintendents found that roughly 80 percent of school districts were already using or planning to use interim assessments during the following school year (Olson, 2005). The theory of action underlying interim assessment programs is that the frequent provision and analysis of student performance data will enhance teachers' knowledge of individual students' skills and therefore help teachers target instruction towards student needs (Perie, Marion & Gong, 2009).

Despite substantial investments in programs aimed at improving teachers' KISS, we know surprisingly little about the direct effect of KISS on student learning. There are only a few well-identified evaluations of interim assessment programs, and findings from these studies are mixed. A small number of quasi-experimental studies fail to find significant interim assessment effects (Henderson, Petrosino, Guckenburg & Hamilton, 2007, 2008; Quint, Sepanik & Smith, 2008), but a randomized control trial (RCT) of a comprehensive data-driven instructional program finds a positive treatment effect on students' math achievement of roughly 0.06 standard deviations (Carlson, Borman & Robinson, 2011). However, due to the breadth of the intervention, it is unclear whether or not interim assessments caused the effect. More generally, interim assessment program evaluations do not examine mechanisms and therefore fail to shed much light on the importance of KISS.

The teacher knowledge literature provides few additional insights into the saliency of KISS. The preponderance of this research focuses on pedagogical content knowledge (Shulman, 1986) and its derivative (Ball, Thames & Phelps, 2008; Hill, Ball & Schilling, 2008). Shulman (1987) defines pedagogical content knowledge (PCK) as "that special amalgam of content and pedagogy that is uniquely the province of teachers" (p. 8). PCK includes: the most useful forms of representation of the most regularly taught topics in a teacher's subject area; a knowledge of what makes learning particular topics easy or difficult; the conceptions and preconceptions that students of different backgrounds and ages bring with them to the learning of these topics; and, knowledge of the strategies most likely to be effective in reorganizing students' understanding (Shulman, 1986).

While PCK was greeted with broad enthusiasm by educational researchers, the construct remained under-specified and little scholarly agreement about the composition of PCK existed into the 2000s (Ball, Thames & Phelps, 2008). In an effort to bring clarity and consensus the field, the Study of Instructional Improvement Group (Ball, 1990) drew on Shulman's (1986) original conceptualization of PCK and findings from context-rich qualitative studies of the types of knowledge required for effective teaching (Borko et al., 1992; Leinhardt & Smith, 1985) to introduce the concept of content knowledge for teaching, or CKT (Ball, Thames & Phelps, 2008; Hill, Ball & Schilling, 2008).

Some of the most well-developed teacher knowledge measures and most convincing empirical findings are based on CKT. For example, Hill, Rowan & Ball (2005) estimates the of effect of content knowledge for teaching mathematics (CKT-M) on students' test score gains, using a CKT-M measure that captures teachers' knowledge of number concepts, operations, patterns, functions and algebra (for an extensive description of the validation work pertaining to the CKT-M measure used in the study, see Hill, Schilling & Ball, 2004). Controlling for student and teacher covariates that are potentially related to CKT-M and student test score growth, this study finds that a standard deviation increase in CKT-M is associated with a gain of about two and one-quarter points on the Terra Nova, roughly equivalent to one-half to two-thirds of a month of student growth.

A random assignment study of a two-week instructional institute designed to improve kindergarten and first grade teachers' phonological and orthographic awareness knowledge provides additional evidence on the importance of CKT (McCutchen et al., 2002). This program addresses an extremely wide array of topics, from teachers' basic knowledge of phonological and orthographic awareness, to helping teachers learn to use children's spelling as a diagnostic tool in assessing their students' phonological awareness. The study finds positive treatment effects on teachers' knowledge of the instructional role of orthographic and phonological information, teachers' instruction (e.g., treatment group kindergarten teachers spent more time on explicit phonological awareness instruction than control group teachers), and student achievement.

Research on teacher knowledge of individual students' skills is far less developed. Nearly all of the work related to KISS comes from the literature on cognitively guided instruction

(Carpenter, Fennema & Franke, 1996). The basic approach of CGI is to combine information on how students generally develop understandings of particular concepts with teachers' knowledge of their own students' thinking to improve instruction and student learning (Carpenter, Fennema, Franke, Levi, & Empson, 2000). In an effort to validate this approach, CGI researchers examined the direct relationship between teachers' knowledge of their own students' thinking and student achievement. In a study based on 40 first grade teachers in Wisconsin and their students, Carpenter, Fennema, Peterson and Carey (1988) finds that teachers' ability to accurately predict whether or not six randomly-selected students in the classroom answered a question correctly is positively and significantly correlated with measures of student computation and problem solving (correlations of 0.32 and 0.31, respectively). A similar study uses data on 25 second grade and 21 fifth grade classrooms and finds a positive association between teachers' ability to predict how students would perform on test items and student achievement in both reading and math (Fisher et al., 1980).

While these studies hint at the importance of KISS, they have significant limitations. The teacher knowledge measure used in the studies described above – namely, a teacher's ability to accurately predict whether or not students answered a question correctly – has not been rigorously examined, and it is based on a coarse understanding of students' skills as well as a small number of teachers and children. We also lack an understanding of KISS is distributed or how teachers develop or use KISS. Finally, extant studies they tell us very little about the causal effect of KISS on student learning. For example, they do not control for variables that are potentially correlated with teachers' knowledge levels and student achievement, and might therefore confound the effect of teacher knowledge with other teacher qualities or student sorting.

The present study builds on the small research base related to KISS. First, I propose an initial definition for teacher knowledge of individual students' skills and describe how KISS is related to CGI. Then, I use nationally-representative date to create and validate KISS measures for kindergarten and first grade teachers. I use these measures to examine the distribution of KISS across schools, within schools, and within classrooms, as well as to investigate potential mechanisms through which teachers might develop KISS and possible instructional uses of KISS. Ultimately, I am interested in the causal effect of KISS on student learning. To estimate this effect, I use a series of student and subject fixed effects models and conduct several falsification tests.

An Initial Definition of KISS

I propose to define KISS as teachers' knowledge of individual students' skills and understanding of the most important concepts and topics in a given subject. A teacher's KISS for a particular student is determined by how close the teacher's perceptions of the student's skills are to reality. For example, if a kindergarten teacher perceives that a child is fully proficient in identifying lower-case letters and the child does indeed have full proficiency in lower-case letter identification, then the teacher has a high level of KISS for that child in lower-case letter identification. Conversely, if the teacher thinks that the same child completely lacks lower-case letter identification skills, then the teacher has weak KISS for that child in identifying lower-case letters.

Teachers' KISS in a given subject is not based on one particular skill. Building on the example above, kindergarten teachers' KISS in reading does not end with their knowledge of students' lower-case letter identification abilities. Kindergarten teachers are responsible for covering many additional early literacy topics, including upper-case letter knowledge, letter sound knowledge, beginning sound awareness, rhyme awareness, nursery rhyme awareness and conventions of print (COP). Only kindergarten teachers with strong knowledge of individual students' proficiency in all relevant early literacy topics have strong KISS in reading. Finally, in addition to child-level KISS, teachers' average level of knowledge of all of the students in their classroom, or teacher-level KISS, is important. For instance, teachers likely draw on their knowledge of all of the students in their classrooms to form instructional groups.

There are a number of possible channels through which teachers' could develop KISS. For example, some teachers' might have received training on how to cultivate this type of knowledge in college, graduate school, or during professional development sessions. Others might develop the ability to generate KISS through on-the-job experience. Gathering and reflecting on student information could be a particularly effective mechanism for developing KISS, given the abundance of potential sources of this type of information. For example, formative assessments, student work samples, and results from interim and prior years' summative assessments could be rich sources of information on students' skills.

KISS is clearly related to cognitively guided instruction. The goal of CGI is to improve teachers' knowledge of how children in general develop conceptual understandings and to use this knowledge along with knowledge of their own students' abilities to make instructional improvements (Carpenter, Fennema, Peterson, Chiang & Loef, 1989). Like KISS, CGI focuses on teachers' knowledge of their own students' abilities. However, teachers' knowledge of their own students' abilities is the primary focus of KISS. Making teachers' knowledge of their own students' abilities a primary focus is important because it places the teacher's attention on identifying the child's skills no matter how the child expresses them, as opposed to looking for a pre-defined set of thinking patterns within a child. Moreover, whereas CGI studies define teachers' knowledge of their own students' abilities as a teacher's ability to accurately predict whether or not students answered a question correctly, KISS requires that teachers' have a strong understanding of students' skills in all important topics in a given subject. KISS is therefore an extension of one of the types of teacher knowledge present in CGI.

Data and Methods

Study Participants

This study uses data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, developed under the sponsorship of the United States (U.S.) Department of Education, Institute of Education Sciences, National Center for Education Statistics (NCES). The ECLS-K is a longitudinal study of children's early school experiences. It followed a nationally-representative cohort of 21,260 students from the 1998-99 school year, when all of the students were in kindergarten, through 2006-07, when most students were in eighth grade. The ECLS-K assessed children's cognitive, socio-emotional, and physical development directly as well as through surveys of the children, their parents, and their teachers. It also collected information on children's home environments and educational activities, school and classroom environments, and teacher qualifications and practices.

The ECLS-K collected data at seven different time points: fall and spring of the 1998-99 school year, when all sampled students were in kindergarten; fall and spring of the 1999-2000 school year, when most students were in first grade (in the fall of first grade, the ECLS-K only collected data on students from a 30 percent subsample of schools); spring of the 2001-02 school year, when most students were in third grade; spring of the 2003-04 school year, when most students were in fifth grade; and, spring of the 2006-07 school year, when most students were in eighth grade. The ECLS-K spring first grade sample was freshened to include children who were not in kindergarten in the U.S. during the 1998-99 school year and is nationally representative of 1999-2000 first graders.

The ECLS-K employed a multistage probability design for sample selection, consisting of three stages. The primary sampling units (PSUs) are counties or groups of counties, selected with probability proportional to size. Size reflects the number of five years olds in the geographic area, adjusted to facilitate oversampling of Asian and Pacific Islanders (APIs). The secondary sampling units (SSUs) are schools within PSUs, selected with probability proportional to a weighted measure of size based on kindergarten enrollment (also modified to facilitate oversampling of APIs). Children within schools represent the final sampling stage units. Within schools, children were selected from one of two independent strata using equal probability systematic sampling (one strata contained APIs and another contained all other children). In general, the ECLS-K aimed to sample 24 children per school; however, there is meaningful variation in the number of students per classroom present in the sample.

The present study uses students' direct cognitive assessment scores, student demographic information, and teacher survey data to investigate teachers' knowledge of individual students' skills and understanding. These data are uniquely well-suited for studying KISS. Creating KISS measures requires objective evaluations of students' skills and knowledge of particular topics, as well as information on teachers' perceptions of children's skills and understanding of the same topics. Fortunately, the ECLS-K has both. Teachers rate students in reading and math, and there is student assessment data in both subjects. The presence of teacher and student data in both subjects allows me to create KISS measures for every student in reading and math.

Detailed information on students' skills and understanding comes from the ECLS-K spring proficiency probability scores. The ECLS-K uses children's raw assessment scores and an item response theory (IRT) model to calculate proficiency probability scores for children in several reading and math areas. These criterion-referenced measures reflect the probability that a child would have passed a proficiency level in a particular topic and can take any value between zero and one. For example, the ECLS-K calculates the probability that children were proficient at identifying upper- and lower-case letters at the end of kindergarten. Probabilities are based on clusters of assessment questions with similar content and difficulty, included at several points along the score scale in the reading and math assessments. They are more reliable than those based on individual items, as they reduce measurement error resulting from correct guesses. Proficiency probability scores are suitable for studying children's specific math and reading knowledge and skills (NCES, 2001).

Teachers' perceptions of students' skills and understanding of math and reading topics come from the Academic Rating Scale (ARS) in the ECLS-K spring teacher questionnaire. The survey asked teachers to rate children's skills, knowledge, and behaviors with respect to math and reading topics. Teachers rated students using a five point scale consisting of the following categories: Not yet, Beginning, In progress, Intermediate, and Proficient. For example, ECLS-K defines "Not yet" as "Child has not yet demonstrated skill, knowledge or behavior," whereas it defines "Proficient" as "Child demonstrates skill, knowledge or behavior completely and consistently" (NCES, 2001, Section 2, p. 15).

I restrict my analyses to kindergarten and first grade as the topical overlap between students' proficiency probability scores and teacher ratings is highest in these grades. I also restrict the data in other ways. I remove students not linked to a particular teacher from the kindergarten sample, since data on these students do not contribute to teacher knowledge measures. I also exclude kindergartners who do not have fall and spring reading and math direct cognitive assessment scores (fall test scores are an important statistical control in a number of my statistical models), as well as students missing teacher's math and reading academic development ratings from the spring. I remove students who changed teachers during the year because teachers had less time to develop knowledge of these students. I also exclude students receiving special services given the possibility that they receive a lot of outside supports that could affect teachers' KISS and student outcomes. Finally, I include only first time kindergartners, as teachers who have many repeating students could have artificially-high levels of KISS. After all restrictions, the kindergarten sample contains information on 13,745 students, 2,598 teachers, and 909 schools.

Given that the ECLS-K only sampled 30 percent of schools in the fall of the first grade year, I do not exclude first graders without initial achievement data. In general, there is less information on kindergartners than fist graders due to the ECLS-K sampling scheme (e.g., data on student mobility in first grade is incomplete). Consequently, the first grade sample is less restricted than the kindergarten sample. In particular, it only excludes students not linked to teachers, students receiving special services, and students missing spring test scores or teacher ratings in either reading or math. After all restrictions, the first grade sample contains information on 12,787 students, 3,649 teachers, and 1,347 schools.

Table 1 describes this study's samples. Students in this study are highly diverse. Roughly half of kindergarteners are females; and, black, Hispanic, Asian and white students make up about 16, 14, three, and 62 percent of the kindergarten sample, respectively. The average ECLS-K direct cognitive assessment *T* scores in reading and math of kindergartners are 50.50 and 51.22.

T scores are versions of the assessment scores standardized within each assessment wave with a mean of 50 and standard deviation of 10. They are suitable for repeated cross-sectional analyses like those conducted in this study (Reardon, Cheadle & Robinson, 2009).

In terms of gender and race, teachers in this study are less diverse than the students in the study. Approximately 98 percent of kindergarten teachers are female and roughly 86 percent are white. However, there is a good amount of variation in teachers' educational backgrounds, experience, and the characteristics of their classrooms. For example, the within-grade teaching experience and class size ranges for kindergarten teachers are one to 20 years and 14 to 25 students, respectively.

Measures

This study's analyses rely on student- and teacher-level KISS measures in reading and math in kindergarten and first grade. I begin by constructing student-level KISS sub-measures in particular reading and math topics. In kindergarten, I create student-level KISS sub-measures in letter recognition, familiarity with the conventions of print, addition and subtraction, and ordinality and sequence. In first grade, I create sub-measures in familiarity with COP, comprehension of words in contexts, ordinality and sequence, and relative size. To create KISS sub-measures, I compare teachers' spring ratings of students' proficiency in the reading and math topics described above with students' actual spring proficiency in the same topics. For example, to create the kindergarten letter recognition sub-measure, I compare teachers' ratings of students' ability to easily and quickly name all upper- and lower-case letters of the alphabet with kindergartners' proficiency probabilities in letter recognition. Because teachers are asked about their students in the spring, I use students' spring proficiency probabilities for making comparisons.

Turning comparisons of teachers' ratings of students' proficiency and students' actual proficiency into KISS measures that are suitable for analysis requires making assumptions about how teacher ratings map on to the continuum of student proficiency. Consider a teacher who rates a student's proficiency in conventions of print as "Not yet." To be able create a usable measure of the teacher's knowledge of that student's COP skills, one must make an assumption about what "Not yet" means in terms of actual student proficiency. If "Not yet" is equivalent to a proficiency probability score of 50 percent or less, and the student's probability proficiency score is 49 percent, then the teacher's rating of the student matches the student's actual proficiency level, and the teacher has strong knowledge of the student. However, if a score of 49 percent is equivalent to "Beginning" proficiency, then the teacher's rating of the student is one category off from the student's actual proficiency level. In this case, the teacher's knowledge of the student is still good, but not as strong as it was in the first example.

For this study's primary analyses, I assume that KISS is defined by the following criteria: "Not yet" proficient includes proficiency probability scores of five percent or less; "Beginning" proficiency includes proficiency probability scores in the range of six to 50 percent; "In progress" includes proficiency probability scores is in the range of 51-70 percent; "Intermediate" includes proficiency probability scores in the range of 71-90 percent; and, "Proficient" includes proficiency probability scores above 90 percent. The linkage for familiarity with conventions of print is somewhat different due to the unavailability of COP proficiency probability scores. Students' familiarity with COP scores reflect the number of COP items out of three that they answered correctly. For familiarity with COP, I link "Not yet" to zero correct answers, "Beginning" to one correct answer, "In progress" and "Intermediate" to two correct answers, and "Proficient" to three correct answers.

Clearly, these schemata are merely two of many possible ways to link teacher ratings to students' actual proficiencies. My justification for choosing the criteria described above is that they are intuitive. For example, it is reasonable to think that a student who is "Not yet" proficient at a particular skill, defined by ECLS-K as the "child has not yet demonstrated skill," has a very low probability of proficiency in the skill, such as five percent or less. Likewise, it is not unreasonable to assume that a child who is "Proficient" in a skill, defined as "child demonstrates skill completely and consistently," has a probability of proficiency of 90 percent and greater. Given that there are many plausible ways to link the data, I replicate study results in auxiliary analyses using alternative KISS measures.

Based on the schemata described above, I construct numerical student-level KISS sub-measures. In particular, I create a four point scale for each of the topics described above, in which a score of four represents an exact match between a teacher's rating of a student's skills and the student's actual skill level; three represents a comparison in which the teacher's rating is one category off from the student's actual skill level; two represents a comparison in which the teacher's rating in two categories off; a one is a comparison in which the teacher's rating is three categories off; and, a zero represents a comparison in which the teacher's rating is four categories off.

For example, if a kindergarten teacher rated a student's letter identification proficiency as "Not yet" in the spring, and the child's letter identification proficiency probability was five percent or less in the spring, then the teacher received a KISS of four for that child in letter identification. If the teacher rated the same child's proficiency as "Beginning," the teacher received a score of three. If the teacher rated the child's proficiency as "In progress," the teacher received a score of two. If the teacher rated the child's proficiency as "Intermediate," the teacher received a score of one. The teacher received a score of zero in letter identification for the child if the teacher rated the child as "Proficient."

Based on this construction, the lowest possible KISS score that a teacher can receive for a student in the first achievement quintile (zero to five percent probability of proficiency) is zero. In particular, a teacher receives this score if she or he rates the student a "Proficient." However, the lowest possible score that a teacher can receive for students in the second and fourth quintiles is one. For example, if a student is in the fourth achievement quintile, at most, a teacher's rating can be three categories off (i.e., when the teacher rates the student as "Not yet" proficient). Similarly, the lowest possible score a teacher can receive for students in the third quintile is two. Finally, as is the case for students in the first quintile, the lowest possible KISS score a teacher can receive for students in the fifth quintile is zero.

I use student-level KISS sub-measures for individual topics to create overall student-level KISS measures in reading and math (KISS-Reading and KISS-Math). To create student-level KISS-Reading measures in kindergarten, I standardize the familiarity with COP and letter identification KISS sub-measures and then take an average of these standardized scores, which I also standardize. Standardizing the measures in this way makes it easy to interpret KISS effect estimates and compare them with estimates of the effects of other educational interventions. To create student-level KISS-Math measures in kindergarten, I standardize the ordinality and sequence and addition and subtraction knowledge sub-measures, take an average of these standardized scores, and then standardize them. I repeat this process for first grade teachers, although as described above, the topics underlying first grade KISS measures are somewhat different than those that comprise the kindergarten measures. To create teacher-level KISS-Reading and KISS-Math measures, I take classroom averages of student-level KISS-Reading and

KISS-Math scores for all of the teachers in the sample. I then standardize these averages to enhance interpretability and comparability.

Table 2 summarizes this study's KISS measures. By construction, all of the measures have a mean of zero and a standard deviation of one. While the distributions of KISS measures are slightly negatively skewed, and the kurtosis of the distribution of the first grade teacher-level reading measure is substantially greater than three (the kurtosis of a normal distribution), in general, the distributions approximate normality. They are all unimodal and the average z-scores at the 10^{th} and 90^{th} percentile across the distributions are about -1.26 and 1.15 standard deviations, respectively (close to the analogous standard normal distribution z-scores of -1.28 and 1.28, respectively).

To further assess the validity of this study's KISS measures, I conduct a series of exploratory and confirmatory factor analyses. Exploratory factor analysis is useful for identifying the underlying dimensionality of data, while confirmatory factor analysis is useful for verifying the dimensions. In the first exploratory analysis, I conduct an iterated principal-factor analysis (IPF) of all student-level reading and math KISS sub-measures in kindergarten. The initial factor solution indicates that two factors account for over 98 percent of the variance in the data (the first factor accounts for about 58 percent of the variance). To get a better qualitative sense of these factors, I rotate the factor loadings using varimax rotation (this type of rotation produces orthogonal factors). The factor loadings clearly indicate that the first factor is defined by the math KISS sub-measures and the second factor is defined by the reading sub-measures. For example, the ordinality and sequence, addition and subtraction, conventions of print, and letter recognition KISS sub-measure loadings on the first factor are: 0.61, 0.64, 0.01 and -0.08, respectively.

Using the same data, I then conduct a principal components analysis and again rotate the factor loadings using varimax rotation. This confirmatory analysis makes it even clearer that the data are defined by KISS-Reading and KISS-Math factors. The COP and letter recognition knowledge KISS sub-measure loadings on the first factor are 0.75 and 0.79, respectively, while the ordinality and sequence and addition and subtraction sub-measure loadings on the second factor are 0.85 and 0.77, respectively. I repeat these procedures for the teacher-level KISS measures in kindergarten, as well as the student- and teacher-level measures in first grade. The results of these analyses are consistent with those described above.

While this study's KISS measures appear to be reliable, they are limited in a couple of ways. First, both KISS-Reading and KISS-Math cover a narrow range of topics (e.g., KISS-Reading in kindergarten only incorporates letter recognition and conventions of print). It would be ideal if KISS measures captured all important topics in a given subject, but unfortunately, the measures in this study are limited by the ECLS-K data. Moreover, due to the ECLS-K sampling strategy, the average KISS of different teachers comes from different numbers of students. Below, I analyze the extent to which study results are sensitive to excluding teachers with few students in the sample.

Methods

The Distribution of KISS across Schools

Given that other important aspects of teacher quality are unevenly distributed (Hill, Rowan & Ball, 2005; Lankford, Loeb & Wyckoff, 2002), I begin my analysis of KISS by examining its distribution across schools by race, socioeconomic status, and achievement. To start, I conduct a series of Student's t-tests to determine whether or not differences in teachers' knowledge of black, Hispanic, Asian and white students are statistically significant. I then examine the distribution of KISS among schools with different levels of initial academic performance. To conduct this analysis, I begin by calculating average fall test scores for schools. Then, I calculate the average KISS for schools in different quantiles of the fall school-average test score distribution, and run a series of Student's t-test to determine whether or not differences in these averages are statistically significant.

The Individual and Classroom Characteristics of High-KISS Teachers

In the next step of my analysis, I examine the relation between KISS and a number of teacher- and classroom-level characteristics, including: number of child development, early childhood education, elementary education, reading teaching methods, and math teaching methods courses completed in college; whether or not the teacher has a graduate degree; teaching certification type (e.g., alternative or regular); whether or not the teacher has an early childhood certification; whether or not the teacher has an elementary certification; total number of years teaching in the grade and overall; class size; and, the within-class standard deviations of fall reading and math scores. For first grade teachers, I also analyze the link between KISS and teachers' majors in undergraduate and graduate school.

Given the non-random sorting of teachers into schools (Lankford, Loeb & Wyckoff, 2002), I constrain this analysis to teachers within the same school by including school fixed effects in the statistical model that I use to estimate these relations. Equation [1] describes this model:

$$KISS_{cjs} = \beta_1 X_{cjs} + \gamma_s + \varepsilon_{cjs} \quad [1]$$

$KISS_{cjs}$ is the average KISS of teacher *c* in subject *j* in school *s*, X_{cjs} is a teacher or classroom characteristic, γ_s is a school fixed effect, and ε_{cjs} is a teacher-level error term.

The Predictors of Student-Level KISS

To examine linkages between student characteristics and teacher knowledge of individual students' skills, I compare students within the same classroom using teacher fixed effects models. Analyzing students in the same classroom enables me to control for the non-random sorting of students within schools. I focus my analysis on student characteristics that could plausibly affect teachers' KISS, including: whether or not the student is the same race as the teacher (virtually all preschool teachers are female and first grade teacher gender is suppressed in the ECLS-K public-use data file, rendering an analysis of the influence of shared gender on KISS of little value); age; gender; socioeconomic status; and fall reading and math test scores. Equation [2] describes the teacher fixed effects model that I use to examine the relations between these student characteristics and KISS:

$$KISS_{icj} = \beta_1 X_{icj} + \gamma_c + \varepsilon_{icj} \quad [2]$$

$KISS_{icj}$ is the KISS for student *i* of teacher *c* in subject *j*, X_{icj} is a student characteristic, γ_c is a teacher fixed effect, and ε_{icj} is a student-level error term.

The Instructional Practices of High-KISS Teachers

In the next step of my analysis, I examine the effects of KISS on teachers' instructional differentiation practices. I begin this examination by analyzing teachers in the same school. In particular, I use a school fixed effects model to estimate the relation between teachers' average KISS across reading and math and their general classroom organization, defined by the minutes per day they allocate to whole class, small group, individual, and child-selected activities. Equation [3] describes this school fixed effects model:

$$Y_{cs} = \beta_1 KISS - bar_{cs} + \gamma_s + \varepsilon_{cs} \quad [3]$$

Y_{cs} is a measure of the general classroom organization of teacher c in school s, $KISS-bar_{cs}$ is the teacher's average KISS across reading and math, γ_s is a school fixed effect, and ε_{cs} is a random teacher-level error term. Confining the analysis to teachers in the same school eliminates the possibility of bias introduced by school-level factors related to teachers' KISS and the extent to which they differentiate instruction, like school-wide interim assessment use.

While the previous analysis controls for school-level confounds, if an unobserved characteristic of teachers in the same school is correlated with teachers' KISS levels and instructional differentiation practices, such as general ability, estimates from school fixed effects models are biased. In the next analysis, I control for possible omitted teacher-level variables by comparing the KISS and instructional differentiation practices of individual teachers in reading to the same teachers' KISS and practices in math. The ECLS-K data enable me to create KISS measures for teachers in reading and math, and they contain information on teachers' use of achievement grouping in both subjects. This feature of the data allows me to estimate a teacher fixed effects model to analyze the effect of KISS on teachers' instructional differentiation practices. Equation [4] describes this model:

$$Y_{cj} = \beta_1 KISS_{cj} + \phi_j + \gamma_c + \varepsilon_{cj} \quad [4]$$

Y_{cj} is a measure of the how many minutes per week teacher c's students spend in achievement groups in subject j, $KISS_{cj}$ is a teacher's average KISS in subject j, ϕ_j is a subject fixed effect, γ_c is a teacher fixed effect, and ε_{cj} is a teacher-level error term. I include subject fixed effects in the model to account for the possibility that teachers differentially group and develop KISS by subject.

The Effect of KISS on Student Achievement

To estimate the effect of KISS on student achievement, I start by simply regressing students' spring ECLS-K direct cognitive assessment scores in reading and math on teachers' average KISS in reading and math (Model 1). Given the possibility that student characteristics impact teachers' KISS levels and student achievement, I add controls for whether or not the student is the same race as the teacher in Model 2, along with student gender, age, and socioeconomic status. In Model 3, I add students' fall test scores; however, making this addition could over-correct for initial achievement since fall assessments were administered as late as early December in kindergarten and as late as November in first grade. I add a vector of teacher characteristic to Model 4, given the potential of correlations between teachers' characteristics, KISS levels, and student achievement. This vector includes teachers' age, experience, educational background (including degrees and college coursework), and certifications. Since classroom characteristics such as class size and within-class variability in initial achievement could affect KISS and student learning, I add these variables to Model 5, which is described by Equation [5]:

$$Y_{icjt} = \beta_1 KISS_{icjt} + X_{icjt}\beta_2 + \beta_3 Y_{icjt-1} + C_{icjt}\beta_4 + \varepsilon_{icjt} \quad [5]$$

Y_{icjt} is the spring (time *t*) standardized ECLS-K direct cognitive assessment *T* score of student *i* of teacher *c* in subject *j*, $KISS_{icjt}$ is the average KISS of student *i's* teacher (*c*) in subject *j*, X_{icjt} is a vector of student *i*'s characteristics, Y_{icjt-1} is the fall (time *t-1*) standardized ECLS-K direct cognitive assessment *T* score of student *i*, C_{icjt} is a vector of teacher and classroom characteristics, and ε_{icjt} is a student-level error term. In all models, I cluster standard errors at the teacher-level given the likelihood that the errors of students in the same classroom are correlated.

While Models 1-5 describe the average and adjusted relations between KISS and student achievement, they do not account for the non-random sorting of teachers into schools, nor do they attend to possible unobservable differences between teachers in the same school related to KISS and student achievement. In the next set of analyses, I control for school- and teacher-level confounds by estimating the effect of KISS on student achievement using a series of teacher fixed effects models. In the first of these models, I regress students' subject-specific test scores on teachers' average subject-specific KISS (Model 1). I add subject fixed effects to Model 2, given the possibility that teachers generate higher KISS and improve student achievement more in one subject than the other. Finally, in Model 3, I include students' fall test scores. Equation [6] describes teacher fixed effects Model 3:

$$Y_{icjt} = \beta_1 KISS_{icjt} + \gamma_c + \phi_j + \beta_2 Y_{icjt-1} + \varepsilon_{icjt} \quad [6]$$

Y_{icjt} is the spring (time *t*) standardized ECLS-K direct cognitive assessment *T* score of student *i* of teacher *c* in subject *j*, $KISS_{icjt}$ is the average KISS of student *i's* teacher in subject *j*, γ_c is a teacher fixed effect , ϕ_j is a subject fixed effect, Y_{icjt-1} is the fall standardized ECLS-K direct cognitive assessment *T* score of student *i*, and ε_{icjt} is a student-level error term.

Estimates from within-teacher models tell us whether or not teachers with high average KISS levels are better at improving student achievement than other teachers. However, they do not control for average differences in how well teachers know particular students. For example, some teachers might have strong relationships (and KISS) with some students and draw on these relationships to improve the learning of these students (e.g., by motivating them). In such cases, one could not separate the effect of teacher-student relationships from that of KISS. In my next set of analyses, I control for differences in how well teachers know particular students by estimating the effect of KISS on student achievement using a series of student fixed effects models.

In the first model, I regress students' spring test scores on teacher's student-level KISS. I add subject fixed effects to Model 2. In my preferred specification (Model 3), I also control for baseline achievement. In this model, described by Equation [7], I compare a teacher's KISS in reading for a particular child and the child's spring reading test score with the teacher's KISS in math for the same child and the child's math score (and repeat the process for all children in the sample):

$$Y_{icjt} = \beta_1 KISS_{icjt} + \gamma_i + \phi_j + \beta_2 Y_{icjt-1} + \varepsilon_{icjt} \quad [7]$$

Y_{icjt} is the spring (time *t*) standardized ECLS-K direct cognitive assessment *T* score of student *i* of teacher *c* in subject *j*, $KISS_{icjt}$ is teacher *c*'s particular (student-level) KISS of student *i* in subject *j*, γ_i is a student fixed effect, ϕ_j is a subject fixed effect, Y_{icjt-1} is the fall standardized ECLS-K direct cognitive assessment *T* score of student *i*, and ε_{icjt} is a student-level error term.

Robustness Checks

Using within-student variation to estimate KISS achievement effects accounts for the strategic sorting of teachers into schools and students into classrooms, as well as average observed and unobserved differences in how well teachers know particular students. However, this study's results rely on particular assumptions, and plausible alternative explanations for the effect of KISS on student learning remain. In addition, different teachers have different student representation in the ECLS-K data, and this differential representation could influence results of teacher-level analyses. In what follows, I describe the methods I use to examine the robustness of study results to alternative assumptions, samples, and explanations.

To begin, I examine whether or not study results are sensitive to different methods of constructing KISS measures. In total, I create three alternative measures. The first measure is the most plausible alternative to the study's main measure but is based on slightly more generous ratings of students' skills. For example, it assumes that "Proficient" includes proficiency probabilities of ten percent or less, above the analogous probability of the main measure of five percent or less. The first alternative measure also has lower thresholds for reaching "In progress" and "Intermediate" status than the main measure: it assumes that that the lower bounds of these ratings are equivalent to proficiency probabilities of 41 and 61 percent, respectively, below the corresponding cut points of the main measure of 51 and 61 percent. The second alternative KISS measure is less plausible than the study's primary measure. For example, its proficiency probability band associated with "Intermediate" skills is 26 to 75 percent. The third alternative measure is an "anti-knowledge" measure. For instance, it assumes that teachers' "Not yet" rating is equivalent to proficiency probability scores of greater than 90 percent. This measure allows me to examine the effect of KISS on student achievement when teachers' ratings of students' skills are wildly inaccurate.

In the next set of specification checks, I revert back to employing the study's main measure. In these tests, I investigate the sensitivity of study results to a number of different restrictions to the analytic sample. To begin, I restrict the sample to only include classrooms with particular numbers of students represented in the ECLS-K data. I start by excluding classrooms below the 25th percentile in terms of the number of students present in the data. In the next two restrictions, I exclude classrooms below the median and 75th percentile in this dimension.

By controlling for fall test scores, many of this study's models go a long way towards addressing the possibility that it is easier for teachers to gain knowledge of students' skills in the subjects that students are stronger in at the time of fall assessment (and that students' relative subject strength, not KISS, is driving results). Nonetheless, I take a number of additional steps to analyze whether or not it's easier for teachers to obtain knowledge of students' skills in students' strong subject.

As a first step, I calculate the percentage of students who have higher baseline test scores in reading than math and for whom the teacher has higher reading than math KISS (and vice versa). In both kindergarten and first grade, about 55 percent of students fall into one of these two categories. That is, for only about 55 percent of the students in the sample do teachers have higher KISS in students' strong subject. For students' relative achievement to be the true driver of results, these percentages would likely have to be far away from 50 percent, the result that would occur by chance.

As a second step in assessing whether or not students' relative subject strength is the real driver of results, I re-estimate the fully-specified teacher and student fixed effects models, restricting the sample to exclude students whose relative strength in a given subject is likely to stand out. In particular, I remove students from the sample if they have relatively high baseline test scores in one subject, if the teacher has relatively high knowledge of them in the same subject, and if the difference between their strong-subject scores and the classroom average score in that subject is greater than the difference between their own strong- and weak-subject scores. For example, if the difference between a student's fall math score and the class average math score is 10 points and the difference between the student's math and reading scores is less than 10 points, I exclude the student from the sample, conditional on the student's teacher having higher knowledge of the student in math than reading (of note, only about 18 percent of all kindergartners and 12 percent of first graders with fall test scores meet these criteria).

The most likely remaining alternative explanation for the effect of KISS on student learning is that teachers who are more skilled at teaching a particular subject gain more KISS and improve student achievement to a greater degree in their strong subject as a strict function of their relative teaching strength. Consider a teacher with strong knowledge of the math curriculum but weak knowledge of the reading curriculum. If this teacher rates all students' spring math skills higher than their spring reading skills because of her strong curricular knowledge in math (e.g., the teacher might feel more optimistic about how students will perform in math in light of her mastery of the curriculum), and if the teacher's students perform better in math than reading because of the teachers' ability to deliver the math curriculum, the effect of KISS is confounded with the effect of curricular knowledge.

If teachers' subject-specific teaching skills are driving effects, it would almost have to be the case that teachers have better KISS in one subject for most if not all of the students in their classrooms (otherwise, subject-specific KISS would not be strongly correlated with subject-specific teaching strength). To examine whether or not teachers' subject-specific strengths are the true driver of effects, I re-estimate the fully-specified student and teacher fixed effects models, restricting the sample in a number of ways. First, I only include teachers who have better knowledge of at least one student in math and another student reading. Then, I only include teachers with relatively stronger knowledge of at least 30% but no greater than 70% of students in any subject. I ultimately restrict the sample to only include teachers who have stronger knowledge of half of their students in reading and stronger knowledge of half of their students in math.

Results

The Distribution of KISS across Schools

As Table 3 indicates, KISS in unequally distributed by student race and socioeconomic status. For instance, the average level of KISS of white first graders in math is 0.09 standard deviation units, well above the corresponding KISS of black and Hispanic students of -0.30 and -0.24 standard deviations, respectively (both significant at the 0.01 level). Similar trends are present in reading and in kindergarten. Moreover, the average level of KISS-Reading of kindergartners in the first socioeconomic status quintile is -0.22 standard deviations, well below average KISS of the wealthiest fifth of students of 0.20 standard deviations (significant at the 0.01 level). Similar patterns exist in math, and the disparities are even more salient in first grade.

Teachers' knowledge of individual students' skills is also unevenly allocated by schools' academic performance (see Table 4). As an example, the average KISS-Math of schools at or below the 10^{th} percentile in terms of first graders' average fall ECLS-K direct cognitive math assessment scores is -0.51 standard deviation units, far below the corresponding average of the highest-performing schools of 0.36 standard deviations (significant at the 0.01 level). Similar trends exist in reading and in kindergarten.

The Individual and Classroom Characteristics of High-KISS Teachers

This study finds no evidence of systematic differences in KISS among teachers in the same school based on observed background characteristics. The types and numbers of courses teachers completed in college hold no statistically-significant relation with teachers' average KISS, nor are there meaningful differences in the average KISS levels of teachers with and without master's degrees or particular types of teaching credentials. Likewise, teaching experience does not seem to be a precursor of KISS. Neither years of within-grade teaching experience nor total years of experience predict this type of teacher knowledge. Finally, class size and within-class student performance heterogeneity do not seem to inhibit or foster the development of KISS. Almost certainly, high-KISS teachers are different than other teachers in important ways. Unfortunately, I do not observe these differences in the ECLS-K data (results are available upon request).

The Predictors of Student-Level KISS

While there are no statistically-significant linkages between the observed characteristics and KISS of teachers in the same school, a number of student characteristics predict teachers' student-level KISS. As Table 5 demonstrates, teachers have particularly strong knowledge of the skills of the older, wealthier, and higher-achieving students in their classrooms. Somewhat similarly, in both kindergarten and first grade, teachers' KISS of female students is higher than that of male students in reading but lower in math. In addition, teacher-student race concordance positively affects teachers' understanding of particular students' skill levels. The most extreme instance of this relation pertains to first graders in math. If a student shares the same race as the teacher, the teacher's knowledge of the student's skills increases by almost 0.18 standard deviations (significant at the 0.01 level). Students' fall ECLS-K direct cognitive assessment scores exhibit the strongest within-class relation to student-level KISS. Across most measures, a standard deviation increase in a kindergartener's fall reading test score is associated with an increase in the teacher's knowledge of that student of around a half of a standard deviation (significant at the 0.01 level).

The Instructional Practices of High-KISS Teachers

Table 6 provides evidence that kindergarten teachers, but not first grade teachers, with high levels of KISS differentiate instruction more than other teachers. Panel A presents results from the school fixed effects model. Within schools, a standard deviation increase in teachers' average KISS across reading and math is associated with an increase in the number of minutes per day that teachers allocate to small group activities of nearly 0.07 standard deviations (significant at the 0.05 level). In addition, the within schools relation between KISS and the number of minutes dedicated to child-selected activities is positive and nearly significant at the 0.10 level (an effect size of about 0.05 standard deviations). Finally, while the estimates of the links between KISS and minutes per day in whole class and individual activities lack statistical significance, their signs are in the expected directions.

Panel B of Table 6 reports estimates from the teacher fixed effects model and provides even stronger evidence of a KISS effect on instructional differentiation in kindergarten. In particular, it indicates that a standard deviation in a teachers' average KISS in one subject, relative to the other subject, is related to an increase in the number of minutes per week students spend in achievement groups in the first subject, relative to second, of roughly 0.03 standard deviations (significant at the 0.05 level). For example, an increase in the difference between a teacher's KISS-Reading and KISS-Math of one standard deviation is associated with an increase in the difference between the minutes per day students spend in reading and math achievement groups of about 0.03 standard deviations. This result is compelling because it accounts for average differences in the extent to which different teachers use instructional groups, as well as average differences in grouping across subjects.

While the results in Table 6 are suggestive of a relationship between classroom practice and KISS, they are limited. The instructional differentiation measures are self-reported and only capture the intensity of differentiation, not the quality. Yet the quality of differentiation is likely more important than intensity, especially in first grade, where achievement grouping in reading is nearly universal (McPartland, Coldiron & Braddock, 1987). First grade teachers with different levels of KISS may differ with respect to the quality of their instructional differentiation.

The Effect of KISS on Student Achievement

Models 1-5 in Table 7 show a consistent positive effect of KISS on students' spring ECLS-K direct cognitive assessment scores. The results of Models 1 and 2 indicate that a standard deviation increase in kindergarten teachers' average KISS is associated with an increase in student achievement of 0.20-0.30 standard deviation units (significant at the 0.01 level). Estimates are even larger in first grade. However, Models 1 and 2 do not control for students' fall test scores, which as shown above, significantly predict KISS. Accounting for fall scores, Model 3 results show that a standard deviation increase in KISS is associated with an increase in kindergartners' spring test scores of roughly 0.08-0.10 standard deviation units (significant at the 0.01 level). First grade KISS effects do not decrease as much kindergarten effects when fall test scores are included in the model. This phenomenon could be due to the fact that fall scores are unavailable for roughly two thirds of the first grade sample. Adding teacher and classroom characteristics to the model has virtually no impact on effect estimates (see Models 4 and 5 in Table 7).

While Models 3-5 in Table 7 eliminate some potential sources of bias, they do not account for the non-random sorting of teachers into schools and students into classrooms on unobserved characteristics that are potentially correlated with both KISS and student learning. Table 8 reports estimates from teacher fixed effects models that control for this type of selection. Within classrooms, the effect of teachers' average KISS on student achievement is about 0.06 standard deviations (significant at the 0.01 level; see Model 1). Adding subject fixed effects does little to change estimates (see Model 2); however, when I include students' initial achievement to model, the effect of KISS on students' ECLS-K direct cognitive assessment scores decreases to around 0.03-0.05 standard deviations (see Model 3).

Table 9 presents student-level KISS achievement effect estimates based on student fixed effects models. Importantly, these models account for the fact that teachers likely have different average levels of knowledge of particular students. The unadjusted effect of a standard deviation increase in KISS on student test scores is roughly 0.10-0.11 standard deviation units (significant at the 0.01 level; see Model 1). Adding subject fixed effects has almost no impact on results (see Model 2). When I add fall test scores to the student fixed effects model, effect estimates only drop slightly to about 0.08-0.09 standard deviations (significant at the 0.01 level; see Model 3).

Results of Robustness Checks

Table 10 reports KISS effect estimates from the fully-specified teacher and student fixed effects models based on alternative KISS measures. KISS effect estimates based on the first alternative measure are very similar to those based on the primary measure. For example, the within-student estimate in first grade is about 0.10 standard deviations (see Model 1). More substantial differences exist between estimates based on the main measure and those based on the second and third alternative measures, which seem less plausible than the study's main measure and the first alternative (see Models 2 and 3). For example, the within-students effect of KISS on first graders' ECLS-K direct cognitive assessment scores based on Alternative KISS Measure 2 and Measure 3 (the "anti-knowledge" measure) are about 0.05 and -0.11 standard deviation units, respectively (both results are significant at the 0.01 level), well below the corresponding estimate based on the study's main measure of about 0.08 standard deviations. Results from this analysis provide additional support for the validity of the study's primary measure. Estimates based on the most plausible alternative to this study's main measure are very close to those based on the study's primary measure; however, effect estimates shrink as the measures become less plausible, and the "anti-knowledge" measure has a negative impact on student learning.

As depicted in Models 4-6 in Table 10, within-student KISS effect estimates modestly increase as the sample becomes more restricted in terms of the number of students sampled by ECLS-K. For instance, the within-students estimates based on kindergarten and first grade classrooms with at least 12 and eight sampled students are about 0.10 and 0.11 standard deviation units, respectively (both results are significant at the 0.01 level; see Model 6). These estimates are greater than the corresponding full-sample estimates of about 0.08 and 0.09 standard deviations.

As Model 7 in Table 10 illustrates, the KISS effect estimates in kindergarten based on restricting the sample to exclude students who really stand out in one subject are slightly larger than those based on the full sample. As an example, the within-students estimate is roughly 0.10 standard deviations (significant at the 0.01 level). On an important related note, the statistically-significant link between kindergartners' fall test scores and teachers' KISS levels disappears when I restrict the sample as described above. The effects in first grade are difficult to interpret given that they reflect a sub-sample sample of the sub-sample of the approximately 27 percent of students with fall test scores.

Table 10 also reports results based on samples that are restricted to include teachers who have better knowledge of at least one student in math and another student reading, teachers with relatively stronger knowledge of at least 30% but no greater than 70% of students in any subject, and teachers who have stronger knowledge of half of their students in reading and half in math. Estimates of the effect of teacher-level KISS in first grade increase as more classrooms are excluded; however, after the initial sample restriction, within-teacher estimates in kindergarten fail to reach significance at conventional levels (see Models 8-11). For kindergarten teachers, it could be the case that relative teaching strength in a particular subject is correlated with average KISS. Yet even among these teachers, teachers' knowledge of students' skills at the individual level remains important. For example, even in the most balanced kindergarten classrooms in terms of the distribution of KISS across subjects, teachers' knowledge of individual students' skills has a positive effect. A standard deviation increase in student-level KISS in these classes improves students' ECLS-K direct cognitive assessment scores by about 0.12 standard deviations (significant at the 0.01 level; see Model 11).

Discussion

Taking motivation from strong empirical evidence on the effectiveness of targeted instruction, this study introduces the concept of teacher knowledge of individual students' skills. I find that not all teachers have strong KISS. As is the case with other teaching skills and qualifications (Hill, Rowan & Ball, 2005; Lankford, Loeb & Wyckoff, 2002), KISS is unequally distributed across schools by student race, wealth, and academic performance. Furthermore, even teachers with high average KISS levels know some types of students better than others. For example, this study finds that teachers' race-driven perceptions of students (Dee, 2005) affect their development of KISS, as teachers have a greater awareness of the skills and understanding of students who share their same race than of other students.

The unequal distribution of KISS across schools and within classrooms is troubling in light of the fact that teacher knowledge of students' skills seems to be important. I find that kindergarten teachers with high levels of KISS target instruction to a greater degree than other teachers and that KISS has a positive effect on student learning of about 0.08-0.09 standard deviations in kindergarten and first grade. To put these gains into perspective, they are roughly 46 to 55 percent as large as the math achievement effect of reducing the size of kindergarten classes by eight students (Word et. al, 1990).

While enhancing teachers' KISS may ultimately prove to be more cost effective than class size reduction or other intensive interventions, the question of how to develop teachers' knowledge of individual students' skills remains unanswered. This study finds no statistically-significant links between teachers' educational background, certification, or teaching experience and KISS, though the available measures of teacher characteristics are clearly incomplete. Identifying the mechanisms through which teachers' develop KISS is important if we want to extend the benefits of targeted instruction to all children.

References

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal, 90,* 449–466.

Ball, D.L., Thames, M.H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, *59*(5), 389-407.

Borko, H., Eisenhart, M. Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23,* 194–222.

Brophy, J., & Good, T.L. (1984). Teacher Behavior and Student Achievement. Occasional Paper No. 73.

Carlson, D., Borman, G.D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378-398.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). *Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction*. The Elementary School Journal, 3-20.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2000). Cognitively Guided Instruction: A Research-Based Teacher Professional Development Program for Elementary School Mathematics. Research Report.

Carpenter, T.P., Fennema, E., Peterson, P.L., & Carey, D.A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, *19*(5), 385-401.

Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499-531.

Center, D.B., Deitz, S.M., & Kaufman, M.E. (1982). Student ability, task difficulty, and inappropriate classroom behavior: A study of children with behavior disorders. *Behavior Modification*, 6: 355-374.

Center for Research on Elementary and Middle Schools. (1987). *School structures and classroom practices in elementary, middle, and secondary schools.* Baltimore, Maryland: McPartland, J.M., Coldiron, J.R., & Braddock, J.H. II.

Clune, W.H., & White, P.A. (2008). *Policy effectiveness of interim assessments in Providence Public Schools* (WCER Working Paper No. 2008-10).

Connor, C.M., Morrison, F.J., & Katch, L.E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, *8*(4), 305-336.

Connor, C.M., Morrison, F.J., & Petrella, J.N. (2004). Effective Reading Comprehension Instruction: Examining Child x Instruction Interactions. *Journal of educational psychology*, *96*(4), 682.

Connor, C.M., Morrison, F.J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology*,*98*(4), 665.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *SCIENCE*, *315*(5811), 464-465.

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., ... & Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*(3), 189-221.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 158-165.

Domas, S.J., & Tiedeman, D.V. (1950). Teacher competence: an annotated bibliography. *Journal of Experimental Education, 19, 101-218*.

Doyle, W. (1977). 4: Paradigms for Research on Teacher Effectiveness. *Review of research in education*, *5*(1), 163-198.

Ehrenberg, R. G., & Brewer, D.J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review*, *13*(1), 1-17.

Ehrenberg, R.G., & Brewer, D.J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, *14*(1), 1-21.

Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R., Cahen, L.S., & Dishaw, M.M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In Denham, C. & Lieberman, A. (Eds.), *Time to Learn* (7-22). Washington, DC: U.S. Government Printing Office.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harv. J. on Legis.*, *28*, 465.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd, ed. *Holding Schools Accountable*. Washington, D.C.: Brookings.

Goldhaber, D. D., & Brewer, D. J. (1996). Evaluating the Effect of Teacher Degree Level on Educational Performance.

Good, T.L., Grouws, D.A., & Ebmeier, H. (1983). *Active Mathematics Teaching*. New York and London: Longman.

Goswami, U. (2004). Neuroscience and education. *British Journal of Educational Psychology*, *74*(1), 1-14.

Hanushek, E.A. (1992). The trade-off between child quantity and quality. *Journal of political economy*, *100*(1), 84-117.

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmarks assessments affect student achievement* (Issues & Answers Report, REL 2007 No. 039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2008). *A second follow-up year for measuring how benchmarks assessments affect student achievement* (Technical Brief, REL 2008 No. 002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Hill, H.C., Ball, D.L., & Schilling, S.G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 372-400.

Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11-30.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational evaluation and policy analysis*, *24*(1), 37-62.

Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, *77,* 247–271.

McCutchen, D., Abbott, R.D., Green, L.B., Beretvas, S.N., Cox, S., Potter, N.S., Quiroga, T., & Gray, A.L. (2002). Beginning Literacy Links Among Teacher Knowledge, Teacher Practice, and Student Learning. *Journal of learning disabilities*, *35*(1), 69-86.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of education review*, *13*(2), 125-145.

Olson, L. (2005, November 30). Benchmark assessments offer regular checkups on student achievement. *Education Week*. Accessed on March 29, 2013, from: http://www.edweek.org/ew/articles/2005/11/30/13benchmark.h25.html.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, *28*(3), 5-13.

Quint, J., Sepanik, S., & Smith, J. (2008). *Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) program in Boston elementary schools.* New York: MDRC.

Reardon, S.F., Cheadle, J.E., & Robinson, J.P. (2009). The effect of Catholic schooling on math and reading development in kindergarten through fifth grade. *Journal of Research on Educational Effectiveness*, 2: 45-87.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, *70*, 256-284.

Shepard, L.A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education, 85*(2), 246-257.

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

Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *The American Economic Review*, 639-652.

Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: ready to learn but further to go. *Developmental psychology*, *33*(4), 711.

Umbreit, J., Lane, K.L., & Dejud, C. (2004). Improving classroom behavior by modifying task difficulty: Effects of increasing the difficulty of too-easy tasks. *Journal of Positive Behavior Interventions*, 6: 13-20.

Vygotsky, L. S. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Wayne, A.J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89-122.

Word, E., Johnston, J., Bain, H. P., Fulton, B. D., Zaharias, J. B., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 class size study: Final summary report, 1985-1990.* Nashville: Tennessee State Department of Education.

Tables and Figures

Table 1

Sample summary statistics (population weighted)

	Kindergarten Sample		First Grade Sample	
Panel A: Students	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.50	-	0.50	-
Black or African-American, non-Hispanic	0.16	-	0.15	-
Hispanic	0.14	-	0.17	-
Asian	0.03	-	0.03	-
White, non-Hispanic	0.62	-	0.60	-
Age in months (at time of spring assessment)	74.51	4.16	87.13	4.22
Socioeconomic status quintile	3.18	1.37	3.16	1.38
Fall reading direct cognitive assessment *T-score*	50.50	9.86	51.38	9.16
Fall math direct cognitive assessment *T-score*	51.22	9.56	51.40	8.96
N	13,745	13,745	12,787	12,787
Panel B: Teachers and classrooms				
Female	0.98	-	N/A	-
Black or African-American, non-Hispanic	0.07	-	0.07	-
Hispanic	0.05	-	0.07	-
Asian	0.02	-	0.00	-
White, non-Hispanic	0.86	-	0.82	-
Age in years	41.19	9.93	40.45	10.93
Reading teaching methods courses completed in college	3.22	1.83	3.61	1.92
Math teaching methods courses completed in college	2.60	1.72	2.73	1.78
Master's degree or higher	0.31	-	0.37	-
Total years of teaching experience	15.03	4.58	12.90	3.92
Class size	19.75	4.99	20.59	3.00
Within-class standard deviation of fall reading *T-* scores	7.95	3.58	6.93	4.05
Within-class standard deviation of fall math *T-* scores	7.82	3.31	6.79	3.66
N	2,598	2,598	3,649	3,649

Notes. The fall kindergarten and first grade Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) direct child assessments were administered as late as early December 1998 and late November 1999, respectively. The ECLS-K limited fall first grade data collection to a subsample of 30 percent of the total schools in the overall sample, covering 27 percent of the base-year students eligible to be interviewed in year two. Fall first grade test scores and associated within-class standard deviations of scores are based on this subsample (samples sizes for the reading and math assessments are 3,705 and 3,747 students, respectively). All other first grade summary statistics reflect spring data, when all students and teachers have sample eligibility. Kindergarten teacher age is calculated by subtracting teacher year of birth from 1998, when the ECLS-K commenced (first grade teacher age is calculated by subtracting teacher year of birth from 1999). First grade teachers' gender is suppressed in the ECLS-K public-use data file. Teachers who took six or more college courses in a particular subject are treated as if they took six courses in that subject (answer options for the ECLS-K teacher survey question on college course include: zero, one, two, three, four, five, and six or more courses). Kindergarten and first grade student data are weighted by ECLS-K population weights C1CW0 and C4CW0, respectively. Kindergarten teacher data are weighted ECLS-K population weight B1TW0. Weights do not exist for teachers beyond kindergarten; the sample of first grade teachers only represents teachers of ECLS-K children eligible for the first grade survey.

Table 2

Summary of teacher knowledge of individual students' skills (KISS) measures (in standard deviations units)

	Percentile				
Panel A: Kindergarten sample	10th	Median	90th	Skewness	Kurtosis
Student-level KISS-Reading (*N*=13,745)	-1.23	0.37	1.19	-0.97	3.93
Student-level KISS-Math	-1.24	0.02	1.21	-0.49	2.90
Teachers' average KISS-Reading (*N*=2,598)	-1.23	0.10	1.12	-0.93	4.95
Teachers' average KISS-Math	-1.27	0.09	1.10	-0.66	3.81
Panel B: First grade sample					
Student-level KISS-Reading (*N*=12,787)	-1.17	-0.13	0.92	-0.20	2.88
Student-level KISS-Math	-1.40	0.20	1.28	-0.51	2.59
Teachers' average KISS-Reading (*N*=3,649)	-1.24	-0.07	1.13	-0.33	3.74
Teachers' average KISS-Math	-1.33	0.14	1.26	-0.55	3.08

Notes. Kindergarten and first grade student-level measures are weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weights C2CW0 and C4CW0, respectively. Kindergarten teacher-level measures are weighted ECLS-K population weight B2TW0. Weights do not exist for teachers beyond kindergarten; the sample of first grade teachers only represents teachers of ECLS-K children eligible for the first grade survey.

Table 3

Mean differences in teacher knowledge of individual students' skills (KISS) by student race and socioeconomic status (in standard deviation units)

	Student Race				
Panel A: Kindergarten sample	White	Black	Hispanic	Other Race	Asian
KISS-Reading	0.00	-0.03	-0.10***	-0.14**	0.27***
KISS-Math	0.07	-0.23***	-0.12***	-0.01	0.08
N	8,297	2,082	1,792	824	727
Panel B: First grade sample					
KISS-Reading	0.05	-0.11***	-0.20***	-0.14***	0.13
KISS-Math	0.09	-0.30***	-0.24***	-0.25***	0.15
N	7,458	1,790	1,995	704	820
	Student Socioeconomic Status Quintile				
Panel C: Kindergarten sample	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
KISS-Reading	-0.22	-0.13**	-0.06***	0.07***	0.20***
KISS-Math	-0.08	-0.11	-0.09	-0.01†	0.22***
N	1,812	2,559	2,764	2,943	3,221
Panel D: First grade sample					
KISS-Reading	-0.34	-0.15***	-0.03***	0.08***	0.27***
KISS-Math	-0.48	-0.21***	-0.03***	0.10***	0.35***
N	1,704	2,168	2,369	2,580	2,969

Notes. In the analysis based on student race, statistical significance refers to the difference between students of a given race and white students. In the analysis based on student socioeconomic (SES) status, statistical significance refers to the difference between students of a given SES quintile and students in quintile one. Estimates for kindergarten and first grade students are weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weights C2CW0 and C4CW0, respectively. Statistical significance levels: †nearly significant at the 0.10 level; *p<0.10; **p<0.05; ***p<0.01

Table 4

Mean differences in schools' average teacher knowledge of individual students' skills (KISS) by school-average initial achievement level (in standard deviation units)

	School Average Performance on Fall Direct Cognitive Assessment					
	10th percentile and below	11th to 25th percentile	26th to 50th percentile	51st to 75th percentile	76th to 90th percentile	90th percentile and above
KISS-Reading (Kindergarten)	-0.20	-0.18	-0.09	0.07***	0.12***	0.22***
N	91	136	227	228	137	90
KISS-Math (Kindergarten)	-0.15	-0.17	-0.17	-0.04	0.13***	0.43***
N	90	137	227	228	136	91
KISS-Reading (First Grade)	-0.39	-0.24	-0.09***	0.04***	0.12***	0.37***
N	42	65	107	107	65	42
KISS-Math (First Grade)	-0.51	-0.44	-0.18***	0.04***	0.21***	0.36***
N	43	65	108	108	65	43

Notes. The analyses of KISS-Reading are based on school-average performance on the fall reading assessment; the KISS-Math analyses are based on math test scores. Kindergarten estimates weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weight S2SAQW0. First grade estimates reflect data for the fall subsample of schools with test score data (approximately 30 percent of the total sample). Weights do not exists for schools in the first grade year; the sample of first grade schools thus represents the 30 percent subsample of schools with ECLS-K children eligible for the first grade survey. Statistical significance refers to the difference between schools in a given average performance range and schools at or below the 10th percentile: *p<0.10; **p<0.05; ***p<0.01

Table 5

The relation between teacher knowledge of individual students' skills (KISS) and student characteristics (in standard deviation units)

Student characteristics:	KISS-Reading (Kindergarten)	KISS-Math (Kindergarten)	KISS-Reading (First Grade)	KISS-Math (First Grade)
Same race as teacher	0.073***	0.067***	0.045*	0.175***
	(0.024)	(0.023)	(0.027)	(0.024)
Female	0.170***	-0.076***	0.105***	-0.025
	(0.017)	(0.016)	(0.019)	(0.017)
Age	0.025***	0.013***	0.004	0.018***
	(0.002)	(0.002)	(0.002)	(0.002)
Socioeconomic status quintile	0.124***	0.027***	0.074***	0.149***
	(0.008)	(0.008)	(0.009)	(0.008)
Fall reading direct cognitive assessment *z*-score	0.470***	0.139***	0.428***	0.511***
	(0.009)	(0.009)	(0.019)	(0.015)
Fall math direct cognitive assessment *z*-score	0.425***	0.203***	0.295***	0.544***
	(0.009)	(0.009)	(0.020)	(0.015)
N	13,745	13,745	12,787	12,787
Model inclusions:				
Teacher fixed effects	X	X	X	X

Notes. The Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) limited fall first grade data collection to a subsample of 30 percent of the total schools in the overall sample, covering 27 percent of the base-year students eligible to be interviewed in year two. For first graders, the relations between KISS and test scores are based on a subsample of these students with both reading and math test scores present (n=3,703). Kindergarten and first grade estimates are weighted by ECLS-K population weights C2CW0 and C4CW0, respectively. Statistical significance: *p<0.10; **p<0.05; ***p<0.01

Table 6

The effects of teachers' knowledge of individual students' skills (KISS) on their instructional differentiation practices (in standard deviation units)

Panel A: School fixed effects model estimates	Minutes per day in whole class activities	Minutes per day in small group activities	Minutes per day in individual activities	Minutes per day in child-selected activities
Kindergarten sample	-0.025	0.067**	0.028	0.052†
	(0.032)	(0.034)	(0.037)	(0.032)
N	2,489	2,471	2,391	2,468
First grade sample	-0.017	-0.012	0.007	0.009
	(0.027)	(0.027)	(0.029)	(0.026)
N	3,513	3,490	3,396	3,441
Model inclusions:				
School fixed effects	X	X	X	X

Panel B: Teacher fixed effects model estimates	Minutes per week in achievement groups
Kindergarten sample	0.028*
	(0.017)
N	2,185
First grade sample	0.004
	(0.018)
N	3,400
Model inclusions:	
Teacher fixed effects	X
Subject fixed effects	X

Notes. In the school fixed effects model, KISS reflects teachers' average KISS across reading and math. To create the minutes per day variables for this model, I use teachers' answers to the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) spring teacher survey question: "In a typical day, how much time do children spend in the following instructional activities?" Answer options include no time, half hour or less, about one hour, about two hours, and three hours or more, which I equate to zero, 30, 60, 120, and 180 minutes per day, respectively. To calculate the minutes per week that teachers' use achievement groups in reading and math (for the teacher fixed effects model), I multiply the frequency per week that teachers report using groups by the minutes per day that they report grouping on the days that they group. Frequency information comes from the ECLS-K spring teacher survey question: "How often do you divide your class(es) into achievement groups for reading and math activities or lessons?" Answer options include never, less than once a week, once or twice a week, three or four times a week, and daily, which I equate to zero, one half, one and a half, three and a half, and five times per week, respectively. Minutes per day in groups on grouping days comes from the ECLS-K spring teacher survey question: "On days when you use achievement grouping, how many groups do you have and how many minutes per day are your class(es) usually divided into achievement groups for reading and math activities or lessons?" Answer options include one to 15 minutes, 16-30 minutes, 31-60 minutes, and more than 60 minute per day, which I equate to seven and a half, 22.5, 45 and 60 minutes, respectively. Kindergarten estimates are weighted by ECLS-K population weight B2TW0. Weights do not exist for teachers beyond kindergarten; first grade estimates only represent teachers of ECLS-K children eligible for the first grade survey. Statistical significance: †nearly significant at 0.10 level; *p<0.10; **p<0.05; ***p<0.01

Table 7

Ordinary least squares estimates of the effect of teachers' average knowledge of individual students' skills (KISS) in reading and math on students' spring reading and math ECLS-K direct cognitive assessment scores (in standard deviation units)

	Model 1	Model 2	Model 3	Model 4	Model 5
<u>*Panel A: Kindergarten sample*</u>					
KISS-Reading	0.249***	0.201***	0.075***	0.074***	0.073***
	(0.015)	(0.013)	(0.009)	(0.009)	(0.009)
KISS-Math	0.296***	0.211***	0.098***	0.100***	0.101***
	(0.015)	(0.012)	(0.008)	(0.007)	(0.007)
N	13,745	13,745	13,745	13,745	13,745
<u>*Panel B: First grade sample*</u>					
KISS-Reading	0.403***	0.309***	0.251***	0.251***	0.252***
	(0.014)	(0.013)	(0.012)	(0.012)	(0.012)
KISS-Math	0.293***	0.196***	0.157***	0.154***	0.155***
	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)
N	12,787	12,787	12,787	12,787	12,787
<u>Model inclusions:</u>					
Student characteristics		X	X	X	X
Fall test scores			X	X	X
Teacher characteristics				X	X
Classroom characteristics					X

Notes. Fall test scores are only available for a subsample of first graders. In models controlling for initial achievement, missing fall test scores are set to zero and a dummy variable that equals one if a student is missing test score data and zero otherwise is included. Kindergarten and first grade Model 1 and 2 estimates are weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weights C2CW0 and C4CW0, respectively. Kindergarten and first grade Model 3-5 estimates are weighted by ECLS-K population weights BYCW0 and C34CW0, respectively. Standard errors clustered at the teacher level. Statistical significance: *p<0.10; **p<0.05; ***p<0.01

Table 8

Teacher fixed effects model estimates of the effect of teachers' average knowledge of individual students' skills (KISS) on students' spring ECLS-K direct cognitive assessment scores (in standard deviation units)

	Model 1	Model 2	Model 3
Kindergarten sample	0.055***	0.055***	0.034***
	(0.008)	(0.008)	(0.005)
N	13,745	13,745	13,745
First grade sample	0.063***	0.063***	0.047***
	(0.010)	(0.010)	(0.009)
N	12,787	12,787	12,787
Model inclusions:			
Teacher fixed effects	X	X	X
Subject fixed effects		X	X
Fall student test scores			X

Notes. First grade fall test scores are only available for students of a subsample of teachers. Model 3 for the first grade sample includes these students as well as first graders without fall test scores and thus only partially controls for initial achievement. Kindergarten and first grade Model 1 and 2 estimates are weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weights C2CW0 and C4CW0, respectively. Kindergarten and first grade Model 3 estimates are weighted by ECLS-K population weights BYCW0 and C34CW0, respectively. Statistical significance: *p<0.10; **p<0.05; ***p<0.01

Table 9

Student fixed effects model estimates of the effect of teacher knowledge of individual students' skills (KISS) on students' spring ECLS-K direct cognitive assessment scores (in standard deviation units)

	Model 1	Model 2	Model 3
Kindergarten sample	0.115***	0.115***	0.094***
	(0.004)	(0.004)	(0.004)
N	13,745	13,745	13,745
First grade sample	0.094***	0.094***	0.078***
	(0.006)	(0.006)	(0.006)
N	12,787	12,787	12,787
<u>Model inclusions:</u>			
Student fixed effects	X	X	X
Subject fixed effects		X	X
Fall test scores			X

Notes. Fall test scores are only available for a subsample of first graders. Model 3 for the first grade sample includes these students as well as first graders without fall test scores and thus only partially controls for initial achievement. Kindergarten and first grade Model 1 and 2 estimates are weighted by Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) population weights C2CW0 and C4CW0, respectively. Kindergarten and first grade Model 3 estimates are weighted by ECLS-K population weights BYCW0 and C34CW0, respectively. Statistical significance: *p<0.10; **p<0.05; ***p<0.01

Table 10

The effects of teacher knowledge of individual students' skills (KISS) on students' spring ECLS-K direct cognitive assessment scores under alternative specifications (in standard deviation units)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Panel A: Kindergarten sample											
Teacher fixed effects model estimates	0.043***	0.033***	-0.049***	0.032***	0.031***	0.032***	0.044***	0.033***	0.021	0.036	0.035
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.011)	(0.005)	(0.007)	(0.014)	(0.024)	(0.033)
Student fixed effects model estimates	0.106***	0.096***	-0.100***	0.094***	0.096***	0.102***	0.101***	0.104***	0.110***	0.118***	0.116***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.007)	(0.004)	(0.004)	(0.006)	(0.009)	(0.013)
N	13,745	13,745	13,745	11,470	7,548	3,454	11,282	11,406	6,329	3,161	1,344
Panel B: First grade sample											
Teacher fixed effects model estimates	0.062***	0.032***	-0.060***	0.052***	0.053***	0.046**	0.018	0.060***	0.056**	0.125***	0.124**
	(0.010)	(0.009)	(0.011)	(0.011)	(0.014)	(0.023)	(0.013)	(0.016)	(0.028)	(0.047)	(0.059)
Student fixed effects model estimates	0.101***	0.054***	-0.109***	0.085***	0.099***	0.107***	0.053***	0.091***	0.085***	0.092***	0.067***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.011)	(0.010)	(0.007)	(0.010)	(0.014)	(0.020)
N	12,787	12,787	12,787	10,446	7,099	3,384	3,165	8,998	5,445	2,711	1,366
KISS measure:											
Study's main KISS measure				X	X	X	X	X	X	X	X
Alternative KISS Measure 1	X										
Alternative KISS Measure 2		X									
Alternative KISS Measure 3			X								
Analytic sample:											
Study's main sample	X	X	X								
Alternative Sample 1				X							
Alternative Sample 2					X						
Alternative Sample 3						X					
Alternative Sample 4							X				
Alternative Sample 5								X			
Alternative Sample 6									X		
Alternative Sample 7										X	
Alternative Sample 8											X
Model inclusions:											
Teacher or student fixed effects	X	X	X	X	X	X	X	X	X	X	X
Subject fixed effects	X	X	X	X	X	X	X	X	X	X	X
Fall test scores	X	X	X	X	X	X	X	X	X	X	X

Notes. Please refer to the text for descriptions of alternative KISS measures. Alternative analytic samples 1-3 include teachers at or above the 25th percentile, the median, and the 75th percentile in terms of the number of students in their classrooms sampled by the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), respectively (four students in kindergarten and three in first grade, seven students in kindergarten and five in first grade, and 12 students in kindergarten and eight in first grade, respectively). Alternative analytic sample 4 excludes students whose strength in one subject relative to the class is greater than their own relative strength in that subject. The 5th alternative analytic sample includes teachers who have better knowledge of at least one student in math and another student reading. Alternative analytic samples 6-8 includes teachers with relatively stronger knowledge of at least 30% but no greater than 70% of student in any subject (a “knowledge mix” of 30-70 or better), a knowledge mix of 40-60 or better, and a 50-50 knowledge mix, respectively. Fall test scores are only available for a subsample of first graders. The models for the first grade sample include these students as well as first graders without fall test scores and thus only partially control for initial achievement. One exception is Model 7, which only includes first graders with fall test scores. Kindergarten and first grade estimates are weighted by ECLS-K population weights BYCW0 and C34CW0, respectively. Statistical significance: *p<0.10; **p<0.05; ***p<0.01