WORKING
PAPER

Measuring the Practices, Philosophies, and Characteristics of Kindergarten Teachers

LAURA S. HAMILTON AND
CASSANDRA M. GUARINO

WR-199-EDU

November 2004

Prepared for the National Center for Education Statistics

RAND EDUCATION

Table of Contents

List of Tables and Figures

Tables

Figures

Acknowledgments

This work was funded by the National Center for Education Statistics. We are grateful to Jerry West and Amy Rathbun for comments on an earlier draft.

I. Introduction

Most parents believe strongly that the quality of their children's teachers is one of the most important determinants of student learning. There is growing empirical evidence that teachers do in fact have strong effects on achievement (Rivkin, Hanushek, & Kain, 1998; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997), but efforts to identify specific characteristics that make a difference have come up short. Most studies that have examined available indicators of teacher preparation or quality, such as certification status and experience, find that the effects of these indicators are either null or very small (Brewer & Goldhaber, 1996; Ferguson, 1991; Hanushek & Pace, 1995; Miller, McKenna, & McKenna, 1998). Together these results suggest that there are characteristics of teachers that influence achievement but that we have not yet figured out what they are.

The Early Childhood Longitudinal Study, Kindergarten Class 1998-99 (ECLS-K), sponsored by the U.S. Department of Education, National Center for Education Statistics, provides an opportunity to extend the existing research on teachers in two significant ways. First, it enables some of the studies of certification, instructional practices, and other influences on achievement to be extended to early elementary school students, a group that has not been the focus of most of this research. Second, it allows researchers to study jointly some issues that are typically addressed separately. For example, most of the literature on teacher certification and other teacher background characteristics fails to tie those characteristics to what teachers do in the classroom. The ECLS-K data permit the exploration of complex models of teachers' effects on student achievement.

In this report we describe exploratory analyses of the ECLS-K data on teachers and teaching. The purpose of these analyses was to create measures of various constructs related to teaching and to explore the possible utility of these measures for future research. Although our explorations are not exhaustive, we examine several categories of variables that would be expected to influence student learning. These include instructional practices, teaching philosophies, school climate, and resources.

The remainder of this report is organized into four sections. In the next section, we briefly discuss the rationale for including each of the categories of constructs. Section III provides a description of the data used in this study. In Section IV, we present the results of the factor analyses and scale construction. This section describes our procedures for constructing scales to measure aspects of instructional practices, philosophy, climate, and resources. The discussion of each set of scales is supported with a brief discussion of the role of those constructs in a conceptual model of teaching effects. Finally, Section V provides a brief summary and a discussion of directions for future research.

II. Background on Teaching Constructs

In this section, we discuss prior research that forms the basis for our rationale for choosing the particular characteristics and constructs explored in this report. We begin by discussing studies relating teacher characteristics and training to instructional quality and student achievement. Next, we discuss background information pertaining to instructional practices and teaching philosophies. Finally, we discuss issues relating to school climate and resources. This section is not intended to provide a comprehensive literature review, but to highlight earlier work that is relevant to the constructs we are examining in this study.

Teacher Characteristics and Pre-service Training

Researchers have attempted to find links between student achievement and the characteristics and training of teachers. Some studies have suggested that background characteristics unrelated to teacher training, such as verbal ability (Ehrenberg and Brewer, 1995) or the selectivity of a teacher's undergraduate institutions (Ehrenberg and Brewer, 1994), are positively correlated with student achievement. Prior research focusing on the impact of credentials and pre-service training on the quality of instruction, however, has found mixed results regarding the impact of the teaching credential and small positive effects with regard to subject area preparation.

A few studies have investigated the effect of a credential on student achievement, though not at the kindergarten level. Its importance has been examined and debated in studies of older students, particularly in relation to learning in the fields of science and mathematics. For example, Hawk, Coble, and Swanson (1985) found that studying with teachers who are fully certified in mathematics instruction tends to raise a student's achievement in that subject. Fetler (1999) found a negative correlation between math scores and the percent of teachers with emergency credentials at the school level. In an analysis using individual student data, Goldhaber and Brewer (2000) found that the performance of high school students on standardized math and science tests did not differ according to whether their teachers held a standard or an emergency credential. They found, however, that students of teachers who were uncertified or who held a private school certification had lower achievement levels than students of teachers with a standard, probationary, or emergency certification in math.

The teaching credential is not a standard measure of preparation, however, since requirements for the credential vary widely from state to state. Thus the type of credential a teacher holds is only a rough proxy for training. Due to the scarcity of data, few studies have been able to assess on a large scale the impact of the amount or type of pedagogical or subject area preparation a teacher might obtain on the learning they impart to students. With regard to subject-matter preparation, Goldhaber and Brewer (2000) found that students of teachers who were certified out of field in math performed worse on standardized tests than students whose teachers had standard credentials. Monk (1994) found a small effect of teacher coursework in math and science on student test scores.

The ECLS-K furnishes information on the credential status of teachers, the types of certifications they hold, and the amount of training they have received in methods of teaching various subject areas, such as reading, math, and science. Therefore, these characteristics and measures are included in our study and analyzed with respect to teacher performance on the various instructional scales we develop.

Instructional Practices

The classroom is an important missing link in much of the research cited in the previous section. Understanding the mechanism through which teacher characteristics influence student achievement requires information about what those teachers do in the classroom. Although a complete and accurate picture of a teacher's classroom practices generally requires labor-intensive data collection methods such as classroom observations, it is possible to obtain some information about instruction through paper-and-pencil questionnaires. This method of data collection limits the kinds of constructs that can be measured; it is probably most effective for obtaining information about the frequency with which teachers engage in clearly defined practices.

There are several examples of the use of this type of measure in the literature, and some evidence that teacher reports of particular types of classroom practices are correlated with student achievement. In particular, a number of studies have examined so-called "reform-based" or "standards-based" instructional practices in mathematics or science; these are practices that are consistent with the kinds of instruction promoted by professional organizations such as the National Council of Teachers of Mathematics and the National Research Council. Cohen and Hill (1998) studied teacher-reported use of reform-based mathematics instructional practices, and found that frequency of use was positively related to scores on the California Learning Assessment System (CLAS) mathematics test at the school level after controlling for demographic characteristics. Klein et al. (2000) examined standards-based mathematics and science practices in several states and districts and found small, positive relationships between students' exposure to these practices and growth in achievement on both multiple-choice and open-response mathematics and science tests. Mayer (1998) found small positive or null relationships between a similar set of practices and student scores on a standardized multiple-choice test in mathematics. Several studies have attempted to distinguish between reform-based practices (e.g., use of cooperative groups, open-ended assessment techniques, inquiry-based instruction) and more traditional approaches (e.g., use of textbooks, lectures, and multiple-choice tests; see Cohen & Hill, 1998; Klein et al., 2000; Smerdon, Burkam, & Lee, 1999). All of these studies used paper-and-pencil questionnaire items similar to those used in the ECLS-K.

Most of this research has focused on upper elementary and secondary students, and relatively few studies of instructional practices have examined nationally representative samples of students or teachers. Given the increased attention to early development of literacy and numeric skills, and the growing (though certainly not complete) professional consensus on how these subjects should be taught, there is a clear need for studies of relationships between instructional practices and student learning in the early elementary grades. The ECLS-K data allow us to begin to fill gaps in the existing literature.

Teaching Philosophies

An accurate understanding of what occurs in the classroom is critical for assessing a teacher's likely impact on student outcomes. For some constructs, however, it may be adequate or even desirable to obtain teachers' views on what they *should* do in the classroom—in other words, to understand the philosophy of teaching that guides their actions. This may be especially true in cases where it is difficult to create questionnaire items that capture a particular type of behavior or tendency.

We examine two sets of measures related to teaching philosophies. The first elicits teachers' views on the relative importance of various student characteristics in determining whether children are ready for kindergarten. In addition to providing information about what teachers view as important, these data may be useful for understanding how teachers allocate their time and efforts in the classroom and the specific ways in which they engage parents. The second set addresses teachers' use of various criteria for evaluating student progress. The use of multiple assessment methods is widely recommended for all students, but perhaps particularly for young children (see, e.g., National Association for the Education of Young Children, 1990). Schools are expected to promote a range of outcomes beyond improvement on standardized test scores, and obtaining information about the relative weights that teachers assign to various outcomes such as social skills is an important first step in exploring the ways in which schools and teachers influence these outcomes. Our analyses of these items builds on work by Rathbun, Walston, and Germino Hausken (2000).

These two sets of items obviously do not provide complete information on a teacher's philosophy of instruction. They address important aspects of it, however, and may prove useful in exploring the various paths by which teachers affect student learning. If examination of these constructs turns out to be fruitful, this would suggest the value in a fuller examination of teaching philosophies in future large-scale survey research.

School Climate and Teacher Satisfaction

There is growing recognition that successful implementation of school reforms depends on effective governance and a positive, professional school climate. Some factors that are especially important include strong leadership from the principal (Berends et al., 2001), opportunities for teachers to collaborate (Newmann et al., 1996), and a governance structure that includes teachers and other local actors in decisions affecting school policies (Bryk et al., 1998). A particularly striking finding in recent research is the importance of teachers' perceptions of the principal's leadership (as measured through questionnaires administered to teachers) in predicting quality of school reform implementation and student achievement (e.g., Berends et al., 2001).

We examined several ECLS-K questionnaire items that elicited teachers' opinions about school climate, including leadership. We also looked at three questions that addressed teachers' overall job satisfaction, a construct that is undoubtedly related to school climate and that may affect student achievement either directly or indirectly (e.g., through effects on teacher retention).

Resources

The importance of school resources has been the focus of vigorous debate for several decades (see, e.g., Hanushek, 1986; Hedges, Laine, & Greenwald, 1994). More recent research has attempted to move beyond the question of whether resources matter to a clearer understanding of which resources matter under what circumstances (Wenglinsky, 1997). Although "resources" can be broadly conceptualized as all inputs to the educational process, including teacher and administrative labor, as well as supplies, facilities, etc., this study uses the term in the narrower sense to refer to physical non-labor inputs. The investigation of resources in this study focuses rather narrowly on teachers' reports of the adequacy or availability of various instructional tools and facilities. Teachers report on whether or not they use and how adequately they are supplied with resources such as textbooks, manipulatives (e.g., books and puzzles), computer equipment, musical instruments, classroom space and heating, etc.

III. Data

The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 was undertaken by the National Center for Education Statistics to collect data on a nationally representative sample of approximately 22,000 children attending kindergarten in 1998-99. The study administered several tests to the kindergarteners near the beginning and end of the kindergarten year, and surveyed their parents, teachers, and school administrators. Data were collected in both the fall and the spring of the 1998-99 school year. The ECLS-K has continued to follow the children into the first grade, and future waves are planned for the third and fifth grades. This report uses data from the kindergarten year only.

The ECLS-K employed a multi-stage probability sample design in selecting a nationally representative sample of children attending kindergarten in 1998-99. The primary sampling units consisted of counties or groups of counties. Within these, the second stage units were schools. 1,277 public and private schools with kindergarten programs were drawn. Then a target sample of approximately 24 children per school was selected. The ECLS-K data were weighted to compensate for differential probabilities of selection at each sampling stage and to adjust for the effects of non-response (U.S. Department of Education, 2001).

The ECLS-K data provide several different sampling weights formulated by NCES, each adapted to different subsamples of the data. Some weights were adapted to the child sample, others to the teacher sample, and others to the school sample. Within these sample categories, there were weights indicated for cross-sectional analysis and weights indicated for longitudinal analysis. For the purposes of a study of self-reported teaching practices and philosophies, the best choice was a weight specifically designed for use with data derived from the teacher questionnaires, both fall and spring. The weight indicated in this situation was B2TW0. This weight was used to generate all analyses presented in this report.

This study drew primarily upon data supplied through the fall 1998 and spring 1999 teacher questionnaires. These were paper-and-pencil surveys designed to elicit information regarding the instructional practices, teaching philosophies, satisfaction levels, perceptions of school organizational characteristics, demographics, and background characteristics of each teacher. A total of 3,305 teachers completed surveys in either spring or fall. Of these, 3,047 were present in the fall, 3,243 were present in the spring (new teachers were added to the spring sample), and 2,985 were present in both spring and fall. The variables we used to create our instructional practice, teaching philosophy, climate, and satisfaction scales are described in detail in the section dealing with the construction of the scales.

IV. Construction of Scales

This section describes our exploratory analyses of selected sets of items from the ECLS-K questionnaires, and our approach to constructing scales from these items. After describing the items, we discuss several common methods of scale construction, describe the approach we used, and present the results of those analyses. This section is divided into four parts. The first covers the items addressing instructional activities and skills emphasized by teachers, which were the main focus of our study. The sections following it describe our analyses of items addressing teaching philosophies, school climate and teacher satisfaction, and school resources. These latter three sections are much briefer than the first because we provide fewer methodological details, the nature of the analysis is less complex, and there are fewer items involved.

Instructional Activities and Curricular Focus

In this section we describe the scales we derived from the questionnaire items on instructional activities and the skills teachers emphasized in their instruction. The factor analytic methodology described in this section is applicable to the analyses of other constructs, presented later.

Description of Questionnaire Items. This portion of the study focused on the set of questionnaire items that addressed instructional activities and skills. All of these appeared in the spring questionnaire. The stem for items 28a through 28w asked teachers "How often do children in this class do each of the following READING and LANGUAGE ARTS activities?" Teachers were asked to rate the frequency of each of 23 activities, using a scale that ranged from 1 ("never") to 6 ("daily"). Intermediate scale points were "once a month or less," "two or three times a month," "once or twice a week," and "three or four times a week." Items 29a through 29s, which focused on skills rather than activities, read as follows: "For this school year as a whole, please indicate how each of the following READING and LANGUAGE ARTS skills is taught in your class(es)." For each of 19 skills, teachers could indicate that the skill is not taught, by selecting either "taught at a higher grade level" or "children should already know", or could indicate the frequency with which the skill was taught, using a five-point scale that ranged from "one a month or less" to "daily." Intermediate scale points are the same as for items 28a through 28w.

The math items were structured the same way. Items 31a through 31q asked "How often do children in this class do each of the following MATH activities?" for each of 17 items, and items 32a through 32cc asked "For this school year as a whole, please indicate how each of the following MATH skills is taught in your class(es)" for each of 29 items. The response scales were the same as those for the corresponding reading and language arts items.

The questionnaire contained one set of items addressing science and social studies. The stem asked "For this school year as a whole, please indicate how each of the following SCIENCE or SOCIAL STUDIES topics or skills is taught in your class(es)." Teachers reported on 22 skills and topics. The computer use items asked teachers "How often do children in your class(es) use computers for the following purposes?" for each of nine purposes.

As discussed above, for the items that addressed skills taught in reading and math, teachers could report that a particular skill was not taught by choosing either "taught at a higher grade level" or "children should already know." For the analyses discussed in this report, which focus on the frequency with which teachers emphasize particular skills or topics, we combined these two options into a single "not taught" category. We then recoded the numerical values for all of the options so that the scale for the "skills" items was similar to that for the "activities" items (i.e., the category representing the greatest frequency received a score of 6, and the "not taught" category received a 1). This enabled us to combine the two sets of items into a single factor analysis. Although the difference between the two "not taught" options may be of substantive interest for some purposes, because our focus was on the amount of exposure students received, combining these categories was appropriate. Therefore the following descriptive information (i.e., item means) and factor analyses are based on sets of items with response options ranging from 1 to 6.

Teachers who were missing data on one or more items were excluded from the initial factor analyses. A total of 3243 teachers were administered the spring teacher questionnaires. The reading factor analysis was conducted on the 2323 teachers who had complete data on all items. The math analysis included 2287 teachers, science included 2451, social studies included 2719, and computer use included 2928.

Several cautions should be observed when interpreting information from these sets of items. The response scales elicit information about frequency, and do not address variation in intensity of instruction or in total time spent in a given day. While the format of these items is reasonable and consistent with other surveys of curriculum and instructional practice, users of the data should keep in mind that some important differences among teachers may not be captured by these measures. In particular, they do not address the quality of instruction. Moreover, because the instructional practice items are only administered in the spring, our data do not capture changes in instructional focus that may take place over the course of the year. Finally, these are self-report data, subject to a variety of possible response biases. On this last point, however, there is evidence that teacher reports of the frequency of instructional practices correspond reasonably well with other sources of evidence of what occurs in classrooms (see, e.g., Mayer, 1999).

<u>Factor Analytic Methods.</u>[1] A common approach for understanding the structure in a set of variables is the use of exploratory factor analytic methods. These methods, which include principal components analysis and factor analysis, are useful for summarizing the correlations among a large set of items and reducing this set to a smaller number of variables, called factors (or components). These factors, which are linear combinations of the observed variables, tend to be much more reliable than the individual items, and can be used for subsequent analysis of relationships among various constructs. There are two major approaches to factor analysis: exploratory, in which the researchers seeks to consolidate and understand a set of variables, and confirmatory, in which the researcher sets out to test a specific theory about relationships among variables. This latter approach is typically implemented through structural equation modeling

[1] The methodological information provided in this section, including the use of cross-validation samples, applies to all of the factor analyses discussed in this and subsequent sections.

(SEM) techniques, using variables that were carefully selected to represent a particular set of constructs. In this report we use the exploratory approach. Although we had some hypotheses regarding how some of the items would function, we did not have a specific theory that drew together the large number of items on the ECLS-K questionnaires, and we were more interested in learning about the ways in which the various instructional emphases and activities tended to cluster together. The scales that we describe below, and the relationships among them, could eventually be examined through a confirmatory factor analysis approach.

To provide a way to test the stability of the solutions, we conducted each factor analysis on a random half-sample of the data and used this to carry out a cross-validation study on each set of items. Before conducting each factor analysis, we randomly split the data into two subsets with an approximately equal number of observations in each. The initial analysis, including all of the exploratory steps, was conducted on one of these halves. Once a reasonable solution was identified, the factor analysis was repeated using the second half of the data. We then examined the patterns of loadings and determined the number of items for which the highest loading occurred on a different factor than in the original dataset. The results are discussed below. In all cases, this number was small, and limited to those items that had relatively weak loadings on all factors in the original analysis. The cross-validation provided evidence regarding the stability of each solution.

All analyses used principal components analysis with orthogonal (varimax) rotation. We explored other factor extraction methods and found that most of the results were relatively robust to choice of method[2]. This finding is consistent with other work that has compared results from different extraction methods (Thompson & Daniel, 1996). Because we expected some of the factors to be correlated, we also examined solutions using various oblique rotations. Conclusions about which items clustered together were consistent with those from the orthogonal rotation, a finding that is also consistent with other research (e.g., Choi, Fuqua, & Griffin, 2001). We focus on the orthogonal solutions to simplify reporting of results. To decide on the number of factors to retain, we initially specified the traditional criterion in which the number of factors retained is equivalent to the number of eigenvalues of the correlation matrix that are greater than one. However, this typically resulted in a larger number of factors than we thought would be useful or informative, so in each case we explored solutions in which fewer factors were retained. We also examined screen plots for each analysis. We based our final choice of solution on these various criteria, with the overarching goal of producing scales that would be informative for future analyses of the ECLS-K data. All analyses were weighted using the B2TWO weights.

Exploratory factor analysis, more so than many other data analytic techniques, is often described as involving a combination of art and science. There are no clear criteria for determining whether a particular solution is the right one. Instead, interpretability is one of the key tests of any factor analysis solution, and it is typically necessary to examine multiple solutions and identify the one that not only has empirical support but that makes substantive sense (Tabachnick & Fidell, 1996). There are often multiple solutions that are reasonable, rather than a single best solution. In the analyses described below we tried to identify scales that would be useful for future studies of teachers and teaching in kindergarten, but the readers should keep in mind that

[2] There were some exceptions in the reading and language arts analysis, discussed below.

there are undoubtedly alternative approaches to combining the questionnaire items, some of which may be preferred over ours for some purposes.

Methods for Creating Scales. After identifying a satisfactory factor solution, we created scales for each teacher to represent the resulting set of factors. A number of methods may be used to combine information from a set of items into a single scale. The most common method for combining items that use Likert scales is to add the scores on the individual items, a process that is typically referred to as summative scoring (or sum scoring). When the resulting scales are based on a factor analysis, the standard approach is to identify the factor on which each item has its highest loading, and include that item in the calculation of a score for that factor. Some researchers have expressed concern that computing a sum is inappropriate for items whose scales are most accurately classified as ordinal rather than interval (Stevens, 1946; Coste et al., 1995). However, others have suggested that meaningful information may in fact be derived from methods that use sums and averages of ordinal scales (Velleman & Wilkinson, 1993). The use of sum scores is widespread, and has the advantage of producing scores that are easily described in terms of the variables from which they are constructed. In addition, most recent studies of instructional practices use sum scores to combine responses to questionnaire items (see, e.g., Cohen & Hill, 1998; Hamilton et al., 2001; Klein et al., 2000; Weiss et al., 1998). For these reasons, we rely on sum scores for most of this report. Because some scales are constructed from more items than others, we use an average rather than a sum for each teacher[3]. For teachers who answered some but not all items on a given scale, we simply used the average score for those items that he or she answered. Given additional time and resources it would be worth exploring alternative imputation approaches to dealing with missing data.

Another approach to scale construction involves the use of factor scores rather than simple sum scores. For example, linear regression can be applied so that each variable is weighted according to its factor loadings. As a result, variables with higher loadings contribute more strongly to the factor score than do variables with low loadings. This method results in scores that are more closely related to the underlying factors than are simple sum scores, but that may be less interpretable because they are linear combinations of all the items included in the factor analysis rather than just those items that are representative of the construct of interest. Factor-weighted sums have been used in some studies of instructional approaches (e.g., Smerdon, Lee, & Burkam, 1999). Because of the relatively large numbers of items included in the factor analyses we report here, and because our purpose was to develop scales that were easy to describe to others and that represent coherent sets of instructional activities or emphases, we decided to use the simple sum score. Additional research to compare these two approaches in this context might be informative.

In recent years there has been increasing interest in the application of item response theory (IRT) models to Likert-type questionnaire items. These models were originally developed to be applied to test-score data (Lord & Novick, 1968). In IRT, an individual's standing on a trait, such as verbal ability, is estimated from his or her pattern of responses to a set of items rather than simply from a raw total score. The relationship between the individual's trait score and his or her response to an item is described by an item characteristic curve that typically has

[3] For simplicity and to maintain consistency with other studies on scaling methods, we continue to refer to these scores as "sum scores" throughout the paper.

a monotonically increasing S shape. Although more commonly known for its applications to dichotomous items, IRT is increasingly being used to model responses to items that have more than two response categories, including personality traits and attitudes (Embretson & Reise, 2000). IRT models offer several improvements over traditional methods for summarizing information on a set of items. In particular, in IRT the standard error of measurement of a scale is not a single quantity, but varies across scores on that scale, reflecting the differences in measurement precision across scores. IRT also allows for the possibility that individuals with the same sum score but different score profiles on the set of items may actually differ on the underlying trait. IRT scores may be especially useful for measuring change in cases where individuals start out at different points along the score distribution, since they are believed to have better equal-interval properties than most sum scores.

To illustrate, consider a scale constructed from three items, each with a five-point (1-5) Likert response scale. Sum scores could take on integer values from 3 to 15, so 13 different scores are possible. A respondent who chose "1" for the first two items and "2" for the third would receive the same score as someone who chose "2" for the first item and "1" for the second and third. In IRT scoring, these two response patterns may receive different scores, depending on the characteristics of the items (e.g., their levels of "difficulty," or in the case of the instructional practice items we are examining here, their frequency distributions). In this three-item example, there are 125 unique response patterns. IRT therefore allows for the possibility of 125 different scores, offering the potential for greater measurement precision than is available with sum scoring.

As an exploratory step, we computed IRT scores for all of the reading and language arts scales described below. We used Samejima's (1969) graded response model. This enabled us to compare distributions on the traditional sum scores with the IRT score distributions, and to examine the strength of the relationships between the two score types. For most of the report, however, we rely on traditional sum scores. Because our exploratory analyses were intended to provide scales that others could use in research with the ECLS-K, we believed that the relatively simple sum score provided an easier way to communicate the meaning of each scale. And because the correlations between corresponding IRT and sum scores were all 0.90 or higher (three of the seven pairs had correlations of 0.99), with most differences occurring near the ends of the distribution, the simple summaries we present are unlikely to be strongly affected by the choice of scaling method. Still, we believe that the IRT approach warrants further exploration with the kinds of items examined in this study.

Reading and Language Arts Skills and Activities. The importance of high-quality, early literacy instruction has been highlighted in several recent reports. One that garnered a good deal of attention from educators and policymakers was written by the Committee on the Prevention of Reading Difficulties in Young Children, convened by the National Research Council (1999). The report underscored the importance of making reading instruction a priority during the kindergarten year, and stated that by the end of that year children should have achieved some phonemic awareness as well as an understanding of how reading can contribute to learning. Many researchers and educators argue that the early elementary years should provide a range of activities that promote phonemic awareness, letter and word recognition, print awareness, comprehension, and related skills. The ECLS-K reading and language arts questionnaire items address a wide variety of such activities and skills, and were designed to be aligned with the skills tapped by the ECLS-K cognitive assessments. Because ECLS-K is a longitudinal study, these items were also administered to first-grade teachers the following year, providing researchers with an opportunity to follow students over time and track their exposure to various instructional approaches.

We conducted a single factor analysis on all of the reading and language arts activities and skills items (i.e., questionnaire items 28 and 29), with the exception of three. We excluded two items that asked teachers how often students listened to stories (28f, "Listen to you read stories where they see the print", and 28g, "Listen to you read stories but they don't see the print"). These items did not fit well empirically with the others, and when they were included in the factor analyses they tended to load on their own factor, with opposite signs. It was not clear how such a factor should be interpreted. The third excluded item asked teachers to report how often students "see/hear stories from story tellers or other artists" (28u). This was a relatively low-frequency activity. Scores on this item were correlated with those on the mixed-achievement groups and peer tutoring items, but conceptually this item did not seem to fit well with the grouping factor. In addition, excluding these three items helped us to maintain consistency with an earlier study that reported a factor analysis of the reading/language arts items (Xue et al., 2001).

Our initial factor solution, using weighted principal components analysis, produced scales similar to those identified by Xue et al. (2001). Inspection of the eigenvalues of the correlation matrix revealed a single dominant factor: The first eigenvalue was 9.80, and the second was 3.22. The factor analysis and inspection of the raw correlation matrix suggest that teachers who report more of one practice or skill tend to report more of everything. We observed a similar pattern for the other subject areas. This may be due in part to variations in how frequently each subject is taught—because some teachers spend more time teaching reading than other teachers, they may report greater frequency of all activities. A later section of this report examines the results separately for full- and half-day kindergarten teachers, which addresses this issue in part. Regardless of the source of these correlations, it is clear that there is extensive overlap among items and scales, and for some purposes it might make sense to combine them into a smaller set of scales.

The criterion of minimum eigenvalue of 1 suggested using a seven-factor solution. We examined solutions that included six, seven, and eight factors. In addition, we applied a weighted principal factor analysis to determine whether this method yielded different results from the principal components analysis. Comparing principal components and principal factors solutions provides one test of the stability of the solution; the two solutions would be expected to be similar when the sample is large, when there are many variables, and when communality estimates are similar (Tabachnick & Fidell, 1996). Our data meet the first two of these criteria but not the third, and partially as a result of this, the two solutions were somewhat different. While several of the factors were the same in both solutions, in the latter analysis only four eigenvalues were greater than one, and inspection of solutions with various numbers of factors (ranging from four to seven) failed to identify the *mixed-achievement grouping* and *reading and writing activities* factors that we describe below. These findings illustrate the ambiguity inherent in using exploratory factor analysis to determine how items should be combined to create scales, and suggest that substantive considerations should guide decisions about scale construction. We do not claim that we have found the "best" or the "right" solution, but we hope that we have identified scales that will be useful for future research of instructional practices and skills in reading and language arts.

In addition to exploring both factor-extraction methods, we conducted the analyses with and without the sample weights. The half-sample cross-validations were also conducted using unweighted data. Although the use of weights is generally preferred when the goal is to provide descriptive information that generalizes to a population, there is disagreement about the need for weights when examining relationships among variables, as in factor analysis. The results with and without weights were similar.

Table 1 presents the factor loadings for the seven-factor weighted principal components analysis solution, using the entire data set[4]. We decided to focus on this solution because it produced interpretable results that were largely consistent with those described by Xue et al. (2001). Because of this consistency we retain some of the names assigned to the factors in this earlier study. The names we assigned the scales are provided at the top of the table, and the highest loading for each item appears in boldface type. In the solution presented in this table, some of the items we assigned to the *reading and writing activities* factor have their highest loading on other factors, but we chose to assign them *reading and writing activities* for substantive reasons and because those items loaded on that factor in our initial half-sample analysis. However, there appears to be significant overlap between this factor and *student-centered instruction*, and combining them into a single scale may be reasonable. For items that are assigned to a factor other than the one on which they had their highest loading, an asterisk marks the factor to which the item was assigned.

Table 2 provides the scales, the corresponding items, and the weighted item means. The scales represent a range of reading and language arts instructional activities and skills. The p*honics* scale includes items addressing teachers' emphasis on activities such as learning and practicing letters, and the importance teachers place on skills such as matching letters to sounds and recognizing letters. The *reading and writing skills* factor includes only items addressing skills taught (i.e., item 29); it measures teachers' reported emphasis on promoting the student's ability to read fluently and to write using standard conventions; e.g., capitalization, punctuation, and conventional spelling. The *reading and writing activities* factor includes only items addressing instructional activities (i.e., item 28). In contrast to the previously discussed factor, this one focuses on specific activities in which students engage to promote reading and writing ability, including reading aloud and writing in journals. *Didactic instruction* includes just three items that appear to measure more traditional language arts instructional activities that students probably do as seatwork: reading from basal texts, working in workbooks or on worksheets, and writing words from dictation, to improve spelling. These three items consistently loaded together across validation samples and regardless of the number of factors specified. *Comprehension* covers skills involved in understanding what is read; e.g., identifying main ideas, making predictions, and using context cues. *Student-centered instruction* appears to address

[4] Recall that our initial analysis was conducted on a random half of the data. We use the full data for the presentation of loadings so that the reader can replicate our results if desired. However, the pattern of loadings in this table is not entirely consistent with those in the original analysis.

activities in which students are responsible for producing something and sharing it with the class; e.g., publishing their own writing, performing plays, and doing an activity or project related to a book or story. The final factor, *mixed-achievement grouping*, consists of two items that assess the frequency of mixed-achievement grouping and peer tutoring.

Cr*oss-validation and separation of activities and skills items*. The cross-validation indicated a reasonable degree of stability for our solution. The *reading and writing skills, didactic instruction,* and *grouping* scales looked virtually identical in the two subsamples; that is, the same sets of items had their highest loadings on these factors in both samples. Three items that we assigned to the *phonics* scale—A2NEWVOC, A2CONVNT, and A2RHYMNG—did not always have their highest loadings on that scale. These items sometimes loaded on the c*omprehension* scale, while the six items that we assigned to the *comprehension* scale loaded on that scale consistently. There were also some changes in the *reading and writing activities* and s*tudent-centered instruction* scales; our original decision to separate these two scales was based on a half-sample analysis in which A2DOPROJ, A2PUBLSH, and A2DICTAT loaded most strongly on the latter[5]. As Table 1 shows, when we conducted the analysis with the full sample, some items that had loaded on *student-centered instruction* instead loaded on *reading and writing activities.* There appears to be more overlap between these two scales than among the other scales in the analysis.

As can be seen from Table 1, the items addressing activities (item 28) and those addressing skills (item 29) tended to separate themselves, with the exception of the *phonics* scale, which includes items from both sets. It is not surprising that they tended to separate, since one set provides information on what teachers do in the classroom, whereas the other focuses on what skills teachers believe are important. However, as the example of *phonics* illustrates, in some cases increased use of certain activities appears to be accompanied by increased emphasis on certain skills. We believe that in these cases it is sensible to combine the two sets of items because together they provide information about a particular approach to teaching reading and language arts.

Scale reliability. The reliability of each scale was estimated using coefficient alpha. Reliability is in part a function of the number of items, so that scales with few items generally have fairly poor reliability. In addition to obtaining an overall alpha coefficient, we examined the change in the alpha when each item was removed, to determine whether any items functioned particularly poorly in the context of the scale to which they were assigned. In all cases, removing an item lowered the scale reliability, as expected. Therefore we retained all of the items in the scales to which we had assigned them. Reliability estimates are presented in the third column of Table 3. As expected, the reliabilities of the didactic instruction and grouping scales are low due to the small numbers of items on these scales. This suggests that these scales should be used with caution. As we discuss in a later section that examines correlations among all of the instruction scales, some of these scales may be combined in ways that would enhance reliability without sacrificing information. However, the two scales with particularly low reliability in our analysis seem to address constructs that are distinct from what is tapped by the other reading and language arts scales. Researchers who are interested in grouping practices in reading, for

[5] This also occurred when we conducted the analysis using the full sample but without sample weights.

example, might be wish to use this limited set of items. An alternative solution is to combine similar items across subjects; we discuss this in the section on correlations among scales.

IRT analyses. For the reading and language arts items only, we applied Samejima's (1969) graded response model to create IRT scores that corresponded to each of our sum scores. We then examined correlations among the two sets of scales, and compared the distributions of corresponding scales. This section is intended to illustrate an alternative approach to scale construction and to describe the degree of correspondence between scales created using the IRT and sum score methods. We believe that the use of IRT scaling for items like those in the ECLS-K questionnaires is a promising method for addressing some of the limitations of sum scores and should be a focus of future research efforts.

We created seven IRT scores, one for each of the reading and language arts scales. Next to the coefficient alpha reliability estimates for the sum scores in Table 3 are the marginal reliability estimates for the IRT scores and the correlations between the corresponding IRT and sum scores. The marginal reliability is a measure of overall reliability for the IRT score, and Table 3 shows these values to be similar to the internal consistency estimates obtained for the sum scores. One of the advantages of the IRT approach, however, is that it can reveal how measurement precision varies along the range of the attribute of interest. Scales are more precise in some regions than others. In particular, precision tends to be lowest near the ends of the distribution, especially when there are floor or ceiling effects (Hambleton, Swaminathan, & Rogers, 1991). To illustrate, Figures 1 through 6 illustrate how the measurement precisions varies along the *student-centered instruction*, *reading and writing activities,* and *phonics* scales, respectively.

Figure 1 shows that the distribution of *student-centered instruction* is relatively symmetric and close to normal, with few respondents scoring at either end of the distribution. The IRT and sum score distributions correspond closely. The standard errors are only slightly larger near the ends of the distribution than in the middle, suggesting that floor and ceiling effects are not a problem (Figure 2). In this case, the overall estimates of reliability and standard error of measurement may be appropriately applied to most points along the response scale.

The distribution of *reading and writing activities* is less symmetric, with more scores at the high end of the range than at the low end (Figure 3). The distribution of the IRT scores is less skewed than that of the sum scores, a result that is not uncommon when scores are clustered in a particular part of the distribution. The IRT approach to assigning scores essentially adjusts for the item difficulties (or item means, in this case) by assigning smaller score increments in regions where item means cluster and larger scores increments in other regions. Because the items in this scale tended to be high-frequency ones, the IRT score distribution pulls many of the scores down and closer to the middle of the score range. The skewness of the distribution leads to a small increase in the standard error of measurement at the high end of the scale (Figure 4).

Phonics provides an extreme example, with a very strongly skewed distribution that indicates that most teachers engage in phonics activities with high frequency (Figure 3). The changes in the standard error of measurement are more pronounced for this scale than for any of the others, and there are fairly large differences in the distributions of IRT and sum scores. This is one reason for the relatively low (0.90) correlation between these two scores. There is a substantial increase in the standard error of measurement as scores on this scale increase, which indicates a lack of precision of this scale among high-scoring teachers. One way to compensate for this in future survey research would be to add additional items that represent less frequently-used phonics approaches. Another would be to change the response scale to distinguish among the large group of teachers who are engaged in phonics activities every day; for example, a duration component could be added.

Again, the purpose of this set of analyses was to illustrate the potential utility of the IRT approach. The IRT methodology provides a method for displaying information that is difficult to obtain through standard methods; e.g., the degree to which measurement precision varies along the measurement scale. Moreover, there is some evidence from other studies that IRT scores based on Likert-type questionnaire items have greater precision than sum scores for detecting differences among groups, and that they are especially useful for measuring change because they are less dependent on initial baseline score than are sum scores (Chan, 2001). Additional research on the application of IRT scoring to the ECLS-K questionnaire items could produce findings and techniques that would enhance future studies of school and teaching effects.

Mathematics Skills and Activities. The mathematics items, like those for reading and language arts, cover a broad range of activities and skills. As with reading and language arts, these items were designed to be aligned with the skills assessed in the ECLS-K cognitive battery and to provide information on both kindergarten and first-grade classrooms. Some of the activities are topic-specific (e.g., “work with counting manipulatives to learn basic operations”) whereas others may be applicable across topic areas (e.g., “explain how a math problem is solved”). Most of the skills items are topic-specific, and range from those that many students have acquired before entering kindergarten (e.g., “recognizing and naming geometric shapes”) to those that are typically covered in later grades (e.g., “subtracting two-digit numbers without regrouping”). We conducted exploratory factor analyses on these items to explore ways in which sets of items clustered together.

Our final set of factor analyses excluded two of the 17 items addressing activities (i.e., item 31), and four of the 26 items addressing skills (i.e., item 32). Items 31e, “use a calculator for math”, and 31j, “engage in calendar-related activities”, had communality estimates of under 0.20 in our preliminary factor analyses, and did not appear to fit well with the rest of our solution. The calculator item did not have much variance, with 78% of teachers reporting that they never used calculators. The calendar item also had very little variance, with daily use reported by 92% of teachers. Similarly, the four items that we excluded from the set addressing skills—mixed operations (item 32q), adding two-digit numbers (item 32y), carrying numbers in addition (item 32z), and subtracting two-digit numbers without regrouping (item 32aa)—represent advanced skills for which approximately 90% of teachers reported no coverage. Despite the low variability of these four items, there may be purposes for which it would be worthwhile to create a scale that consists of these four advanced skills items, particularly when examining student achievement growth at some of the higher proficiency levels or for research that follows students into first grade. For our purposes, since we were mainly interested in exploring variability among the full kindergarten teacher sample, we chose to exclude them.

Using the remaining 40 items, we explored several solutions using a random sample of approximately half the data, and later conducted a cross-validation using the other half. As in the reading and language arts data, a single dominant factor characterized this set of mathematics items. Based on a weighted principal components analysis, the first eigenvalue of the correlation matrix was 9.9, and the second was 2.8. Using a criterion of one as a minimum eigenvalue, nine factors were initially retained. We examined the patterns of loadings on these nine factors, and also inspected seven-, six-, five-, and four-factor solutions. The five-factor solution produced results that were most easily interpreted. With larger numbers of factors there were generally one or more factors that included only a single item. Table 4 presents the loadings for the five-factor solution for the entire sample (again, recall that the original analysis was based on half the data and produced slightly different results). In addition to the cross-validation discussed below, we examined the stability of the solution by conducting both weighted and unweighted principal components and principal factor analyses. The solutions in all cases were quite similar. In particular, the weighted principal factor analysis suggested that five factors should be retained (based on the minimum eigenvalue criterion), and the set of factors identified based on the five-factor solution was nearly identical to the set identified using principal components analysis.

We discuss six scales. For the most part these are consistent with the factor solution presented in Table 4, except that we created an additional scale, *mixed-achievement grouping,* to maintain consistency with the reading and language arts solution. Mean scores and variable labels for each of these items are given in Table 5. Some of our scales focus on particular content, whereas others related more to the structure of the classroom instruction. The first factor, n*umbers and geometry*, consists of eleven items, three of which address classroom activities designed to promote understanding of numbers, operations, and geometry. The other eight items on this scale come from the "skills" item set (31) and address instructional objectives related to numbers and geometry. The second factor, *advanced numbers and operations*, contains five items from the "skills" item set. In contrast to the first factor, these items emphasize skills related to somewhat advanced numeracy concepts; e.g., reading two- and three-digit numbers. *Traditional practices and computation*, the third factor, consists of three "activities" and four "skills" items. This factor appears to identify teachers who take a fairly didactic approach to instruction and who emphasize computational facility. Emphasis on recognizing the value of coins and currency also loaded on this factor. The first three items from this factor are similar to those that load on the reading and language arts didactic instruction scale. They include having students do math worksheets, do math problems from their textbooks, and complete math problems on the chalkboard. This set of three items could be treated as a separate scale by researchers who are interested in distinguishing instructional approach from content taught. However, these items consistently loaded with the other skills-oriented items on this scale, and combining the two sets of items produces a scale with a reasonable degree of reliability (see below).

We labeled the fourth factor *student-centered math instruction*. This scale is similar to the similarly-named scale we identified for reading, and consists of six items that emphasize having students take an active role in their own learning, beyond doing seatwork. Our fifth scale, *mixed-achievement grouping*, consists of two items: use of mixed-achievement groups and peer tutoring. As mentioned above, in order to maintain some consistency with the reading/language arts solution and provide opportunities for examining the similarity of teachers' approaches across subjects, we separated the “mixed-achievement groups” and “peer tutoring” items from the other student-centered instruction items even though they loaded on the student-centered math instruction factor. Removing the two grouping items from the student-centered instruction scale in math results in a very small decrease in scale reliability (the coefficient alpha reliability decreases from .77 to .74) but does not substantively alter the correlations between student-centered instruction and other variables. As we discuss later, the grouping scales for reading/language arts and math are highly correlated with one another, suggesting that teachers who use these strategies in one subject tend to use them in both.

Finally, *measurement and advanced topics* includes one item from the “activities” set—work with rulers, measuring cups, spoons, or other measuring instruments—and eight “skills” items emphasizing topics other than the basic numbers and geometry topics addressed by the first and second factors. These include fractions, data collection, probability, and estimation, along with measurement-specific skills such as “using measuring instruments accurately.”

The National Council of Teachers of Mathematics (NCTM) states that the Number and Geometry content standards should form the core of mathematics instruction in the early grades, and that other content, such as data analysis and algebra, should be woven into this core (NCTM, 2000). Other published guidelines for early elementary mathematics instruction, such as the C*ore Knowledge Sequence* (1999) and many of the state-developed content standards, promote a similar emphasis on numbers and geometry. The first factor that we identified, therefore, may be considered an indicator of teachers’ emphasis on this core set of concepts and skills. As we discussed above, our scales should not be considered the only or even the best set of measures. These scales were derived empirically, and provide an indication of what practices tend to be associated with one another. For some purposes, however, users of the ECLS-K data might find it informative to combine items based on the content or skill area they represent. For example, to obtain an indicator of teachers’ emphasis on numbers and operations (as defined, for example, by NCTM), one could select some of the items that we include in our *numbers and geometry* scale, and add a few that are not represented there, such as emphasis on fractions or place value.

Cross-validation and separation of activities and skills items. Our cross-validation indicated a stable solution. Items that loaded on different factors in the three solutions (the two half-samples and the full sample) tended to be those that had nearly equal loadings on more than one factor in all solutions. For example, though we assigned A2TELLTI (telling time) to the *measurement and advanced topics* factor, in some cases it loaded more highly on the *traditional* factor. A2MANIPS (working with counting manipulatives), which we assigned to *numbers and geometry*, also had a high loading on *student-centered instruction,* and A2MTHGME (play math-related games), which we assigned to *student-centered instruction*, also loaded highly on *numbers and geometry.* In all, for four of the 40 items, the factor on which the item had its highest loading failed to be consistent across the three solutions.

Several of the factors included items addressing both activities and skills. As with reading and language arts, we conducted separate factor analyses for these two sets of items but decided to retain our original scales. Separating the two sets would result in a larger number of factors and did not appear to enhance interpretability, and the high correlations between some sets of activities and skills items suggested that there would be advantages to combining them.

Scale reliability. For each of the six scales, we used coefficient alpha to estimate internal consistency reliability, and we examined the effects of removing each item on the scale's overall reliability. As with reading and language arts, in every case, removing an item resulted in a reduction in scale reliability. Table 6 presents the alpha coefficients, which indicate reliability levels of at least 0.7 for all but the *mixed-achievement grouping* factor. Below we discuss the possibility of combining the two mixed-achievement grouping items in mathematics with corresponding items in reading; because of the high correlation among all of these items, a single mixed-achievement grouping scale has a fairly high internal consistency and may be preferable to using two separate scales.

Science and Social Studies Skills. These analyses included items 34a through 34v, which asked teachers to report the frequency with which they taught various science and social studies topics and skills. Using the same half-sample cross-validation approach described earlier, we conducted one set of analyses for the 14 science items and another for the 8 social studies items (see Table 7 for item descriptions, labels, and means). In science, there was strong support for a single dominant factor: The first two eigenvalues of the correlation matrix were 5.7 and 1.4. We examined a three-factor solution (see Table 8) based on the criterion of minimum eigenvalue of one. In this solution, weather and hygiene clustered together in one factor, with temperature having nearly equal loadings on this factor and the physical science factor described below. As shown in Table 7, teachers reported that weather and hygiene were the most frequently taught topics among all the science and social studies items. The attention to weather probably reflects the prevalence of a daily "circle time" or similar activity in which teachers and students discuss calendar, weather, and current events. The other two factors appeared to represent a distinction between life/earth sciences and physical science. Sum scores created based on these factors were moderately correlated with each other (weighted correlations were 0.65 for the life/earth and physical science factors, and 0.42 and 0.44 respectively between these two factors and the weather, temperature, and hygiene factor). A two-factor solution resulted in one factor on which human body, plants, weather, and hygiene loaded highly; all other items except temperature loaded on the other factor. Temperature did not load highly on either factor.

A two-factor solution was initially examined for social studies; the first two eigenvalues were 3.7 and 1.0. The items addressing history, community, map-reading, and cultures loaded on the first, and items on laws, ecology, geography, and social-problem solving loaded on the second. These results are presented in Table 9. The correlation between the sum scores constructed from scores on these items was 0.59.

Cross-validation suggested that both the science and social studies solutions were stable: With one exception, all items had their strongest loadings on the same factor in both validation samples and the full sample. The exception was the "temperature" item in science, which in one half sample loaded most strongly with the hygiene and weather items. Despite this stability, several of the factors are difficult to interpret. In particular, it is not clear what construct is represented by teachers' emphasis on weather, temperature, and hygiene. In addition, most of the scales that would result if we used the factor solutions would have only three or four items, leading to low reliability. These considerations suggest that a single scale for science and another for social studies would be appropriate for most research purposes. In fact, the responses teachers give to these two sets of items may best be captured by one scale that includes both subjects. In a factor analysis that combined the two sets of items, there was again a single dominant factor (the first two eigenvalues were 8.2 and 1.9), and a two-factor solution did not produce subject-specific factors. Therefore an alternative to constructing science and social studies scales would be a single scale that encompasses both. For the correlations presented later in this report, we chose to keep the science and social studies factors separate, but the reader should keep in mind that they are highly correlated. Internal consistency reliability coefficients are 0.86, 0.82, and 0.90 for the science, social studies, and combined scales, respectively (see Table 12).

Computer Use. Items 33a through 33i asked teachers to report the frequency of use of computers for each of nine purposes. Several teachers reported not using computers for any of the purposes included in this questionnaire: 316 of the 3243 teachers in this sample, or almost ten percent, responded "never" to all nine computer items. These items were also examined in a study by Taylor, Denton, and West (2001), which reported data on the individual items focusing on use of computers for reading, mathematics, science, and social studies. This study found that about two-thirds of teachers reported using computers at least once a week for reading and slightly fewer did so for math, whereas computer use for the other subjects was less frequent. Taylor et al. also observed more frequent computer use and greater availability of computers in public schools than in private schools.

We applied factor analysis to explore ways of combining information across this set of items. As might have been expected, a single dominant factor was identified for computer use. Because the first two eigenvalues were 4.1 and 1.2, we initially explored a two-factor solution, which seemed to reflect differences in the frequency of computer use for various purposes. The first factor included the more frequent activities: computer use for reading, math, keyboarding skills, art, and learning games. The second factor included science, social studies, music, and accessing information, the less-frequently reported activities (see Table 10 for the factor analysis results and Table 11 for item means). The correlation between the sum scores created from these two factors was 0.55. Since the scales were highly correlated and seemed to reflect frequency differences rather than clearly interpretable patterns of use, we decided to construct a single scale that represents frequency of computer use. The internal consistency reliability of the total scale was 0.84 (see Table 12). Of course, as with the other sets of items we examined, there may be specific analytic tasks that would benefit from selecting a subset of these items, but that decision should be based on the substance of the item rather than just on the correlations among the items.

Correlations among Instruction Scales. Table 13 provides correlations among all of the scales addressing instructional activities and skills, including computer use. Consistent with the previous discussion, this table shows small to moderate positive correlations among all of the instruction scales. The presence of only positive correlations suggests that teachers' responses may be driven partly by response sets, or general tendencies to respond at either the high or low end of the scale. The positive correlations may also reflect differences in frequency with which particular subjects are taught, though this provides a better explanation for within-subject correlations than cross-subject correlations. Finally, it is possible that length of the kindergarten day is driving these relationships in part; some teachers may do more of everything than other simply because they have more time. A subsequent section of this report examines the factor structures and correlations separately for full- and half-day kindergarten teachers.

Examining the patterns of correlations suggests ways to combine scales both within and across subjects. Among the reading and language arts scales, the highest correlation (0.61) is that between *student-centered instruction* and *reading and writing activities*. The next largest correlations are those between r*eading and writing skills* and *reading and writing activities* and between r*eading and writing skills* and *comprehension*; both correlations are 0.52. This is not surprising given the content of the scales. In mathematics, all of the correlations among *numbers and operations*, *student-centered instruction*, and *measurement and advanced topics* are higher than 0.50, as is the correlation between *student-centered instruction* and *mixed-achievement grouping.* Thus, even though most of these scales showed a reasonable degree of stability as separate factors in the factor analyses described earlier, combining them to form a smaller set of more reliable scales is probably justifiable in most cases.

An inspection of correlations among scales for different subjects suggests some additional pairs of scales that could be combined. One of the highest correlations in Table 13 is that between the *mixed-achievement grouping* scales in mathematics and reading. This correlation of 0.67 is especially noteworthy given the low reliability of these two scales, which would tend to depress the correlation. This suggests that teachers who use mixed-ability grouping strategies tend to do so for both reading and math, and that it may be worthwhile to create a single scale that addresses grouping in both subjects. The internal consistency reliability for this four-item measure is 0.74, which is a substantial improvement over the reliabilities of the individual scales. *Student-centered instruction* in the two subjects correlated at 0.56, and the correlation between *didactic instruction* in reading and *traditional practices and computation* in math was 0.51. Like the grouping scales, both of these pairs of scales could be combined into single scales that capture stylistic tendencies of teachers that cross subject lines. The internal consistency reliabilities are 0.81 and 0.76 for the combined student-centered and didactic/traditional scales, respectively. Finally, consistent with the factor analysis results discussed earlier, the science and social studies scales are highly correlated (0.69) and may be combined into a single scale that has an internal consistency reliability of 0.90. In all of these cases, inspection of item intercorrelations and the effects of each item's removal on the scale's alpha coefficient indicates that these combined scales function reasonably well.

The choice of which scales to use should depend on analytic purposes. For example, if the objective is to understand what factors lead teachers to adopt a particular style, or to understand the overall classroom environment, combining items across subjects might produce the most reliable and informative scale. For the researcher who is interested in predicting student achievement in a particular subject, the use of a scale constructed only from the items addressing that subject might be preferred for substantive reasons and to ensure interpretability.

Length of Kindergarten Day. We examined the factor structures of the instructional activity and skill items separately for teachers who reported teaching in full-day and half-day kindergarten programs. We used the KGCLASS variable to determine whether each teacher should be assigned to the full-day group or to the half-day group, and we deleted the eight teachers who reported teaching both full- and half-day classes. It should be noted that although half-day teachers on average reported fewer hours of total instruction than full-day teachers, there was overlap, with some half-day teachers reporting as much as five or six hours, and some full-day teachers reporting as few as two. More detailed information about time can be obtained from questionnaire items that ask teachers to report how frequently and for how long they teach each of a number of subjects.

For reading and language arts, we examined seven-factor solutions in each of these groups to determine whether the patterns of loadings would be similar to those obtained with the full sample. The analyses were conducted using the 905 half-day teachers and the 1418 full-day teachers who had complete data on this set of items. Several of the factors were similar to the originals in both the full- and half-day samples: *Reading and writing skills*, *phonics*, *comprehension*, and *mixed-achievement grouping* were nearly identical. In the analysis using half-day teachers, *didactic instruction*, *student-centered instruction*, and *reading and writing activities* failed to emerge as distinct factors. The use of workbooks loaded negatively on the factor that included many of the *student-centered instruction* items, and use of basal readers loaded on a factor that also included items that involved reading silently or reading books students had chosen. Thus there seemed to be a "reading" factor that did not emerge in the full sample. Among full-day teachers, the scales looked very similar to those we obtained with the full sample. In particular, the *reading and writing activities* and *student-centered instruction* scales contained the same sets of items that we had assigned to these scales in the total sample, despite the general lack of stability that characterized these scales.

The math analysis included 892 half-day teachers and 1393 full-day teachers. Both the full- and half-day solutions were largely consistent with the original solution. Among half-day teachers, all but two items loaded on the same factors as in the total sample. Among full-day teachers, the *numbers and geometry* factor split into two, with the first 5 items in Table 4 loading on one factor and the next 6 on another factor. This latter factor also included most of the items that loaded on the original *measurement and advanced topics* factor. The *traditional practices and computation*, *advanced numbers and operations,* and *student centered instruction/grouping* factors were similar to those obtained in the total sample.

The science, social studies, and computer use solutions were consistent with the original, with two exceptions. First, in history, "map reading" loaded on the second factor instead of the first in the half-day sample only. Second, in science, solar systems loaded with the physical science items instead of with the life science items in the full-day sample. As in the total sample, a single factor for each subject fit the data well, and a single factor combining science and social studies also appeared to be reasonable in both the full- and half-day samples. Correlations among all of the instruction scales were also examined separately for full- and half-day kindergarten teachers, and the magnitudes and directions of the correlations in both samples were similar to those in the original.

It is difficult to know the source of the differences in some of the factor solutions; they could arise in part from the use of smaller and perhaps less representative samples of teachers. The generally consistent results suggest that, for the most part, the structures we observed in our original analyses are applicable to both full- and half-day kindergarten classrooms.

Evaluation Strategies and Views on Readiness[6]

Items 3a through 3j in the fall teacher questionnaire asked teachers to rate the importance of various factors in evaluating the children in their classes, using a four-point scale ranging from "not important" to "extremely important." Teachers could also mark "not applicable"; we excluded these responses from our analysis. Rathbun, Walston, and Germino-Hausken (2000) examined teachers' responses to the individual items as a function of teacher background and school characteristics. We build on this work, conducting a factor analysis to explore the degree to which clusters of evaluation criteria would emerge. Table 14 lists these items and the mean scores. The item that asked about the importance of the student's improvement or progress over past performance had a low communality estimate—0.20—suggesting that it functions differently from the other items in this set. Therefore we eliminated this item from our analysis. We also eliminated the "other method" option. Our analyses produced a two-factor solution for the remaining items (see Table 15). The first factor represents an achievement orientation, with emphasis on the child's achievement relative to the rest of the class and relative to local, state, or professional standards. Teachers tended to assign lower ratings to these two items than to the others. The second focuses on aspects of students' performance other than achievement, including effort, participation, and attendance. Scales derived from these two factors, plus a teacher's score on the "individual improvement" item, together would provide a summary of that teacher's relative priorities when evaluating students.

A related set of items (7a through 7m), also administered in the fall, asked teachers how important they believe various characteristics are for children to be ready for kindergarten. The five-point scale ranges from "not important" to "essential." Mean scores on these items are presented in Table 16. We obtained a two-factor solution for these items after eliminating "knows the English language," which had a low communality estimate. This item may function differently from the others in part because many teachers lack experience with children whose first language is not English. Our first factor consists of four items that focus on specific academic skills—"can count to 20 or more," "is able to use pencils and paint brushes," "knows most of the letters of the alphabet," and "identifies primary colors and shapes"—plus the item "has good problem-solving skills." The second factor consists of seven items that relate to social or behavioral skills, such as "takes turns and shares," "is not disruptive of the class," and "can follow directions." The "problem-solving" item, which we assigned to the first factor, loaded nearly as strongly on the second factor, probably because problem solving ability is relevant to the social as well as the academic domain; that is problem-solving skills are necessary for success in academic activities as well as in social interactions (see Table 17).

Table 18 presents the internal consistency reliability estimates for the evaluation and readiness scales. As expected, the two-item scale that measures teachers' emphasis on achievement-related evaluation criteria has a lower reliability than the others. The internal consistency estimate for the academic readiness scale improved very slightly when the problem-solving item

[6] In this and the following two sections, we omit discussion of the cross-validation results because in every case the solutions were consistent across the half-samples and full samples (i.e., all items loaded on the same factors in all samples).

was removed (not shown in the table), providing further evidence that this item functions somewhat differently from the other items on this scale.

Table 19 provides correlations among the evaluation and readiness scales. The simple correlations suggest that the readiness and evaluation criteria scales tap different dimensions of teachers' philosophies, though the low reliability of the achievement evaluation scale makes these results difficult to interpret. The two readiness scales were reasonably highly correlated, suggesting that there may be a general construct underlying teachers' responses to all of the readiness items.

School Climate and Teacher Satisfaction

In this section we discuss items 10 through 13 from the fall teacher questionnaire (Part B). This set of items elicits teachers' opinions on issues related to the professional and academic climate of the school. We also examine teacher satisfaction as reported in items 14a through 14c. In items 10a through 10g, teachers are asked how much they agree with each of a set of statements about school spirit, child misbehavior, collegial relationships, and parent support. Items 11 and 12 ask teachers to assess the amount of influence they have over (1) school policy and (2) classroom issues such as skills to be taught, teaching techniques, and discipline. In Items 13a through 13f, teachers report their extent of agreement with six statements about school environment, including the adequacy of academic standards, agreement over the school's mission, and the effectiveness of the school's administrator. All items use a 5-point scale. The items and means are presented in Table 20.

We conducted one factor analysis using items 10a through 10f, and separate analysis for 13a through 13f. A third factor analysis combined these two sets of items, and a fourth incorporated items 11 and 12 into this analysis. Based on the results of these analyses and inspections of item intercorrelations, we focus on a three-factor solution that includes all of the items, presented in Table 21. The first factor, *leadership,* focuses on teachers' evaluations of their administrators' performance and their own role in determining school policy and mission. Given the importance of leadership in determining the quality of school reform implementation (Berends et al., 2001), this scale may prove especially useful for efforts to model relationships among school and classroom characteristics and student achievement. We call the second factor *efficacy.* It includes teachers' report of how much influence they have over their own classrooms, as well as their perceptions of factors that interfere with their teaching (see, e.g., Hoy & Woolfolk, 1993). The third factor, *professional climate*, focuses on how teachers relate to their colleagues as well as their perceptions of parental support and the adequacy of standards at the school. Scores on this scale may be an indicator of the degree to which teachers feel they are respected and treated as professionals. Related constructs, such as teacher collegiality, have been shown to be important predictors of student achievement (see, e.g., Bryk, Lee, & Holland, 1993; Louis & Marks, 1998).

We also created a separate scale that assessed teacher career satisfaction. Items 14a through 14c asked teachers the extent to which they agreed with each of the following three statements: "I really enjoy my present teaching job"; "I am certain I am making a difference in the lives of the children I teach"; and "If I could start over, I would choose teaching again as my career." As shown in Table 14, scores on these items tended to cluster near the high end of the scale. As expected, scores on these three items were highly correlated with one another. Table 22 presents internal consistency reliability estimates for the three climate scales and the satisfaction scale, and Table 23 presents the scale intercorrelations. All four scales are moderately positively correlated with one another, indicating that teachers who are satisfied with their jobs tend to perceive a better professional climate than those who are not.

Of course, the reasons for these relationships are unknown. Given the importance of attracting and retaining good teachers, a potentially worthwhile next step would be to model teacher satisfaction as a function of school characteristics such as climate and resources, to better understand the factors that are related to teachers' career satisfaction. In addition, it would be useful to examine how climate and satisfaction relate to teachers' professional development experiences, including the specific training activities addressed in items 40a through 40j (from the spring questionnaire) and collaborative opportunities addressed in items 39a through 39d (also from the spring questionnaire).

Resources

There is great interest in understanding the kinds of school resources that are available to children and how these are distributed according to neighborhood and family characteristics such as socioeconomic status. A study by West & Denton (2001) examined administrators' reports of the availability of a number of facilities including libraries, playgrounds, computer labs, and art and music rooms, as well as their evaluations of school climate problems such as crime and violence. They found that although basic facilities such as classrooms and libraries were available to almost all children, facilities that promote enriching opportunities were less prevalent for students from at-risk families than for other children. Similarly, children from at-risk families were more likely to attend schools that lacked a safe and resource-rich learning environment than were other children.

Rathbun, Walston, and Germino-Hausken (2000) examined another set of resource items from the fall teacher questionnaire. Items 2a through 2k asked teachers whether their classrooms included each of a set of 11 interest areas or centers, including reading area with books, water or sand table, and computer area. The average teacher reported having nine of these 11 centers, with science-related centers less available than those related to reading, math, the arts, and computers.

To build on this work, we examined items 26a through 26r on the spring teacher questionnaire, which asked teachers to rate the adequacy of 18 resources including instructional materials, technology, and facilities. Teachers could choose the option "I don't use these at this grade level," or rate the adequacy on a four-point scale ranging from "never adequate" to "always adequate." Because we were interested in constructing scales to measure teachers' perceptions of the adequacy of resources and not their use of them, and because we wanted to avoid large numbers of missing values in the factor analysis, we removed from the analysis those items for which 10 percent or more teachers reported no use. The percentages of teachers reporting no use were particularly large for textbooks (63%), materials for teaching LEP children (42%), and materials for teaching children with disabilities (35%). For the remaining items the factor analysis was conducted for all teachers who had responses to all items (i.e., records for teachers who reported no use for at least one of these items were deleted). The items and their means are presented in Table 24, and the factor analysis results are presented in Table 25.

We identified three factors. The first, which included the majority of the items, emphasized materials that are directly related to instructional activities, other than computers: manipulatives, audiovisual equipment, videotapes and films, paper and pencils, ditto or photocopier equipment, art materials, and musical recordings. The second factor included the two computer-related items: computer equipment and computer software. The third factor reflected more general facilities: heat and air-conditioning, classroom space, and child-sized furniture.

Internal consistency estimates are provided in Table 26 for the three resource scales. The two-item computer scale had an extremely high internal consistency because of a strong relationship between teacher reports regarding the adequacy of software and hardware. Table 27 provides correlations among four resource-related scales: the three from the factor analysis described above, plus a fourth scale that indicates the total number of interest areas or centers teachers have in their classrooms; these are the items examined in the study by Rathbun, Walston, and Germino Hausken (2000). Adequacy of instructional resources was moderately correlated with adequacy of both computers and facilities. Correlations between each of the three "adequacy" scales and the total number of interest areas or centers in the classroom were quite small, suggesting that the scales are tapping different aspects of resource availability.

Several other teacher questionnaire items address resources; these include reports of frequency of visits to school library or media center, access to arts instruction, and frequency of use of the resources and materials included in items 26a through 26r. Future work should examine relationships among the various resource-related variables from both the teacher and administrator questionnaires.

V. Summary And Applications To Future Research

Summary

In this study we explored the feasibility of constructing measures for each of several categories of constructs related to teachers and teaching. A set of exploratory factor analyses produced several scales that address teachers' reported use of various instructional practices and their emphases on different types of skills. Seven scales were identified for reading and language arts, six for mathematics, and one each for science, social studies, and computer use. Items addressing the latter three subject areas were fairly unidimensional within each subject, perhaps reflecting differences in availability of instructional materials and time spent on those subjects. All of the instruction scales were positively correlated with one another, and most of these correlations were moderate in magnitude. In other words, teachers who report emphasizing more of one thing than their colleagues tend to report emphasizing more of everything. Without more information it is impossible to determine whether this represents a type of response bias or whether it reflects real differences among teachers. It is possible that time spent on teaching is one factor driving the positive correlations, but inspection of scale score correlations for full- and half-day kindergarten teachers separately does not provide evidence that this is the case.

The pattern of correlations also suggested ways in which some of our scales might be combined to produce measures that have greater reliability than the separate scales but that reflect similar constructs. In particular, teachers' use of student-centered strategies and mixed-ability grouping tended to be somewhat consistent across reading and mathematics, which suggests that those pairs of scales could be combined to provide descriptors of overall classroom environment across subjects.

In addition to the items on instructional emphases and skills, several sets of items addressing a subset of teachers' philosophies, school climate, teacher satisfaction, and resource adequacy were examined. Within the first category, two scales representing different approaches to evaluating students and two that assessed teachers' views of the importance of different criteria for determining kindergarten readiness were identified. In both of these sets of items, the two scales that emerged revealed contrasts between an emphasis on academic skills and an emphasis on social or behavioral skills. The strongest correlation among this set of four scales was the correlation between the two readiness scales, suggesting that teachers varied in the overall importance they attached to readiness criteria in general.

Three scales related to school climate were identified. The first, *leadership*, is one that may be especially important for future research efforts given the accumulating evidence of the contributions of leadership quality to the success of education reform initiatives. We also examined a three-item scale that focused on teachers' levels of job satisfaction and found that it was moderately positively correlated with all three climate scales, not surprisingly. Finally, we identified three resource-related scales that assessed teachers' views of the adequacy of various resources and facilities.

As discussed in the section on scale construction, we do not claim that the approach to combining items used in this report is the only or even the best approach. Some alternative approaches (e.g., combining similar scales across subjects) are suggested, but the ultimate choice of scales should depend on the purposes of the research. It is also important to acknowledge the limitations of the measures examined. Perhaps most important is that self-report data inevitably fail to capture some important aspects of the constructs of interest. In particular, they do not measure differences in the quality of instruction provided. In addition, this study utilized a fairly straightforward strategy for combining items to create scales; the simple sum score may not be the best metric for synthesizing information across items but has the advantage of being easy to communicate. Explorations of an IRT-based approach suggested that this method warrants further research. Despite the limitations of the study, it is hoped that the scales identified will prove useful in future research efforts that seek to understand ways in which teachers promote student learning using large databases such as ECLS-K. We now turn to a brief discussion of what we believe are important next steps in the analyses of these data.

Applications to Future Research

The scales we have constructed provide a tool for distinguishing potentially important differences in instructional styles and approaches. To understand the mechanisms underlying these relationships, there are several clear next steps that could be taken. In this section we briefly discuss possible avenues for future research.

First, it would be informative to conduct similar analyses using the first-grade ECLS-K data to determine whether the patterns of relationships among items and scales for the kindergarten teachers are similar to those for first-grade teachers. As we discussed earlier, the correlational pattern for the instructional scales in the kindergarten data—specifically the finding that all scales were positively correlated, both within and across subjects—is somewhat different than what is typically found in later elementary and secondary grades, but may be partly a function of time spent teaching. The ECLS-K study is following the kindergartners into the first, third, and fifth grades, offering an opportunity to examine changes in the instructional practices to which those students are exposed. An important first step in conducting such longitudinal analyses would be to understand the structure of the instructional practice and skill items in the first-grade data.

Second, research is needed to examine how scales that we identified are distributed across different types of schools and teachers. In particular, it would be of interest to policy makers to understand how scores on the various instructional scales might differ by the wealth, minority status, and location of schools. In addition, given the importance of strong leadership and academic climate, descriptive analyses of the non-instructional scales for teachers in different kinds of schools and neighborhoods would be worthwhile. Similarly, differences in teachers' perceptions of the adequacy of resources across different kinds of schools should be investigated. By using the resource-related scales we constructed, combined with other information on resources from the teacher and administrator questionnaires, researchers could obtain rich descriptions of the kinds of resources available to children in different types of schools.

Third, the large number of constructs we examined and the complex nature of relationships linking them to one another and to student achievement call for the application of modeling strategies that could illuminate the mechanisms by which various characteristics of teachers and teaching may influence student achievement. The longitudinal nature of the ECLS-K, combined with the richness of student and teacher data it provides, make it an ideal data set for examining these relationships. We have completed an important step in such modeling; namely the creation of scales. Efforts to understand the sources of the differences in scale scores that we observed and to link these teaching influences to student achievement should be the focus of future research.

Table 1.——Factor Loadings for Reading and Language Arts Instructional Activities and Skills Items, Full Sample (n=2323): 1999

Item number	Variable name	Reading and writing activities	Phonics	Didactic instruction	Compre-hension	Student-centered instruction	Reading and writing skills	Mixed-ability grouping
29f	A2SYLLAB	0.09	0.02	0.11	-0.20	-0.21	**-0.61**	0.01
29m	A2PNCTUA	0.22	0.14	0.04	-0.22	0.18	**-0.68**	0.02
29n	A2COMPSE	0.40	0.05	-0.03	-0.18	0.05	**-0.72**	-0.01
29o	A2WRTSTO	0.25	-0.03	0.00	-0.14	-0.23	**-0.66**	0.03
29p	A2SPELL	0.13	0.07	0.17	-0.03	-0.04	**-0.71**	0.08
29q	A2VOCAB	-0.03	0.17	0.09	-0.20	-0.03	**-0.52**	0.33
29r	A2ALPBTZ	-0.01	0.06	0.20	-0.08	-0.20	**-0.53**	0.17
29s	A2RDFLNT	0.13	0.01	0.45	-0.14	-0.03	**-0.56**	0.03
28a	A2LERNLT	0.05	**0.76**	0.05	-0.10	-0.05	0.04	0.00
28b	A2PRACLT	0.09	**0.69**	0.22	-0.04	-0.13	-0.09	0.03
28c	A2NEWVOC	0.11	**0.39**	0.08	-0.32	-0.13	-0.11	0.26
28e	A2PHONIC	0.11	**0.73**	0.15	-0.09	0.02	-0.07	0.11
29a	A2CONVNT	0.07	**0.40**	-0.12	-0.36	0.01	-0.14	0.12
29b	A2RCGNZE	0.02	**0.76**	-0.11	-0.08	0.02	-0.02	0.04
29c	A2MATCH	0.09	**0.80**	0.00	-0.12	0.01	-0.09	0.05
29d	A2WRTNME	0.08	**0.54**	-0.04	-0.14	-0.19	-0.08	0.01
29e	A2RHYMNG	0.11	**0.40**	0.08	-0.32	-0.32	-0.18	0.04
28j	A2BASAL	0.02	-0.00	**0.76**	0.01	-0.06	-0.14	-0.02
28l	A2WRKBK	-0.29	0.23	**0.56**	-0.07	0.05	-0.04	0.08
28m	A2WRTWRD	0.06	0.14	**0.45**	0.03	-0.07	-0.35	0.17
29g	A2PREPOS	-0.11	0.16	0.04	**-0.44**	-0.37	-0.34	0.05
29h	A2MAINID	0.10	0.04	0.08	**-0.68**	-0.30	-0.22	0.01
29i	A2PREDIC	0.26	0.13	-0.04	**-0.72**	-0.14	-0.12	0.10
29j	A2TEXTCU	0.18	0.05	0.10	**-0.69**	-0.07	-0.27	0.10
29k	A2ORALID	0.17	0.24	-0.01	**-0.65**	0.17	-0.10	0.20
29l	A2DRCTNS	0.10	0.30	0.01	**-0.59**	0.18	-0.10	0.11
28i	A2READLD	*0.38	0.22	**0.41**	-0.27	0.02	-0.24	0.03
28k	A2SILENT	**0.43**	0.03	0.34	-0.22	0.25	-0.13	0.18
28n	A2INVENT	**0.74**	0.17	-0.03	-0.18	-0.01	-0.21	0.09
28o	A2CHSBK	**0.59**	0.19	0.15	-0.32	0.19	-0.09	0.16
28p	A2COMPOS	**0.68**	0.05	0.04	-0.07	-0.18	-0.30	0.14
28t	A2JRNL	**0.71**	0.05	-0.09	-0.14	-0.10	-0.21	0.00
28d	A2DICTAT	**0.46**	0.22	-0.01	-0.15	*-0.45	-0.11	0.03
28h	A2RETELL	0.30	0.23	0.18	-0.39	**-0.47**	-0.03	-0.01
28q	A2DOPROJ	**0.42**	-0.01	0.05	-0.32	*-0.21	-0.04	0.25
28r	A2PUBLSH	**0.53**	-0.01	-0.01	-0.02	*-0.38	-0.19	0.22
28s	A2SKITS	0.28	0.04	0.00	-0.15	**-0.50**	-0.09	0.22
28v	A2MXDGRP	0.18	0.09	-0.05	-0.19	0.02	-0.03	**0.73**
28w	A2PRTUTR	0.16	0.10	0.14	-0.11	-0.13	-0.16	**0.67**

NOTE: Each item's highest loading appears in boldface type.
* indicates item assigned to a factor other than the one on which it had its highest loading.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 2.—Reading and Language Arts Scales and Item Means: 1999

Scale	Item number	Variable name	Description	Mean score
Reading and writing skills	29f	A2SYLLAB	Reading multi-syllable words, like adventure	2.7
	29m	A2PNCTUA	Using capitalization and punctuation	4.1
	29n	A2COMPSE	Composing and writing complete sentences	3.5
	29o	A2WRTSTO	Composing and writing stories with an understandable beginning, middle, and end	2.2
	29p	A2SPELL	Conventional spelling	2.9
	29q	A2VOCAB	Vocabulary	4.3
	29r	A2ALPBTZ	Alphabetizing	2.1
	29s	A2RDFLNT	Reading aloud fluently	3.1
Phonics	28a	A2LRNLT	Work on learning the names of letters	5.8
	28b	A2PRACLT	Practice writing the letters of the alphabet	5.5
	28c	A2NEWVOC	Discuss new or difficult vocabulary	5.4
	28e	A2PHONIC	Work on phonics	5.7
	29a	A2CONVNT	Conventions of print	5.4
	29b	A2RCGNZE	Alphabet and letter recognition	5.8
	29c	A2MATCH	Matching letters to sounds	5.7
	29d	A2WRTNME	Writing own name (first and last)	5.5
	29e	A2RHYMNG	Rhyming words and word families	4.6
Didactic instruction	28j	A2BASAL	Read from basal reading texts	2.4
	28l	A2WRKBK	Work in a reading workbook or on a worksheet	4.1
	28m	A2WRTWRD	Write words from dictation, to improve spelling	2.9
Comprehension	29g	A2PREPOS	Common prepositions, such as over and under, up and down	3.9
	29h	A2MAINID	Identifying the main idea and parts of a story	4.1
	29i	A2PREDIC	Making predictions based on text	4.8
	29j	A2TEXTCU	Using context cues for comprehension	4.4
	29k	A2ORALID	Communicating complete ideas orally	5.3
	29l	A2DRCTNS	Remembering and following directions that include a series of actions	5.2
Student-centered language arts instruction	28d	A2DICTAT	Dictate stories to a teacher, aide, or volunteer	3.9
	28h	A2RETELL	Retell stories	4.2
	28q	A2DOPROJ	Do an activity or project related to a book or story	4.0
	28r	A2PUBLSH	Publish their own writing	2.4
	28s	A2SKITS	Perform plays and skits	2.2
Reading and writing activities	28i	A2READLD	Read aloud	4.7
	28k	A2SILENT	Read silently	4.1
	28n	A2INVENT	Write with encouragement to use invented spellings, if needed	4.6
	28o	A2CHSBK	Read books they have chosen for themselves	5.0
	28p	A2COMPOS	Compose and write stories or reports	3.3
	28t	A2JRNL	Write stories in a journal	3.7
Mixed-achievement grouping	28v	A2MXDGRP	Work in mixed-achievement groups on language arts activities	4.4
	28w	A2PRTUTR	Peer tutoring	3.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 3.—Reliabilities and Correlations for Reading and Language Arts Sum Scores and IRT Scores: 1999

Scale	Number of items	Alpha for sum score	Marginal reliability for IRT score	Correlation between sum and IRT score
Reading and writing skills	8	0.83	0.85	0.97
Phonics	9	0.75	0.68	0.90
Didactic instruction	3	0.50	0.53	0.99
Comprehension	6	0.77	0.82	0.94
Student-centered instruction	5	0.69	0.72	0.99
Reading and writing activities	6	0.77	0.82	0.96
Mixed-achievement grouping	2	0.48	0.51	0.99

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 4.—Factor Loadings for Mathematics Instructional Activities and Skills Items, Full Sample (n=2287): 1999

Item number	Variable name	Numbers and Geometry	Traditional practices and computation	Advanced numbers and operations	Measurement and advanced topics	Student-centered instruction/ grouping
31a	A2OUTLOU	**0.51**	-0.02	-0.28	0.15	0.26
31b	A2GEOMET	**0.57**	-0.08	-0.01	-0.05	0.30
31c	A2MANIPS	**0.55**	0.10	-0.10	-0.01	0.44
32a	A2QUANTI	**0.56**	0.12	-0.25	0.08	0.26
32b	A21TO10	**0.53**	0.37	-0.13	0.11	0.13
32f	A2SHAPES	**0.67**	0.16	-0.04	-0.09	-0.01
32g	A2IDQNTY	**0.55**	0.15	-0.24	-0.26	0.09
32h	A2SUBGRP	**0.73**	0.06	0.00	-0.28	0.10
32i	A2SZORDR	**0.76**	0.13	0.04	-0.26	0.05
32j	A2PTTRNS	**0.54**	-0.05	-0.25	-0.25	0.06
32u	A2ORDINL	*0.36	0.08	-0.23	**-0.38**	0.05
32c	A22S5S10	0.14	0.18	**-0.60**	-0.08	0.16
32d	A2BYD100	0.14	0.06	**-0.64**	-0.07	0.07
32n	A2PLACE	-0.02	0.14	**-0.63**	-0.19	0.07
32o	A2TWODGT	0.11	0.14	**-0.64**	-0.15	0.11
32p	A23DGT	0.01	-0.01	**-0.75**	-0.14	0.02
31k	A2MTHSHT	0.08	**0.61**	0.07	0.13	0.00
31l	A2MTHTXT	-0.04	**0.63**	0.15	0.03	0.01
31m	A2CHLKBD	0.05	**0.61**	-0.03	-0.03	0.27
32e	A2W12100	0.12	**0.52**	-0.27	-0.01	0.07
32k	A2REGZCN	0.22	**0.44**	-0.24	-0.33	0.00
32l	A2SNGDGT	0.19	**0.58**	-0.34	-0.13	0.20
32m	A2SUBSDG	0.12	**0.60**	-0.27	-0.19	0.20
31d	A2MTHGME	**0.51**	0.01	-0.07	-0.09	*0.45
31f	A2MUSMTH	0.29	-0.06	0.09	-0.18	**0.37**
31g	A2CRTIVE	0.28	-0.02	0.07	-0.22	**0.43**
31i	A2EXPMTH	0.14	0.27	-0.23	-0.25	**0.52**
31n	A2PRTNRS	0.21	0.28	-0.04	-0.21	**0.63**
31o	A2REALLI	0.13	0.15	-0.14	-0.30	**0.60**
31p	A2MXMATH	0.14	-0.01	-0.13	-0.06	**0.63**
31q	A2PEER	0.12	0.15	-0.11	-0.11	**0.58**
31h	A2RULERS	0.38	0.08	0.04	**-0.39**	0.27
32r	A2GRAPHS	0.26	-0.13	-0.29	**-0.54**	0.17
32s	A2DATACO	0.22	-0.14	-0.28	**-0.58**	0.17
32t	A2FRCTNS	0.14	0.36	-0.10	**-0.53**	0.02
32v	A2ACCURA	0.29	0.23	-0.04	**-0.60**	0.10
32w	A2TELLTI	0.22	**0.54**	-0.12	*-0.42	0.02
32x	A2ESTQNT	0.24	0.02	-0.16	**-0.62**	0.25
32bb	A2PRBTY	-0.02	0.08	-0.09	**-0.56**	0.27
32cc	A2EQTN	-0.07	0.29	-0.21	**-0.46**	0.26

NOTE: Each item's highest loading appears in boldface type.
* indicates item assigned to a factor other than the one on which it had its highest loading.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 5.—Mathematics Scales and Item Means: 1999

Scale	Item number	Variable name	Description	Mean score
Numbers and geometry	31a	A2OUTLOU	Count out loud	5.6
	31b	A2GEOMET	Work with geometric manipulatives	4.3
	31c	A2MANIPS	Work with counting manipulatives to learn basic operations	4.8
	32a	A2QUANTI	Correspondence between number and quantity	5.1
	32b	A21TO10	Writing all numbers between 1 and 10	4.6
	32f	A2SHAPES	Recognizing and naming geometric shapes	4.1
	32g	A2IDQNTY	Identifying relative quantity (e.g., equal, most, less, more)	4.3
	32h	A2SUBGRP	Sorting objects into subgroups according to a rule	3.9
	32i	A2SZORDR	Ordering objects by size or other properties	3.7
	32j	A2PTTRNS	Making, copying, or extending patterns	4.3
	32u	A2ORDINL	Ordinal numbers (e.g., first, second, third)	3.9
Advanced numbers and operations	32c	A22S5S10	Counting by 2s, 5s, and 10s	4.1
	32d	A2BYD100	Counting beyond 100	3.0
	32n	A2PLACE	Place value	3.0
	32o	A2TWODGT	Reading two-digit numbers	4.6
	32p	A23DGT	Reading three-digit numbers	2.7
Traditional practices and computation	31k	A2MTHSHT	Do math worksheets	4.2
	31l	A2MTHTXT	Do math problems from their textbooks	2.1
	31m	A2CHLKBD	Complete math problems on the chalkboard	2.9
	32e	A2W12100	Writing all numbers between 1 and 100	2.3
	32k	A2REGZCN	Recognizing the value of coins and currency	3.4
	32l	A2SNGDGT	Adding single-digit numbers	4.0
	32m	A2SUBSDG	Subtracting single-digit numbers	3.5
Student-centered math instruction	31d	A2MTHGME	Play math-related games	4.5
	31f	A2MUSMTH	Use music to understand math concepts	2.7
	31g	A2CRTIVE	Use creative movement or creative drama to understand math concepts	2.6
	31i	A2EXPMTH	Explain how a math problem is solved	3.9
	31n	A2PRTNRS	Solve math problems in small groups or with a partner	3.5
	31o	A2REALLI	Work on math problems that reflect real-life situations	3.9
Mixed-achievement grouping	31p	A2MXMATH	Work in mixed achievement groups on math activities	4.1
	31q	A2PEER	Peer tutoring	3.1
Measurement and advanced topics	31h	A2RULERS	Work with rulers, measuring cups, spoons, or other measuring instruments	2.9
	32r	A2GRAPHS	Reading simple graphs	3.7
	32s	A2DATACO	Performing simple data collection and graphing	3.2

Table 5 (cont'd)

Table 5. (cont'd)—Mathematics Scales and Item Means: 1999

Scale	Item number	Variable name	Description	Mean score
	32t	A2FRCTNS	Fractions (e.g., recognizing that 1/4 of a circle is colored)	2.0
	32v	A2ACCURA	Using measuring instruments accurately	2.6
	32w	A2TELLTI	Telling time	3.3
	32x	A2ESTQNT	Estimating quantities	2.0
	32bb	A2PRBTY	Estimating probability	1.7
	32cc	A2EQTN	Writing math equations to solve word problems	1.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 6.—Coefficient Alpha Reliabilities for Mathematics Scales: 1999

Scale	Number of items	Alpha
Numbers and geometry	11	0.83
Advanced numbers and operations	5	0.75
Traditional practices and computation	7	0.73
Student-centered math instruction	6	0.74
Mixed-achievement grouping	2	0.56
Measurement and advanced topics	9	0.81

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 7.—Science and Social Studies Scales and Item Means: 1999

Scale	Item number	Variable name	Description	Mean score
Science skills	34a	A2BODY	Human body	2.8
	34b	A2PLANT	Plants and animals	3.3
	34c	A2DINOSR	Dinosaurs and fossils	2.1
	34d	A2SOLAR	Solar system and space	2.0
	34e	A2WTHER	Weather (e.g., rainy, sunny)	5.0
	34f	A2TEMP	Understand and measure temperature	2.7
	34g	A2WATER	Water	2.4
	34h	A2SOUND	Sound	2.1
	34i	A2LIGHT	Light	2.0
	34j	A2MAGNET	Magnetism and electricity	1.8
	34k	A2MOTORS	Machines and motors	1.5
	34l	A2TOOLS	Tools and their uses	2.0
	34m	A2HYGIEN	Health, safety, nutrition, and personal hygiene	4.1
	34u	A2SCMTHD	Scientific method	1.9
Social studies skills	34n	A2HISTOR	Important figures and events in American history	2.6
	34o	A2CMNITY	Community resources (e.g., grocery store, police)	2.9
	34p	A2MAPRD	Map-reading skills	2.1
	34q	A2CULTUR	Different cultures	2.9
	34r	A2LAWS	Reasons for rules, laws, and government	3.0
	34s	A2ECOLOG	Ecology	2.5
	34t	A2GEORPH	Geography	2.4
	34v	A2SOCPRO	Social-problem solving	3.7

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 8.—Factor Loadings for Science Skills Items, Full Sample (n=2451): 1999

Item number	Variable name			
34g	A2WATER	**0.67**	0.13	-0.36
34h	A2SOUND	**0.76**	0.11	-0.29
34i	A2LIGHT	**0.80**	0.09	-0.26
34j	A2MAGNET	**0.66**	-0.01	-0.32
34k	A2MOTORS	**0.76**	0.00	-0.09
34l	A2TOOLS	**0.69**	0.13	-0.21
34u	A2SCMTHD	**0.49**	0.28	-0.07
34f	A2TEMP	**0.50**	0.48	0.08
34e	A2WTHER	0.02	**0.81**	-0.09
34m	A2HYGIEN	0.11	**0.62**	-0.42
34a	A2BODY	0.28	0.28	**-0.67**
34b	A2PLANT	0.15	0.27	**-0.75**
34c	A2DINOSR	0.33	-0.08	**-0.68**
34d	A2SOLAR	0.46	-0.06	**-0.59**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 9.—Factor Loadings for Social Studies Skills Items, Full Sample (n=2719): 1999

Item number	Variable name		
34n	A2HISTOR	**0.80**	0.11
34o	A2CMNITY	**0.80**	0.09
34p	A2MAPRD	**0.54**	0.41
34q	A2CULTUR	**0.68**	0.36
34r	A2LAWS	0.30	**0.66**
34s	A2ECOLOG	0.40	**0.67**
34t	A2GEORPH	0.46	**0.62**
34v	A2SOCPRO	-0.05	**0.81**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 10.—Factor Loadings for Computer Use Items, Full Sample (n=2928): 1999

Item number	Variable name		
33a	A2LRNRD	**0.85**	0.18
33b	A2LRNMTH	**0.86**	0.21
33e	A2LRNKEY	**0.62**	0.22
33f	A2LRNART	**0.59**	0.38
33h	A2LRNGMS	**0.79**	0.15
33c	A2LRNSS	0.36	**0.75**
33d	A2LRNSCN	0.39	**0.74**
33g	A2LRNMSC	0.07	**0.68**
33i	A2LRNLAN	0.01	**0.62**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 11.—Computer Use Item Means: 1999

Item number	Variable name	Description	Mean score
33a	A2LRNRD	To learn reading, writing, or spelling	3.7
33b	A2LRNMTH	To learn math	3.5
33c	A2LRNSS	To learn social studies concepts	2.0
33d	A2LRNSCN	To learn science concepts	2.1
33e	A2LRNKEY	To learn keyboarding skills	2.7
33f	A2LRNART	To create art	2.6
33g	A2LRNMSC	To compose and/or perform music	1.3
33h	A2LRNGMS	For enjoyment (e.g., games)	3.7
33i	A2LRNLAN	To access information (e.g., to connect to Internet or local network)	1.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 12.—Coefficient Alpha Reliabilities for Science, Social Studies, and Computer Use Scales: 1999

Scale	Number of items	Alpha
Science skills	14	0.86
Social studies skills	8	0.82
Science and social studies skills	22	0.90
Computer use	9	0.85

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 13.—Correlations Among Instruction Scales (n=2950): 1999

	RWS	RWA	RPHN	RDID	RCOM	RSTU	RGRP	MNUM	MADV	MTRD	MSTU	MGRP	MMEA	SCI	SOC	COMP
RWS	1.00															
RWA	.52	1.00														
RPHN	.32	.37	1.00													
RDID	.35	.16	.21	1.00												
RCOM	.52	.48	.48	.17	1.00											
RSTU	.41	.61	.40	.11	.50	1.00										
RGRP	.30	.41	.28	.11	.35	.37	1.00									
MNUM	.34	.37	.57	.15	.49	.48	.30	1.00								
MADV	.37	.30	.25	.13	.28	.23	.21	.32	1.00							
MTRD	.44	.19	.30	.51	.26	.20	.13	.36	.32	1.00						
MSTU	.43	.47	.42	.19	.52	.56	.40	.59	.29	.39	1.00					
MGRP	.33	.35	.29	.13	.35	.36	.67	.37	.23	.23	.51	1.00				
MMEA	.47	.45	.33	.18	.46	.47	.30	.55	.43	.42	.56	.36	1.00			
SCI	.32	.24	.24	.16	.33	.32	.24	.42	.17	.31	.39	.30	.49	1.00		
SOC	.35	.31	.26	.12	.37	.32	.27	.40	.23	.25	.40	.32	.48	.69	1.00	
COMP	.23	.28	.19	.10	.26	.34	.19	.29	.15	.19	.36	.21	.36	.24	.22	1.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 14.—Evaluation Item Means: 1999 (Question 3 from Fall Questionnaire B: How important is each of the following in evaluating the children in your class(es)?)

Scale	Item number	Variable name	Description	Mean score
Evaluation based on academic achievement	3a	B1TOCLAS	Individual child's achievement relative to the rest of the class	2.5
	3b	B1TOSTND	Individual child's achievement relative to local, state, or professional standards	2.6
Evaluation based on other factors	3d	B1EFFO	Effort	3.6
	3e	B1CLASPA	Class participation	3.3
	3f	B1ATTND	Daily attendance	3.5
	3g	B1BEHVR	Classroom behavior or conduct	3.5
	3h	B1COPRTV	Cooperativeness with other children	3.5
	3i	B1FLLWDR	Ability to follow directions	3.6
Evaluation based on improvement	3c	B1IMPRVM	Individual improvement or progress over past performance	3.6

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 15.—Factor Loadings for Evaluation Items, Full Sample (n=3123): 1999

Item number	Variable name	Evaluation: Social/Behavioral	Evaluation: Achievement
3d	B1EFFO	**0.68**	0.01
3e	B1CLASPA	**0.70**	0.13
3f	B1ATTND	**0.67**	0.18
3g	B1BEHVR	**0.79**	0.12
3h	B1COPRTV	**0.76**	0.02
3i	B1FLLWDR	**0.71**	0.13
3a	B1TOCLAS	0.12	**0.81**
3b	B1TOSTND	0.05	**0.83**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 16.—Readiness Item Means: 1999 (Question 7 from Fall Questionnaire B: How important do you believe the following characteristics are for a child to be ready for kindergarten?)

Scale	Item number	Variable name	Description	Mean score
Academic criteria	7b	B1CNT20	Can count to 20 or more	2.6
	7d	B1PRBLMS	Has good problem-solving skills	3.2
	7e	B1PENCIL	Is able to use pencils and paint brushes	3.2
	7j	B1ALPHBT	Knows most of the letters of the alphabet	2.8
	7l	B1IDCOLO	Identifies primary colors and shapes	3.1
Social and behavioral criteria	7a	B1FNSHT	Finishes tasks	3.5
	7c	B1SHARE	Takes turns and shares	3.9
	7f	B1NOTDSR	Is not disruptive of the class	4.0
	7h	B1SENSTI	Is sensitive to other children's feelings	3.7
	7i	B1SITSTI	Sits still and pays attention	3.7
	7k	B1FOLWDR	Can follow directions	4.0
	7m	B1COMM	Communicates needs, wants, and thoughts verbally in primary language	4.1
Other	7g	B1ENGLAN	Knows the English language	3.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 17.—Factor Loadings for Readiness Items, Full Sample (n=3060): 1999

Item number	Variable name	Readiness: Academic	Readiness: Social/Behavioral
7b	B1CNT20	**0.86**	0.10
7d	B1PRBLMS	**0.49**	0.42
7e	B1PENCIL	**0.68**	0.34
7j	B1ALPHBT	**0.88**	0.15
7l	B1IDCOLO	**0.80**	0.23
7a	B1FNSHT	0.37	**0.55**
7c	B1SHARE	0.23	**0.72**
7f	B1NOTDSR	0.14	**0.74**
7h	B1SENSTI	0.15	**0.69**
7i	B1SITSTI	0.30	**0.69**
7k	B1FOLWDR	0.30	**0.72**
7m	B1COMM	0.15	**0.62**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 18.—Coefficient Alpha Reliabilities for Evaluation and Readiness Scales: 1999

Scale	Number of items	Alpha
Evaluation: Achievement	2	0.55
Evaluation: Social/Behavioral	6	0.82
Readiness: Academic	5	0.85
Readiness: Social/Behavioral	7	0.85

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 19.—Correlations Among Evaluation and Readiness Scales (n=3198): 1999

	Eval_ach	Eval_soc	Readi_acad	Readi_soc
Eval_ach	1.00			
Eval_soc	0.22	1.00		
Readi_acad	0.22	0.17	1.00	
Readi_soc	0.21	0.31	0.60	1.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 20.—School Climate Item Means: 1999

Scale	Item number	Variable name	Description	Mean score
Leadership	13b	B1MISSIO	There is broad agreement among the entire school faculty about the central mission of the school	4.0
	13c	B1ALLKNO	The school administrator knows what kind of school he/she wants and has communicated it to staff	4.1
	13d	B1PRESSU	The school administrator deals effectively with pressures from outside the school (for example, budget, parents, school board) that might otherwise affect my teaching	3.9
	13e	B1PRIORI	The school administrator sets priorities, makes plans, and sees that they are carried out	4.0
	13f	B1ENCOUR	The school administrator's behavior toward the staff is supportive and encouraging	4.1
	11	B1SCHPLC	Teachers' influence over school policy	3.4
Efficacy	10b	B1MISBHV	The level of child misbehavior in this school interferes with my teaching	2.1
	10c	B1NOTCAP	Many of the children I teach are not capable of learning the material I am supposed to teach them	1.9
	10f	B1PAPRWR	Routine administrative duties and paperwork interfere with my job of teaching	3.0
	12	B1CNTRLC	Teachers' influence over classroom	4.4
Professional climate	10a	B1SCHSPR	Staff members in this school generally have school spirit	4.1
	10d	B1ACCPTD	I feel accepted and respected as a colleague by most staff members	4.4
	10e	B1CNTNLR	Teachers in this school are continually learning and seeking new ideas	4.3
	10g	B1PSUPP	Parents are supportive of school staff	4.0
	13a	B1STNDLO	The academic standards at this school are too low	1.8
Satisfaction	14a	B1ENJOY	I really enjoy my present teaching job	4.5
	14b	B1MKDIFF	I am certain I am making a difference in the lives of the children I teach	4.8
	14c	B1TEACH	If I could start over, I would choose teaching again as my career	4.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 21.—Factor Loadings for Climate Items, Full Sample (n=2794): 1999

Item number	Variable name	Leadership	Efficacy	Prof. Climate
13b	B1MISSIO	**0.51**	-0.05	0.45
13c	B1ALLKNO	**0.84**	0.01	0.19
13d	B1PRESSU	**0.86**	-0.16	0.14
13e	B1PRIORI	**0.86**	-0.05	0.16
13f	B1ENCOUR	**0.80**	-0.18	0.17
11	B1SCHPLC	**0.42**	-0.33	0.25
10b	B1MISBHV	-0.11	**0.62**	-0.14
10c	B1NOTCAP	-0.03	**0.60**	-0.19
10f	B1PAPRWR	-0.27	**0.61**	0.07
12	B1CNTRLC	0.12	**-0.64**	0.09
10a	B1SCHSPR	0.27	-0.15	**0.64**
10d	B1ACCPTD	0.16	-0.14	**0.72**
10e	B1CNTNLR	0.22	0.04	**0.80**
10g	B1PSUPP	0.17	-0.40	**0.50**
13a	B1STNDLO	-0.27	0.21	**-0.50**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 22.—Coefficient Alpha Reliabilities for School Climate and Satisfaction Scales: 1999

Scale	Number of items	Alpha
Leadership	6	0.84
Efficacy	4	0.52
Professional climate	5	0.74
Satisfaction	3	0.72

NOTE: Negatively-worded items were reverse-coded.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 23.—Correlations Among School Climate and Satisfaction Scales (n=2922): 1999

	Lead	Efficacy	Climate	Satisf.
Lead	1.00			
Control	0.37	1.00		
Climate	0.58	0.38	1.00	
Satisf.	0.30	0.31	0.32	1.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 24.—Factor Loadings for Resource Items, Full Sample (n=2422): 1999

Item number	Variable name	Instructional Resources	Computers	Facilities
26d	A2MANIPU	**0.49**	-0.23	0.22
26e	A2AUDIOV	**0.56**	-0.36	0.09
26f	A2VIDEO	**0.57**	-0.38	0.15
26i	A2PAPER	**0.67**	-0.08	0.09
26j	A2DITTO	**0.66**	-0.08	0.17
26k	A2ART	**0.73**	-0.13	0.22
26m	A2RECRDS	**0.66**	-0.23	0.14
26g	A2COMPEQ	0.10	**-0.92**	0.09
26h	A2SOFTWA	0.14	**-0.92**	0.06
26p	A2HEATAC	0.09	-0.25	**0.62**
26q	A2CLSSPC	0.07	-0.06	**0.79**
26r	A2FURNIT	0.30	-0.05	**0.73**

NOTE: Each item's highest loading appears in boldface type.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 25.—Resource Item Means: 1999 (Question 26 from Spring Questionnaire A: In general, how adequate is each of the following for your class(es)? (scale ranges from 2 to 5))

Scale	Item number	Variable name	Description	Mean score
Instructional resources	26d	A2MANIPU	Manipulatives (e.g., blocks, puzzles)	4.7
	26e	A2AUDIOV	Audiovisual equipment (e.g., VCR)	4.5
	26f	A2VIDEO	Videotapes and films	4.4
	26i	A2PAPER	Paper and pencils	4.8
	26j	A2DITTO	Ditto or photocopier equipment	4.5
	26k	A2ART	Art materials, paints, clays	4.6
	26m	A2RECRDS	Musical recordings	4.3
Computers	26g	A2COMPEQ	Computer equipment	4.1
	26h	A2SOFTWA	Computer software	4.0
Facilities	26p	A2HEATAC	Heat and air-conditioning	4.2
	26q	A2CLSSPC	Classroom space	4.2
	26r	A2FURNIT	Child-sized furniture	4.6

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 26.—Coefficient Alpha Reliabilities for Resource Scales: 1999

Scale	Number of items	Alpha
Resources: Instructional	7	0.73
Resources: Computers	2	0.93
Resources: Facilities	3	0.58

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Table 27.—Correlations Among Resource Scales (n=2962): 1999

	Res_instr	Res_comp	Res_facil	Centers
Res_instr	1.00			
Res_comp	0.37	1.00		
Res_facil	0.39	0.15	1.00	
Centers	0.10	0.14	0.01	1.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Restricted-Use data files.

Figure 1.—Distributions of IRT and Sum Scores for *Student-Centered Instruction*

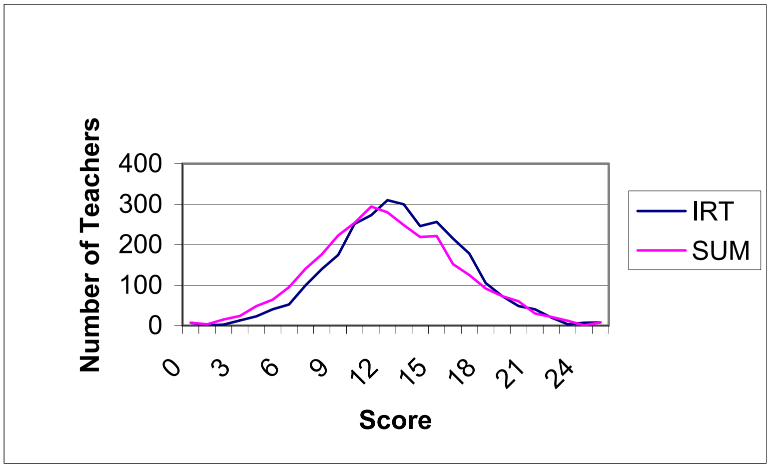
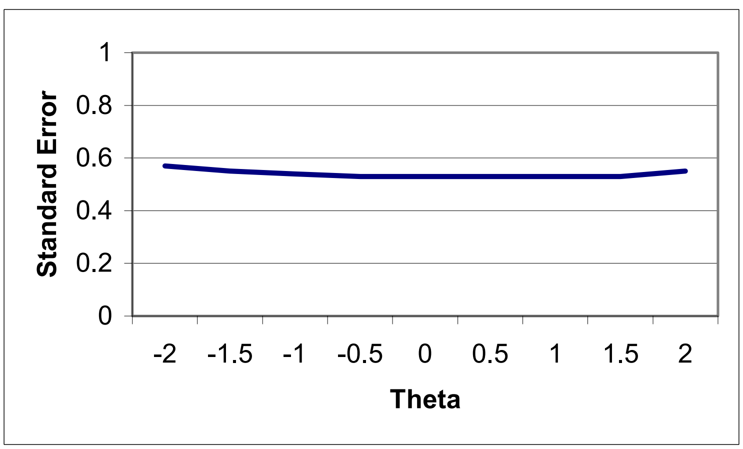


Figure 2.—Standard Errors of *Student-Centered Instruction* IRT Scores



Note: Theta is a standardized IRT score for each teacher.

Figure 3.—Distributions of IRT and Sum Scores for *Reading and Writing Activities*

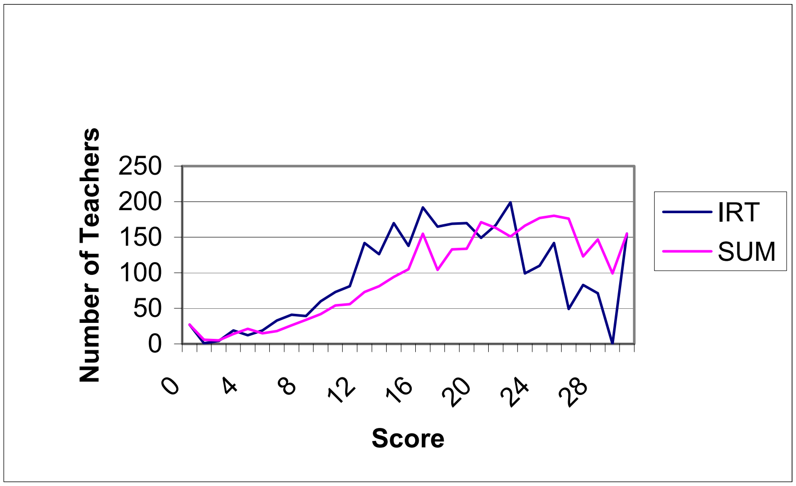


Figure 4.—Standard Errors of *Reading and Writing Activities* IRT Scores

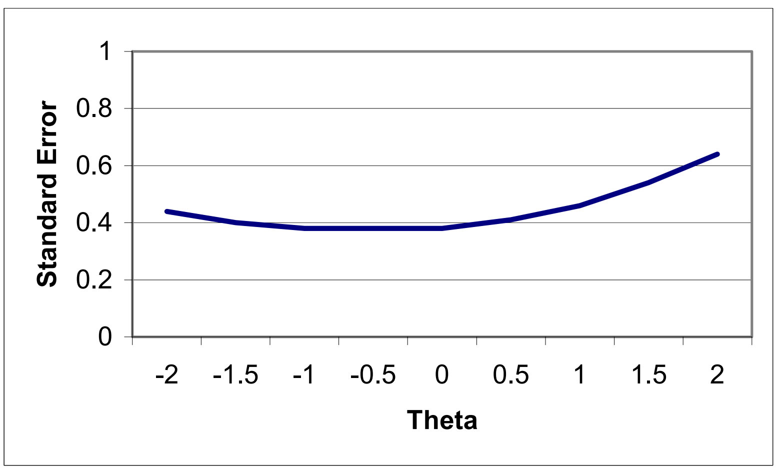


Figure 5.—Distributions of IRT and Sum Scores for *Phonics*

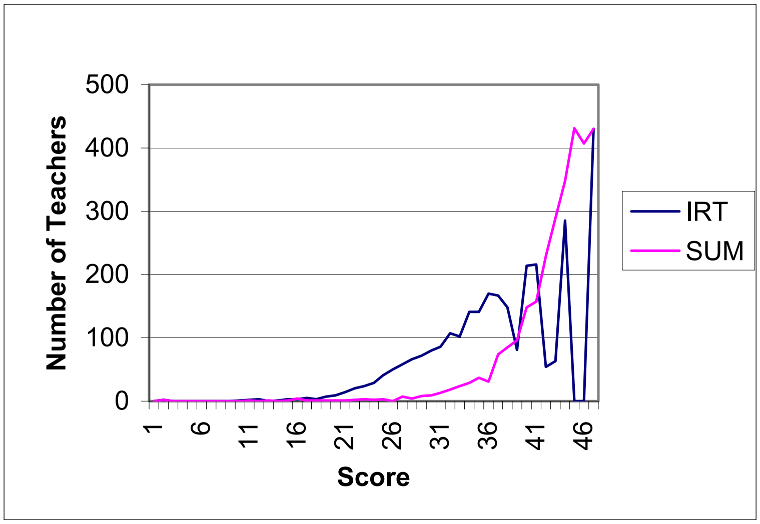
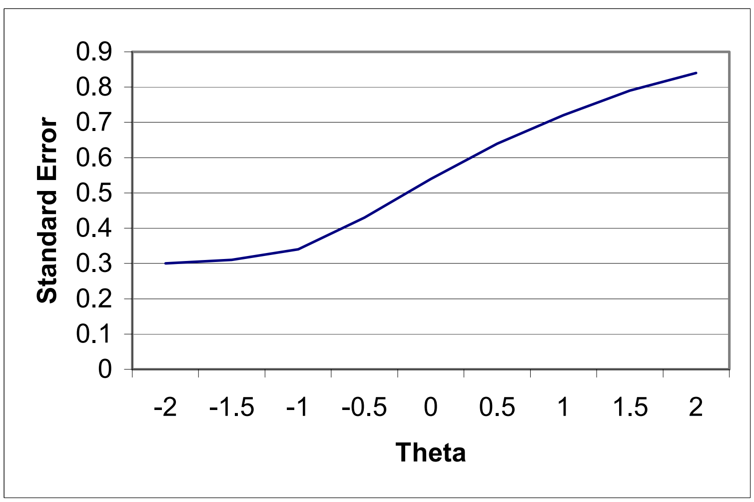


Figure 6.—Standard Errors of *Phonics* IRT Scores



REFERENCES

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up* (MR-1145-EDU). Santa Monica, CA: RAND.

Brewer, D.J. & Goldhaber, D.D. (1996). Educational achievement and teacher qualifications: New evidence from microlevel data. In B.S. Cooper & S. Speakman (Eds.), *Advances in Educational Productivity* (6th ed., pp.243-264). Greenwich, CT: Jai Press.

Bryk, A., Lee, V. & Holland, P. (1993). Catholic schools and the common good. Cambridge, MA: Harvard University.

Bryk, A. S., Sebring, P. B., Kerbow, D., Rollow, S., & Easton, J. Q. (1998). Charting Chicago school reform: Democratic localism as a lever for change*. Boulder, CO: Westview.*

Chan, K. S. (2001). *Scoring methods and measurement properties of health status measures: An empirical comparison of summative and item response theory scoring for the SF-36 Physical Functioning Scale*. Doctoral Dissertation, Johns Hopkins University.

Choi, N., Fuqua, D. R., & Griffin, B. W. (2001). Exploratory analysis of the structure of scores from the Multidimensional Scales of Perceived Self-Efficacy. *Educational and Psychological Measurement, 61,* 475-489.

Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance: Mathematics reform in California* (CPRE Policy Brief). Philadelphia: Consortium for Policy Research in Education.

Core Knowledge Foundation. (1999) *Core knowledge sequence* (9th ed.). Charlottesville, VA: Author.

Coste, J., Fermanian, J., & Venot, A. (1995). Methodological and statistical problems in the construction of composite measurement scales: A survey of six medical and epidemiological journals. *Statistics in Medicine, 14,* 331-345.

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

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

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*, 465-498.

Fetler, M. (1999) High school characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9).

Goldhaber, D. and Brewer, D. (2000) Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22: 129-145.

Hambleton, R. K., Swaminathan, H., Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.

Hamilton, L. S., McCaffrey, D., Klein, S. P., Stecher, B. M., Robyn, A., & Bugliari, D. (2001). Teaching Practices and Student Achievement: Evaluating Classroom-Based Education Reforms (DRU-2603). Santa Monica, CA: RAND.

Hanushek, E. (1986). The economics of schooling: Production and efficiency in the public schools. *Journal of Economic Literature*, *24*, 1141-1178.

Hanushek, E. & Pace, R. (1995). Who chooses to teach (and why). *Economics of Education Review*, *14*(2), 101-117.

Hawk, P., Coble, C., and Swanson, M. (1985) Certification: It does matter. *Journal of Teacher Education*. 36(3), 13-15.

Hedges, L., Laine, R., & Greenwald, R. (1994). A meta-analysis of the effects of differential school inputs on student outcomes. *Educational Researcher*, *23* (3), 5-14.

Hoy, W.K. & Woolfolk, A.E. (1993). Teachers' sense of efficacy and the organizational health of schools. *Elementary School Journal, 93,* 355-72.

Louis, K. S., and H. M. Marks. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education, 106,* 532-575.

Klein, S., Hamilton, L., McCaffrey, D., Stecher, B., Robyn, A., & Burroughs, D. (2000). *Teaching practices and student achievement: Report of first-year results from the Mosaic study of Systemic Initiatives in mathematics and science* (MR-1233-EDU). Santa Monica, CA: RAND.

Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Mayer, R. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis, 20,* 53-73.

Mayer, R. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21,* 29-45.

Miller, J.W., McKenna, B.A., & McKenna, M.C. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49* (3), 165-176.

Monk, D. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 12(2): 125-145.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council, Committee on the Prevention of Reading Difficulties in Young Children (1999). *Starting out Right: A guide to promoting children's reading success.* Washington, DC: National Academy Press.

Newmann, F. M., & Associates (eds., 1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco: Jossey Bass.

Rathbun, A. H., Walston, J. T., & Germino Hausken, E. (2000). *Kindergarten teachers' use of developmentally appropriate practices: Results from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (1998). *Teachers, schools, and academic achievemen*t. NBER Working Paper No. W6691.

Samejima F (1969) Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research Center.

Smerdon, B. A., Burkam, D. T., & Lee, V. E. (1999). Access to constructivist and didactic teaching: Who gets it? Where is it practiced? *Teachers College Record*, *101*, 5-34.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677-680.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics.* New York: HarperCollins College Publishers.

Taylor, N. Z., Denton, K. L., & West, J. (2001). *A new teaching tool: Computers and technology in the classroom.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56,* 197-208.

U.S. Department of Education, National Center for Education Statistics (2001). *ECLS-K base year public-use data file user's manual* (NCES 2001-029). Washington, DC: NCES.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician, 47,* 65-72.

Weiss, I. R., Montgomery, D. L., Ridgway, C. J., & Bond, S. L. (1998). *Local Systemic Change through Teacher Enhancement: Year three cross-site report.* Chapel Hill, NC: Horizon Research, Inc.

Wenglinsky, H. (1997). When money matters: *How educational expenditures improve student performance and how they don't* (ETS Policy Information Report). Princeton, NJ: Educational Testing Service.

West, J., & Denton, K. L. (2001). *Vulnerable children and their families: What schools offer kindergartners and their families.* Paper presented at the Annual Meeting of the Society for Research in Child Development, Minneapolis.

Wright, S. P., Horn, S. P. & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-67.9

Xue, Y., Meisels, S., Lee, V. E., Burkam, D. T., & Dorfman, A. (2001). *Differential effects of phonics instruction on children's literacy learning in kindergarten.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle.