**Current Issue:**
**Volume 13 Number 2**

**Past Issues**

**Beyond This Issue**

**Search ECRP**

Sign up for the
**Email Newsletter**

**About ECRP**

**Sponsoring ECRP**

**Information for Authors**

**ECRP Survey**

Send comments
to the ECRP Editor.

**Volume 13 Number 2**
©The Author(s) 2011

# The Starting Line: Developing a Structure for Teacher Ratings of Students' Skills at Kindergarten Entry

## Jessica Goldstein & D. Betsy McCoach
### University of Connecticut

## Abstract

Developmentally appropriate, psychometrically sound instruments are needed to assess young children and evaluate learning programs. In the United States, little guidance exists on the development and use of large-scale assessments that cover the broad range of skills that encompass young children's development. In 2005 and 2006, the State of Connecticut passed legislation requiring the implementation of a statewide developmentally appropriate assessment that "measures a child's level of preparedness for kindergarten." In response to this legislation, the Connecticut State Department of Education developed a Kindergarten Entrance Inventory. The Inventory was designed to provide a statewide snapshot of the skills that children demonstrate, based on teachers' observations, at the beginning of the kindergarten year. This article investigates teacher ratings of children's skills at kindergarten entry in one large urban district using a series of exploratory and confirmatory factor analyses. Analyses indicate that readiness evaluations should address the following skills: expressive language, receptive language, responses to stories, familiarity with books, familiarity with letters, emergent writing, counting, shapes and patterns, measurement, fine motor skills, gross motor skills, conflict resolution, social engagement, engagement with self-selected activities, and creative skills.

## Introduction

Although annual testing requirements mandated in the *No Child Left Behind Act of 2001* (NCLB) begin in third grade, educators in the United States are placing a renewed emphasis on education in the primary grades because it serves as the foundation for all future learning. Measurement of young children's educational development is a critical piece of any comprehensive assessment system, yet it differs a great deal from the measurement protocols used with older children. Scott-Little, Kagan, and Clifford (2003) suggest that young children learn in a manner that is more episodic than older students and that multiple means of assessment are necessary to gain a full understanding of their knowledge. For young children, a single assessment administered at one point in time cannot accurately reflect their development. Moreover, the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) have established boundaries on appropriate uses of assessments in early childhood. These guidelines state that the appropriate use of assessments in early childhood is to guide teaching and learning, to identify children who may require focused interventions, and to improve educational programs and development interventions (NAEYC & NAECS/SDE, 2009). From a measurement perspective, it is clear that that these divergent objectives require unique assessment tools, and standardized measures for this population are not readily available.

Developmentally appropriate, psychometrically sound instruments are needed to monitor young children and evaluate the effectiveness of their early childhood learning programs. Yet in the research literature and in practice, little guidance exists on the development and use of large-scale assessments that address children's

emotional, cognitive, and physical development. This paper describes an empirical investigation of the structure of teacher ratings of students' skills at kindergarten entry based on one implementation of a state measure. Though the results of the study have implications for the validity of the current instrument, we believe subscales scores from this measure can be used as a reporting structure for similar instruments designed to assess kindergarten students' skills.

## Understanding Kindergarten Students' Skills

The creation of two national data sets as well as growing interest in the instruction and assessment of young children have spawned a small body of research to describe the skills that students demonstrate at the start of the kindergarten year. The U.S. Department of Education's National Center for Education Statistics (NCES) developed a data set called the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) that looks at children's health, development, and education during the formative years from birth through kindergarten entry. Denton Flanagan and McPhee (2009) found that upon kindergarten entry, children born in 2001 demonstrated reading and mathematics knowledge and skills that varied by their race/ethnicity, family type, poverty status, primary home language, and their primary early care and education setting the year prior to kindergarten. Specifically, White and Asian children had higher reading and mathematics assessment scores than did Black, Hispanic, or American Indian/Alaska Native children. Also, children in households with two parents, with incomes at or above the poverty threshold, or with English as a primary home language had higher reading and mathematics scores than their counterparts. The authors also found that children who had participated in regular early care and education arrangements the year prior to kindergarten scored higher on the reading and mathematics assessments than children who had not. Similar patterns were found for children's fine motor skills; children with higher scores on fine motor skill assessments tended to be female, White or Asian, living in two-parent households, living in households with incomes at or above the poverty threshold, and had participated in regular early care and education arrangements the year prior to kindergarten.

An earlier but similar study, the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), followed a nationally representative sample of 22,000 kindergartners from the fall of 1998 through their fifth-grade year. West, Denton, and Germino-Hausken (2000) reported on students' skills at kindergarten entry. In early literacy, 66% were proficient in recognizing their letters, 29% were proficient in understanding beginning sounds, and about 17% were proficient in understanding ending sounds. In math, nearly all kindergartners were proficient in identifying numbers and shapes, 58% were proficient in understanding relative size, and 20% were proficient in understanding ordinal sequence. With regard to social skills, teachers reported that about 75% of first-time kindergartners were accepting of peer ideas and were able to form friendships. Of the students in the sample, teachers reported that 71% of first-time kindergartners persisted at tasks often or very often, 75% seemed eager to learn, and 66% were able to pay attention most of the time. Rathbun and West (2004) used ECLS-K data to describe children's gains in reading and mathematics from the start of kindergarten through third grade.

In addition to describing students' skills, ECLS-K includes data on teacher perceptions of kindergarten readiness. Lin, Lawrence, and Gorell (2003) conducted one such study. In defining readiness, kindergarten teachers tended to emphasize the social demands of schooling over academic skill development. Specifically, readiness definitions centered on a child's social behaviors such as "tells wants and thoughts," "not disruptive of the class," "follows directions," and "takes turns and shares." Less frequently mentioned social skills were "sits still and alert," "finishes tasks," "has problem-solving skills," and "is sensitive to others." In their study, teachers were less likely to include more academic skills such as "counts to 20 or more," "knows most of the alphabet," "names colors and shapes," and "uses pencil and brushes."

In addition to the design of ECLS-K and publications that followed, other early childhood experts have attempted to bring a common language on early development and kindergarten readiness to the field. Such efforts are grounded in a multidimensional perspective of development that includes five dimensions: (1) physical and motor development, (2) social and emotional development, (3) approaches toward learning (i.e., creativity, initiative, attitudes toward learning, task mastery), (4) language, and (5) cognition and general knowledge (Kagan, Moore, & Bredekamp, 1995; Love, 2001; Meisels, 1999). This multifaceted structure accounts for the contributions of families and early education programs to children's development and emphasizes a child's orientation toward learning and being part of a group. Academic knowledge is viewed as only one component of a broad, diverse skill set.

Research suggests that kindergarten teachers support this view. One study found that the top three qualities that public school kindergarten teachers consider essential for school readiness are that a child be physically healthy, rested, and well nourished; be able to communicate needs, wants, and thoughts verbally; and be enthusiastic and curious in approaching new activities (Heaviside & Farris, 1993). A decade later, further research confirmed that teacher perceptions of kindergarten success rest on the child's health, social competence, ability to communicate, and ability to follow directions (Lin, Lawrence, & Gorrell, 2003; Wesley & Buysse, 2003). Other studies suggest that parents and preschool teachers place greater emphasis on academic competencies and basic knowledge, such as letters of the alphabet, than kindergarten teachers (Harradine & Clifford, 1996; Hains, Fowler, Schwartz, Kottwitz, & Rosenkoetter, 1989; West, Germino-Hausken, & Collins,

1993).

State early learning standards that define expectations for children's learning and development prior to kindergarten entry can also be viewed as a conceptualization of the state-level expectations for kindergarten students' skills. These standards documents are important to understanding the measurement of kindergarten readiness because they represent a bridge from early learning to formal schooling. Scott-Little, Kagan, and Frelow (2006) conducted a content analysis of 46 early learning standards documents developed by state-level organizations available for review in January 2005 and found that state early learning standards are more focused on language and cognitive skills than on the other domains. The authors suggest that this emphasis may derive from efforts to link early learning standards with K-12 standards, as well as from more academically oriented content being pushed down into the early years as it relates to achievement in later grades. The authors also examined the depth and breadth of topics covered within each domain. Within the physical health and motor development domain, they found that motor skills (gross, fine, oral, sensory) and functional performance/self-help skills have been the subject of far more standards items than physical fitness or overall health. Social skills with peers was the indicator category most often reflected in standards items within the social-emotional domain. Other indicators included the expression of emotions, self-concept, and comprehension of the feelings of others. Few states include standards related to the ability to develop relationships with peers and adults and the child's self-efficacy. In the approaches to learning domain, the following four indicators were approximately equally represented: (1) approach to reflection and interpretation; (2) curiosity about new tasks and challenges; (3) capacity for invention and imagination; and (4) initiative, task persistence, and attentiveness. Within the language and communication domain, 16 different indicators addressed either verbal language or early literacy skills. Within the cognition and general knowledge domain, almost 80% of the cognitive standards items were coded as either knowledge of the physical world or logico-mathematical knowledge.

Early childhood experts tell us that a multifaceted view of the child is imperative, but limited operational guidance for policy makers and practitioners is available regarding creation of measures of young children's knowledge and skills. Data from the ECLS-B and ECLS-K offer perspectives on children's abilities, but the assessment and skill inventory techniques across the domains are not readily available for replication. At present, 25 states have assessments of school readiness, and an additional four states have assessments in development; most are a single teacher checklist of students' skills (Stedron & Berger, 2010). Yet few relevant large-scale studies have been published. Information on students' skills at kindergarten entry is necessary to inform efforts to establish expectations for kindergarten readiness and children's preparedness to learn. The current study was designed to explore the empirical structure of the domains used to define students' skills at kindergarten entry based on teacher ratings using a sample of students in one state. Specifically, we examined the structure of teacher ratings using exploratory and confirmatory factor analysis in an attempt to classify students' skills. Our goal was to describe the skills of one sample of students as they started kindergarten. Our hope is that this effort will inform other local research and policy initiatives built on the notion of kindergarten readiness.

## Methods

This section includes an overview of the instrument and data collection techniques, the study participants, and statistical analyses used to examine the data.

### Instrumentation

In 2005 and 2006, the State of Connecticut passed legislation requiring the implementation of a statewide developmentally appropriate assessment that "measures a child's level of preparedness for kindergarten." In response to this legislation, the Connecticut State Department of Education developed a Kindergarten Entrance Inventory. The Inventory was designed to provide a statewide snapshot of the skills that students demonstrate, based on teachers' observations, at the beginning of the kindergarten year. The indicators for the Inventory were developed from the Connecticut Preschool Curriculum Framework and State Curriculum Standards for language arts and mathematics and are based on Connecticut's educational standards. A group of preschool and kindergarten teachers, representing urban and suburban districts, special education, and English language learners, reviewed the indicators and provided the Department of Education with their recommendations on the appropriateness of the indicators for a measure of this nature. The indicators that were selected for the Inventory are a result of the input from this committee.

Components of the Curriculum Framework and Standards were selected for the Inventory to represent the most important skills that students need to demonstrate at the beginning of kindergarten. These skills and behaviors are defined by three to five specific indicators in six domains—Language Skills, Literacy Skills, Numeracy Skills, Physical/Motor Skills, Creative/Aesthetic Skills, and Personal/Social Skills. As an example, the Language domain includes the following indicators: participates in conversations, retells information from a story read to him/her, follows simple two-step verbal directions, speaks using sentences of at least five words, communicates feelings and needs, and listens attentively to a speaker. This study is an analysis of the relationships among the indicators. The instrument was first used in fall 2007.

In the state's implementation of the instrument, each teacher is required to classify the students in his/her class(es) into three performance levels by domain; i.e., each teacher assigns each student one rating for each of the six domains. Teachers are asked to assign a rating from 1 to 3 based on the consistency with which the student demonstrates the skills and the level of instructional support required for skill demonstration. The rating scale has three levels:

> Level 1: Students at this level demonstrate emerging skills in the specified domain and require a large degree of instructional support.

> Level 2: Students at this level inconsistently demonstrate the skills in the specified domain and require some instructional support.

> Level 3: Students at this level consistently demonstrate the skills in the specified domain and require minimal instructional support.

No guidance is offered to teachers regarding how to assign a rating for a student who has variable abilities on a set of skills within a single domain.

In fall 2009, administrators in one large urban district petitioned the state to complete the Kindergarten Entrance Inventory at both the indicator and the domain levels (only domain-level data are required by the state). The administrators felt that the data provided from this use of the Inventory would have greater utility than the data from the instrument in its original form. Data from this implementation were used for the current study. In total, these teachers assigned ratings to each of their students on 32 indicators across 6 domains.

### Participants

Ninety-five kindergarten teachers in 24 different elementary schools assigned ratings to 1,670 students. The number of students assigned to each teacher ranged from 1 to 34, with a median of 18. Five teachers taught half-day programs, and eight teachers submitted data for fewer than 10 students. Demographic data for the students was provided by the district. The data showed that 49% of the students were female, 27% had limited English proficiency, 9% received special education services, and 97% were eligible for free/reduced price lunch. A majority of the students were Hispanic (60%). Of the remaining students, 32% were Black, 6% were White, 2% were Asian, and 0.1% were American Indian.

### Data Analyses

The Kindergarten Entrance Inventory was designed to measure kindergarten readiness, which is considered a latent, or unobserved, variable. In this data set, readiness was measured on a 3-point rating scale based on the consistency and the level of independence with which students demonstrated a set of observable skills and knowledge at kindergarten entry. Exploratory and confirmatory factor analytic procedures were used to bring structure to a definition of kindergarten readiness using the indicator scores. The data from the 2009 data collection were randomly split into two samples for these analyses. The first subsample was used for the exploratory analysis; this data structure from the exploratory analysis was confirmed using the second subsample. Though the results of the study have implications for the validity of the current instrument, we believe subscales scores from this measure can be used as a reporting structure for similar instruments designed to assess kindergarten students' skills.

Initially, an exploratory factor analysis using principal axis factoring (PAF) with direct oblimin rotation was used on the first sample. In PAF, factors are defined based on covariation among the indicators. Similarly, direct oblimin rotation assumes that the factors are correlated, which was evident in preliminary correlational analyses. Several criteria were used to define the appropriate number of factors within a data set, including the Kaiser-Guttman rule (Kaiser, 1991), the Scree Plot (Thompson, 2004), and a Parallel Analysis (Hayton, Allen, & Scarpello, 2004; Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Next, we conducted multiple confirmatory factor analyses (CFA) using Mplus (Muthén & Muthén, 2007) to examine the fit of the factor structure to the second data sample. Initially, the student ratings were analyzed in a nonhierarchical manner and were treated as continuous. We acknowledge that these data are more appropriately classified as ordered categorical. However, we present an analysis based on continuous treatment of the data to allow for the examination of modification indices. The modification indices offer a preliminary understanding of the manner in which teachers viewed the Inventory indicators because they reflect covariance among the indicators. Modification indices are not available in MPlus when the data are treated as ordinal.

Following the single-level CFA, we examined an ordinal treatment of the data in a multilevel context. We believe that this multilevel technique is more appropriate because it accounts for the clustering of the ratings by teacher. Multilevel CFA (MCFA) explicitly models the factor structure at both the within-teacher and the between-teacher levels. There are several reasons to anticipate that the data would exhibit non-independence. First, despite the state's training and professional development efforts, each teacher may have a unique interpretation of the instrument. One teacher may have a more rigid interpretation of "minimal instructional support" or "consistently demonstrates" than the next. Second, teachers' interpretations of the rating scale

may be based on a comparison of one child to the pool of students in a given classroom. Another teacher may interpret the scale in the context of an ideal for all kindergarten students across the state. Finally, the multilevel framework accounts for the natural correlation that may exist among students in the same classroom. In our MCFA, we evaluated model fit using several common fit indices, including the root mean square error of approximation (RMSEA), the Tucker-Lewis index (TLI), and the comparative fit index (CFI). We also examined the standardized regression weights (pattern matrix), the squared multiple correlations of the indicators, and the standardized residuals.

In a multilevel CFA, the variance-covariance matrix is decomposed into two matrices—one that captures the within-teacher variances and covariances and one that captures the between-teacher variances and covariances. The proportion of between-school variance to total variance is the intraclass correlation (ICC), which increases as a result of both between-teacher heterogeneity and within-teacher homogeneity. The ICC ranges from 0 to 1. Higher ICCs indicate that a greater proportion of item variance lies between teachers. There is some degree of homogeneity among students who are rated by a given teacher and/or some degree of heterogeneity across teachers in terms of their student ratings. In other words, knowing who a student's teacher is can help predict students' scores on the assessment. If the ICC were 0, students in the same classroom would be no more similar than students in other classrooms. If the ICC were 1, students in the same classroom would be complete replicates of one another. If teachers assigned identical ratings to all of their students, the ICC would also be 1.

The number of estimable parameters was limited by the number of teachers ($n$ = 84). For that reason, it was necessary to impose some constraints on the multilevel CFA model. First, we constrained the between and within loadings for all indicators to be equal. In addition, we constrained the error variances at the between level to be zero. Constraining the error variances at the between level to be zero implies that all of the variability in the group means can be explained by differences in the common factor means. Hox (2002) states that fixing residual variances to zero at the between level is often necessary in MCFA when sample sizes at Level 2 are small and the true between-group variance is close to zero. In contrast, allowing between-level residuals implies that some group level variance is specific to each measured variable (Kamata, Bauer, & Miyazaki, 2008). These are common constraints (Kamata et al., 2008).

## Results

The mean and standard deviation for each item are indicated in Table 1. Overall, the indicators in the Physical/Motor and Creative domains had higher means, and the indicators in the Literacy and Numeracy domains had lower means. The sample of students was split randomly, and one subsample was used for each analysis.

**Table 1**
Item Stems, Means, and Standard Deviations ($n$ = 1659)

| Indicator Stem | M | SD |
|---|---|---|
| Lang1/Participate in conversations | 2.10 | 0.77 |
| Lang2/Retell information from a story read to him/her | 1.81 | 0.74 |
| Lang3/Follow simple two-step verbal directions | 2.12 | 0.74 |
| Lang4/Speak using sentences of at least 5 words | 2.11 | 0.79 |
| Lang5/Communicate feelings and needs | 2.07 | 0.74 |
| Lang6/Listen attentively to a speaker | 2.02 | 0.75 |
| Lit1/Hold a book and turn pages from the front to the back | 2.32 | 0.73 |
| Lit2/Understand that print conveys meaning | 2.07 | 0.77 |
| Lit3/Explore books independently | 2.19 | 0.74 |
| Lit4/Recognize printed letters, especially in their name and familiar printed words | 2.00 | 0.78 |
| Lit5/Match/connect letters and sounds | 1.82 | 0.76 |
| Lit6/Identify some initial sounds | 1.86 | 0.77 |
| Lit7/Demonstrate emergent writing | 1.73 | 0.72 |
|  |  |  |
| Num1/Count to 10 | 2.36 | 0.76 |
| Num2/Demonstrate one-to-one correspondence while counting | 2.15 | 0.78 |
| Num3/Measure objects using a variety of everyday items | 1.73 | 0.70 |
| Num4/Identify simple shapes | 2.15 | 0.76 |
| Num5/Identify patterns | 1.95 | 0.74 |
| Num6/Sort and group objects by size, shape, function, or other attributes | 1.95 | 0.73 |
| Num7/Understand sequence of events | 1.74 | 0.70 |
|  |  |  |
| PerSoc1/Engage in self-selected activities | 2.35 | 0.67 |
| PerSoc2/Interact with peers to play or work cooperatively | 2.23 | 0.69 |
| PerSoc3/Use words to express own feelings or to identify conflicts | 2.11 | 0.73 |
| PerSoc4/Seek peer or adult help to resolve a conflict | 2.14 | 0.72 |
| PerSoc5/Follow classroom routines | 2.20 | 0.71 |

| | | | |
|---|---|---|---|
| Phys1/Run, jump, or balance | | 2.51 | 0.61 |
| Phys2/Kick or throw a ball, climb stairs, or dance | | 2.49 | 0.63 |
| Phys3/Write or draw using writing instruments | | 2.30 | 0.73 |
| Phys4/Perform tasks, such as completing puzzles, stringing beads, or cutting with scissors | | 2.28 | 0.73 |
| | | | |
| Creat1/Draw, paint, sculpt, or build to represent experiences | | 2.19 | 0.70 |
| Creat2/Participate in pretend play | | 2.29 | 0.63 |
| Creat3/Enjoy or participate in musical experiences | | 2.37 | 0.67 |

## Exploratory Factor Analysis (EFA)

We conducted an exploratory factor analysis with the indicator-level data, using principal axis factoring (PAF) with direct oblimin rotation. In these data, three factors were suggested based on the Kaiser-Guttman rule (eigenvalue greater than one), and two factors were indicated by the scree plot. In a parallel analysis, a parallel set of random "noise" data is created and compared to the extracted factors from these data. It is expected that no factors will be present in the random noise data and that legitimate factors in the research data should have eigenvalues greater than the means/percentile data of the eigenvalues in the random data. Based on the eigenvalue means and the percentile data, the analysis indicated the presence of three factors in the instrument. We opted to move forward with a three-factor solution because our primary goal was to examine how teachers use the set of indicators to define kindergarten readiness. After the rotation and extraction of three factors, a Kaiser-Meyer-Olkin Measure of Sampling Adequacy was obtained (KMO = .97) and was considered "marvelous" by the KMO guidelines (Pett, Lackey, & Sullivan, 2003).

In general, the new factors subsumed the original scales. Based on the item correlations, the Literacy, Language, and Numeracy domains were combined into an "academic readiness" factor. Indicators from the Physical/Motor and Creative/Aesthetic domains were combined into a second factor, which we refer to as "readiness for activities." Indicators from the Personal/Social scale hung together on a third factor, with two indicators from the Language domain, which relate to engaging with others. We refer to this factor as "social readiness." The three factors and the item loadings are indicated in Table 2.

**Table 2**
Pattern Matrix from the EFA

| | Factor | | |
|---|---|---|---|
| **Indicator Stem** | **1** | **2** | **3** |
| Lit5/Match/connect letters and sounds | **0.94** | | |
| Lit6/Identify some initial sounds | **0.91** | | |
| Num7/Understand sequence of events | **0.85** | | |
| Lit4/Recognize printed letters, especially in their name and familiar printed words | **0.85** | | |
| Num6/Sort and group objects by size, shape, function, or other attributes | **0.83** | | |
| Num5/Identify patterns | **0.8** | | |
| Lit7/Demonstrate emergent writing | **0.78** | | |
| Num2/Demonstrate one-to-one correspondence while counting | **0.77** | | |
| Num4/Identify simple shapes | **0.75** | | |
| Lit2/Understand that print conveys meaning | **0.74** | | |
| Num3/Measure objects using a variety of everyday items | **0.74** | | |
| Lang2/Retell information from a story read to him/her | **0.72** | | 0.17 |
| Num1/Count to 10 | **0.67** | 0.18 | |
| Lang4/Speak using sentences of at least 5 words | **0.56** | | 0.32 |
| Lit1/Hold a book and turn pages from the front to the back | **0.55** | 0.21 | 0.16 |
| Lang1/Participate in conversations | **0.54** | | 0.31 |
| Lang3/Follow simple two-step verbal directions | **0.54** | | 0.32 |
| Lit3/Explore books independently | **0.53** | | 0.27 |
| | | | |
| Phys2/Kick or throw a ball, climb stairs, or dance | | **0.89** | |
| Phys1/Run, jump, or balance | | **0.87** | |
| Phys4/Perform tasks, such as completing puzzles, stringing beads, or cutting with scissors | 0.2 | **0.61** | |
| Phys3/Write or draw using writing instruments | 0.2 | **0.58** | |
| Creat3/Enjoy or participate in musical experiences | | **0.51** | 0.26 |
| Creat2/Participate in pretend play | | **0.47** | 0.35 |
| Creat1/Draw, paint, sculpt, or build to represent experiences | 0.23 | **0.4** | 0.3 |
| | | | |
| PerSoc1/Engage in self-selected activities | | 0.35 | **0.49** |
| PerSoc2/Interact with peers to play or work cooperatively | | 0.21 | **0.71** |
| PerSoc4/Seek peer or adult help to resolve a conflict | | | **0.85** |

| | | |
|---|---|---|
| PerSoc3/Use words to express own feelings or to identify conflicts | | **0.83** |
| PerSoc5/Follow classroom routines | | **0.68** |
| Lang5/Communicate feelings and needs | 0.39 | **0.5** |
| Lang6/Listen attentively to a speaker | 0.42 | **0.49** |

## Single-level Confirmatory Factor Analysis (CFA)

While the primary purpose of the CFA was to confirm the three-factor data structure indicated in the EFA, the estimation results provide useful guidance on relationships among individual indicators as well as variability that can be attributed to the teacher. Maximum likelihood estimation was used to estimate the models. As stated earlier, the data sample was randomly split for the EFA and CFA. The second sample was used in this analysis.

First, the hypothesized model based on the factor structure detailed in Table 1 was tested. Language 5 and 6 loaded onto two factors during the EFA. In order to keep the subscales completely contained to only one factor, both Language 5 and 6 were specified to load only onto the Social Readiness factor. The results indicated misfit between model and data, $\chi^2$(461, $N$ = 797) = 5195.25, $p$ < .001, $\chi^2$/df = 11.27, Tucker-Lewis Index (TLI) = .805, comparative fit index (CFI) = .819, and root mean-square-error-of-approximation (RMSEA) = .114 CI (.111, .116). Given that the chi-square statistic is sensitive to large sample sizes, acceptable model fit would be indicated by TLI values above .95, CFI values above .95, and RMSEA values below .06 or a confidence interval that contains .05 (Browne & Cudeck, 1993; Hu & Bentler, 1999).

We examined the modification indices to help us better understand the structure and functioning of the instrument. It is believed that the modification indices reflect teachers' use of the instrument for several reasons. Correlated errors are produced when the residual of one indicator is associated with the residual of another indicator. In this context, correlated errors may result when a teacher assigns identical ratings on multiple indicators, perhaps based on prior knowledge or an assumption of general ability rather than on an assessment of the stated skill. Alternatively, correlated errors may result when indicators actually addressed the same material. Model misfit may also be the result of an item loading on more than one factor. Three indicators loaded on two factors. The correlated errors and cross loadings are informative from an instrument development perspective because they help identify either indicators with redundant language or redundant treatment of the indicators by the participating teachers. Table 3 indicates groupings of indicators within each domain based on the modification indices. This table also includes a column labeled "Potential Subdomain." In general, the modification indices suggested groupings of indicators that had similar content. Addressing the correlated errors and cross-loadings led to improved model fit, $\chi^2$ (440, $N$ = 797) = 1948.81, $p$ < .001, $\chi^2$/df = 4.43, Tucker-Lewis Index (TLI) = .935, comparative fit index (CFI) = .942, and root mean-square-error-of-approximation (RMSEA) = .066 CI (.063, .069).The implications and utility of this subdomain are addressed in the Discussion.

**Table 3**
Redundant Indicators-Based Modification Indices

| Original Domain | Potential Subdomain | Original Indicators |
|---|---|---|
| Language | Expressive language | Lang4/Speaks using sentences of at least 5 words<br>Lang5/Communicates feelings and needs<br><br>Lang1/Participates in conversations |
| | Receptive language | Lang6/Listens attentively to a speaker<br><br>Lang3/Follows simple two-step verbal directions |
| | Retelling stories | Lang2/Retells information from a story read to him/her |
| Literacy | Familiarity with books | Lit1/Holds a book and turn pages from the front to the back<br>Lit2/Understands that print conveys meaning<br><br>Lit3/Explores books independently |
| | Familiarity with letters | Lit5/Match/connect letters and sounds<br><br>Lit4/Recognize printed letters, especially in their name and familiar printed words<br><br>Lit6/Identify some initial sounds |
| | Emergent writing | Lit7/Demonstrate emergent writing |
| Numeracy | Counting | Num1/Count to 10<br><br>Num2/Demonstrate one-to-one correspondence while counting |

| | | |
|---|---|---|
| | Shapes/Patterns | Num4/Identify simple shapes |
| | | Num5/Identify patterns |
| | | Num6/Sort and group objects by size, shape, function, or other attributes |
| | | Num7/Understand sequence of events |
| | Measurement | Num3/Measure objects using a variety of everyday items |
| Physical/ Motor | Fine motor skills | Phys3/Write or draw using writing instruments |
| | | Phys4/Perform tasks, such as completing puzzles, stringing beads, or cutting with scissors |
| | Gross motor skills | Phys1/Run, jump, or balance |
| | | Phys2/Kick or throw a ball, climb stairs, or dance |
| Personal/ Social | Conflict resolution | PerSoc3/Use words to express own feelings or to identify conflicts |
| | | PerSoc4/Seek peer or adult help to resolve a conflict |
| | Engagement | PerSoc2/Interact with peers to play or work cooperatively |
| | | PerSoc5/Follow classroom routines |
| | Self-selected activities | PerSoc1/Engage in self-selected activities |
| Creative/ Aesthetic | Creative/Aesthetic | Creat1/Draw, paint, sculpt, or build to represent experiences |
| | | Creat1/Participate in pretend play |
| | | Creat1/Enjoy or participate in musical experiences |

## Multilevel Confirmatory Factor Analysis (MCFA)

Each factor was estimated separately because the number of estimable parameters was limited by the number of teachers. Many, but not all, of the modification indices required at the single-level analysis were required to achieve model fit in the multilevel context. In addition, we kept the residual error variances at the between level to be zero and freed the variances of several items based on the modification indices (as indicated in Table 3). Given the constraints of the model, each factor exhibited acceptable measures of model fit. The results of each model, including the correlated errors and treatment of the residual variances, are included in Table 4.

**Table 4**
Results from Individual Models for Each Factor

| | Three Models | | |
|---|---|---|---|
| | **Academic Readiness** | **Social Readiness** | **Readiness for Activities** |
| Indicators | Lang1 – Lang4 Lit1 – Lit7 Num1 – Num7 | Lang5 – Lang 6 PerSoc1 – PerSoc5 | Phys1 – Phys4 Creat1 – Creat4 |
| $\chi^2$ (df) | 304.145 (31) | 39.216 (7) | 46.129 (5) |
| RMSEA | 0.105 | 0.076 | 0.102 |
| CFI | 0.973 | 0.886 | 0.892 |
| TLI | 0.994 | 0.951 | 0.934 |
| **Correlated errors** | | | |
| Within cluster | LIT6 with LIT5 LIT3 with LIT1 LANG4 with LANG1 NUM2 with NUM1 NUM7 with NUM3 | PerSoc4 with PerSoc3 | Phys1 with Phys2 Creat3 with Creat2 Phys3 with Phys4 |
| Between clusters | LIT6 with LIT5 LIT3 with LIT1 LANG4 with LANG1 NUM2 with NUM1 NUM7 with NUM3 | PerSoc4 with PerSoc3 | n/a |
| **Residual variances** | | | |
| Between clusters | Lang1 Lit1 – Lit3, Lit5, Lit7 Num1 – Num3 Num5 – Num7 | PerSoc1 PerSoc 3 | Creat1 – Creat2 Creat4 |

We also calculated the intraclass correlations (ICCs) for each indicator. As mentioned earlier, higher ICCs indicate that a greater proportion of item variance lies between teachers. There is some degree of homogeneity among students who are rated by a given teacher and/or some degree of heterogeneity across teachers in terms of their student ratings. For example, if some teachers tended to give higher ratings than other teachers, or if there were differences in the interpretations of the meanings of some of the items among teachers, those conditions could result in higher ICCs. Table 5 lists indicators by ICC. Indicators with lower ICCs were interpreted more consistently (i.e., with less variability) across the sample of teachers. These items were less teacher dependent.

**Table 5**
ICCs for the Items on the Kindergarten Entrance Inventory

| Indicator Stem | Teacher N | ICC |
|---|---|---|
| Lit4/Recognize printed letters, especially in their name and familiar printed words | 84 | 0.135 |
| Lang5/Communicate feelings and needs | 84 | 0.162 |
| Per5/Follow classroom routines | 84 | 0.167 |
| Lit5/Match/connect letters and sounds | 84 | 0.168 |
| Lang6/Listen attentively to a speaker | 84 | 0.173 |
| Lit6/Identify some initial sounds | 84 | 0.178 |
| Lang2/Retell information from a story read to him/her | 84 | 0.179 |
| Num4/Identify simple shapes | 84 | 0.192 |
| Lang1/Participate in conversations | 84 | 0.198 |
| Per2/Interact with peers to play or work cooperatively | 84 | 0.198 |
| Lang4/Speak using sentences of at least 5 words | 84 | 0.210 |
| Per4/Seek peer or adult help to resolve a conflict | 84 | 0.218 |
| Per3/Use words to express own feelings or to identify conflicts | 84 | 0.241 |
| Phys4/Perform tasks, such as completing puzzles, stringing beads, or cutting with scissors | 84 | 0.248 |
| Lang3/Follow simple two-step verbal directions | 84 | 0.249 |
| Phys3/Write or draw using writing instruments | 84 | 0.255 |
| Num2/Demonstrate one-to-one correspondence while counting | 84 | 0.268 |
| Creat3/Enjoy or participate in musical experiences | 84 | 0.281 |
| Lit2/Understand that print conveys meaning | 84 | 0.287 |
| Num7/Understand sequence of events | 84 | 0.290 |
| Num1/Count to 10 | 84 | 0.291 |
| Lit7/Demonstrate emergent writing | 84 | 0.294 |
| Creat1/Draw, paint, sculpt, or build to represent experiences | 84 | 0.300 |
| Per1/Engage in self-selected activities | 84 | 0.300 |
| Num5/Identify patterns | 84 | 0.303 |
| Phys1/Run, jump, or balance | 84 | 0.308 |
| Creat2/Participate in pretend play | 84 | 0.321 |
| Lit3/Explore books independently | 84 | 0.332 |
| Phys2/Kick or throw a ball, climb stairs, or dance | 84 | 0.332 |
| Num6/Sort and group objects by size, shape, function, or other attributes | 84 | 0.346 |
| Lit1/Hold a book and turn pages from the front to the back | 84 | 0.362 |
| Num3/Measure objects using a variety of everyday items | 84 | 0.372 |

## Discussion

The Kindergarten Entrance Inventory was designed to provide a snapshot of students' skills at the start of the kindergarten year. The analyses presented here provide insight into the manner in which teachers make judgments about their students' readiness for kindergarten at the start of the year. The data suggest that when evaluating children's skills at kindergarten entry, teachers use more global evaluation schema of students' skills than is presented in the six-domain structure of the original instrument. Specifically, teacher judgments were centered around three factors in the EFA: students' academic readiness, their social readiness, and their readiness for nonacademic activities. This finding may be a result of either teachers' understanding of their students' skills at the start of the year or the structure of the instrument. Perhaps the same instrument used later in the year, when teachers have a more complex understanding of their students' abilities, would yield a different factor structure. Alternatively, the structure may be an artifact of the rating scale. It is possible that the 3-point ordinal scale encourages gross judgments of students' skills.

From the CFA, it was evident that teachers assign similar ratings on indicators with similar content. It is clear that the statewide implementation of the Kindergarten Entrance Inventory requires teachers to assign a single

rating at the domain level to a divergent set of skills. One salient example of this phenomenon is the Physical/Motor domain, which includes two indicators that relate to fine motor skills and two indicators that relate to gross motor skills. The CFA results may also indicate that the original meaning of some of the original indicators may be lost when presented in this format. The modification indices suggested correlated errors between these two language indicators—"speaks using sentences of at least 5 words" and "communicates feelings and needs." A cursory glance at these two prompts might suggest similar content, and the correlated errors suggest that teachers assigned similar ratings to each prompt. The curriculum framework used to write this indicator elaborates that a child may be able to communicate his or her feelings and needs using hand gestures or even sounds, without using words. This interpretation is not clear from the term "communicate." Other domains cover a hierarchy of skills. Perhaps an indicator that states "matches/connects letters and sounds" is not necessary alongside an indicator that states "identifies some initial sounds." A student who cannot identify initial sounds will not be able to associate sounds and letters.

The ICCs provide insight into the teachers' understanding of the individual indicators. A low ICC has two interpretations in this context. First, it could mean that the students had the most similar ratings on these indicators. Alternatively, it may mean that such indicators were interpreted most consistently across teachers. Two similar indicators with low ICCs are "match/connect letters and sounds" (ICC = .168, $M$ = 1.82, $SD$ = .76) and "identify some initial sounds" (ICC = .178, $M$ = 1.86, $SD$ = .77). In this case, the relatively lower mean and higher standard deviation might lead us to believe that the low ICC was the result of consistent interpretation and student variability. The specificity of the language in these indicators is also of interest. The indicator with the highest ICC, "measure objects using a variety of everyday items" (ICC = .372), had a lower mean (1.73) and standard deviation (.70). In this context, the ICC may reflect inconsistent interpretation across teachers or that all teachers tended to rate their students similarly. For this indicator, the language is quite vague. "Objects" and "everyday items" are not defined. In addition, there is no clarification offered with regard to the frequency or accuracy with which students are asked to perform these activities.

Finally, we can comment on the individual indicators with an understanding that these are data from one urban district at one point in time. Within the Literacy domain, teachers gave the highest ratings to the indicator that addressed students' familiarity with books and the lowest ratings to their emergent writing skills. The mean ratings on all indicators were very close to the midpoint of the scale in the Language domain, with the highest ratings on the indicators relating to participation in conversations and speaking in sentences of at least five words. For Numeracy, teacher ratings were highest for the "count to 10" indicator and lowest for the indicator relating to sequence of events. All of the mean ratings for the Personal/Social, Physical/Motor, and Creative domains were above the midpoint of the scale. These high ratings may have also contributed to the division of the academic and nonacademic indicators in the factor analysis. In the Personal/Social domain, "engage in self-selected activities" had the highest mean rating. The indicator relating to running, jumping, and balancing had the highest mean rating and lowest standard deviation of any on the instrument. Within the Creative domain, the indicator related to students' enjoyment of or participation in musical experiences had the highest rating. These results represent one school district and cannot be extrapolated beyond this population.

## Implications

Although our work began as a validation of the structure of one state's instrument, the results can be used to develop a more detailed measure of kindergarten readiness. The domain and subdomain structure is an outline for instrument developers. Our findings provide the foundation for a categorization of skills to be measured at kindergarten entry. Based on the indicators used in the Inventory, evaluations of students' educational development at the start of the kindergarten year should include a rating or measure of the following constructs:

- expressive language
- receptive language
- responses to stories
- familiarity with books
- familiarity with letters
- emergent writing
- counting
- shapes and patterns
- measurement
- fine motor skills
- gross motor skills
- conflict resolution
- social engagement

- engagement with self-selected activities
- creative skills

A measure based on these constructs would allow teachers to furnish a more detailed picture of individual students' development than ratings at the domain level (e.g., language, literacy, numeracy, etc.). Still, we caution that the indicators of Connecticut's Inventory provide some initial descriptors for such an instrument. Further work with teachers and early childhood researchers would be necessary to bring more description and definition to each of these constructs.

This study also offers structural guidance for researchers, evaluators, and administrators designing teacher rating scales for young children. First, specific language is necessary to achieve consistency in the utilization of the instrument across raters. If "communication" is intended to include nonverbal gestures, it should be noted. In this study, more specific indicators produced more consistent ratings. Moreover, in a teacher-driven rating scale, redundant items should be eliminated to ease the burden of data collection. Second, our results highlight issues with the number of points on the rating scale. In this study, teachers were asked to use a coarse 3-point rating scale to evaluate their students on very specific indicators. In addition, the rating scale was designed to represent the extent to which students exhibited the specified skills both independently and consistently over time. Reliable administration of the assessment requires performance descriptors that measure only one construct. With that change, an expanded, 4-point rating scale would produce more variable ratings, which would allow for more complex analyses. One example of such a scale might include one 4-point scale of consistency (not at all, some of the time, most of the time, all of the time) and one for independence.

## Limitations

The current study had several limitations. The sample of 1,600 students across 83 schools limited both the analytical techniques that could be used and the power of the current analyses. Some of the redundancies and inconsistencies evident in the analyses may result from inappropriate use of the indicator-level information, i.e., the indicators were included to describe the domains and not to guide or define individual students' skills. Finally, these data represent one group of urban students in a diverse state. In this study, 97% of the students were eligible for free or reduced price lunch, 60% were Hispanic, and 32% were Black. In 2006 data from the state, 27% of students were eligible for free or reduced price lunch, 19% were Hispanic, and 14% were Black (Connecticut State Department of Education, 2007). This difference is notable and limits the generalizability of our results.

## Acknowledgments

## References

Browne, Michael W., & Cudeck, Robert. (1993). Alternative ways of assessing model fit. In Kenneth A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Connecticut State Department of Education. (2007). *Data bulletin: Kindergarten, 2006-2007.* Hartford: Connecticut State Department of Education.

Denton Flanagan, Kristin, & McPhee, Cameron. (2009, October). *The children born in 2001 at kindergarten entry: First findings from the kindergarten data collections of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2010-005). Washington, DC: National Center for Education Statistics.

Fabrigar, Leandre R.; Wegener, Duane T.; MacCallum, Robert C.; & Strahan, Erin J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Hains, Ann Higgins; Fowler, Susan A.; Schwartz, Ilene S.; Kottwitz, Esther; & Rosenkoetter, Sharon. (1989). A comparison of preschool and kindergarten teacher expectations for school readiness. *Early Childhood Research Quarterly*, *4*(1), 75-88.

Harradine, Christine C., & Clifford, Richard M. (1996, April). *When are children ready for kindergarten? Views of families, kindergarten teachers, and child care providers.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Hayton, James C.; Allen, David G.; & Scarpello, Vida. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205.

Heaviside, Sheila, & Farris, Elizabeth. (1993, September). *Public school kindergarten teachers' views on children's readiness for school* (NCES 93-410). Washington, DC: National Center for Education Statistics.

Hox, Joop J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hu, Li-tze, & Bentler, Peter M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Kagan, Sharon Lynn; Moore, Evelyn; & Bredekamp, Sue (Eds.). (1995, June). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Washington, DC: National Education Goals Panel.

Kaiser, Henry F. (1991). Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychological Reports, 68*(3), 855-858.

Kamata, Akihito; Bauer, Daniel J.; & Miyazaki, Yasuo. (2008). Multilevel measurement model. In Ann A. O'Connell & D. Betsy McCoach (Eds.), *Multilevel analysis of educational data* (pp. 345-388). Charlotte, NC: Information Age.

Lin, Huey-Ling; Lawrence, Frank R.; & Gorrell, Jeffrey. (2003). Kindergarten teachers' views of children's readiness for school. *Early Childhood Research Quarterly*, *18*(2), 225-237.

Love, John M. (2001, December). *Instrumentation for state readiness assessment: Issues in measuring children's early development and learning*. Princeton, NJ: Mathematica Policy Research.

Meisels, Samuel J. (1999). Assessing readiness. In Robert C. Pianta & Martha J. Cox (Eds.), *The transition to kindergarten* (pp. 39-66). Baltimore, MD: Paul H. Brookes.

Muthén, Linda K., & Muthén, Bengt O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Authors.

National Association for the Education of Young Children (NAEYC) & National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2009). *Where we stand on curriculum, assessment, and program evaluation*. Retrieved August 10, 2011, from http://www.naeyc.org/files/naeyc/file/positions/StandCurrAss.pdf

Pett, Marjorie A.; Lackey, Nancy R.; & Sullivan, John J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks, CA: Sage.

Rathbun, Amy, & West, Jerry. (2004). *From kindergarten through third grade: Children's beginning school experiences.* (NCES 2004-007). Washington, DC: National Center for Education Statistics.

Scott-Little, Catherine; Kagan, Sharon Lynn; & Clifford, Richard M. (Eds.). (2003). *Assessing the state of state assessments: Perspectives on assessing young children.* Greensboro: University of North Carolina, SERVE.

Scott-Little, Catherine; Kagan, Sharon Lynn; & Frelow, Victoria Stebbins. (2006). Conceptualization of readiness and the content of early learning standards: The intersection of policy and research? *Early Childhood Research Quarterly, 21*(2), 153-173.

Stedron, Jennifer, & Berger, Alexander. (2010, August). *NCSL technical report: State approaches to school readiness assessment.* Retrieved August 10, 2011, from http://www.ncsl.org/documents/Educ/KindergartenAssessment.pdf

Thompson, Bruce. (2004). *Exploratory and confirmatory factor analysis.* Washington, DC: American Psychological Association.

Wesley, Patricia W., & Buysse, Virginia. (2003). Making meaning of school readiness in schools and communities. *Early Childhood Research Quarterly*, *18*(3), 351-375.

West, Jerry; Denton, Kristin; & Germino-Hausken, Elvira. (2000). *America's kindergartners.* (NCES 2000-070). Washington, DC: National Center for Education Statistics.

West, Jerry; Germino-Hausken, Elvie; & Collins, Mary. (1993, September). *Readiness for kindergarten: Parent and teacher beliefs* (NCES 93-257). Washington, DC: National Center for Education Statistics.

## Author Information

Dr. Jessica Goldstein is an assistant professor in residence in the Measurement, Evaluation, and Assessment Program at the University of Connecticut. Dr. Goldstein's research interests include the validity of large-scale assessment systems for special populations and the use of alternative measures of student achievement for school accountability.

Jessica Goldstein
Assistant Professor in Residence
Measurement, Evaluation, and Assessment Program
Department of Educational Psychology
University of Connecticut
249 Glenbrook Road, Unit 2064
Storrs, CT 06269
Email: Jessica.Goldstein@uconn.edu

Dr. D. Betsy McCoach is an associate professor in the Measurement, Evaluation, and Assessment Program at the University of Connecticut. She has extensive experience in hierarchical linear modeling, instrument design, factor analysis, and structural equation modeling. She has coauthored over 50 articles and book chapters.

D. Betsy McCoach
University of Connecticut
249 Glenbrook Road, Unit 2064
Storrs, CT 06269-2064
Telephone: 860486-0183
Fax: 860-486-0180
Email: betsy.mccoach@uconn.edu

University of Illinois at Urbana-Champaign
College of Education
Early Childhood and Parenting Collaborative