Scientific evidence for the validity of the New Mexico Kindergarten Observation Tool

Katie Dahlke
Rui Yang
Carmen Martínez
Suzette Chavez
Alejandra Martin
Laura Hawkinson
American Institutes for Research

Joseph Shields
Marshall Garland
Jill Carle
Gibson Consulting Group

Key findings

This study examined the construct validity of the 2015 New Mexico Kindergarten Observation Tool (KOT), an observational measure of students' knowledge and skills completed by the kindergarten teacher at the beginning of the year. Key findings include:

- Construct validity analyses supported an overall general school readiness score and two domain scores: a cognitive school readiness domain score and a noncognitive school readiness domain score. The analyses did not support the six domain scores identified by the developer.
- KOT domain scores were moderately correlated with scores from an established measure of early literacy skills, and the correlation patterns support the conclusion that the KOT domains measure distinct dimensions of school readiness.
- Rating categories were distinct (that is, no category was redundant) and ordered appropriately (that is, teachers used higher rating categories for students with higher overall ability).
- Substantial classroom-level variation was found for KOT domain scores and item ratings. Such variation is not uncommon among observational measures or indirect assessments, but it raises questions about the extent to which scores measure students' true abilities.

Summary

The New Mexico Public Education Department developed the Kindergarten Observation Tool (KOT) as a multidimensional observational measure of students' knowledge and skills at kindergarten entry. It was intended to assess knowledge and skills in six domains:
- Physical development, health, and well-being.
- Literacy.
- Numeracy.
- Scientific conceptual understanding.
- Self, family, and community.
- Approaches to learning.

The primary purpose of the KOT is to inform instruction, so that kindergarten teachers can use the information about their students' knowledge and skills from the KOT to inform their curricular and pedagogical decisions. Stakeholders also are interested in using data from the KOT for other purposes, such as assessing student readiness for school statewide and identifying disparities in students' readiness for school across the state.

This study examined the construct validity of the KOT to determine whether data from a field test supported using the KOT to measure six school readiness domains and, if not, what domain structure the data best supported. The study was conducted in response to the New Mexico Prekindergarten Research Alliance's and New Mexico Public Education Department's interest in evidence for the KOT's validity and reliability. The study team conducted exploratory and confirmatory factor analyses to identify the latent constructs that the KOT measured. The analyses identified the sets of items that were most related to one another and that could be used to develop domain scores. The study team also assessed reliability to confirm that items in the identified domains consistently measured the same construct. In addition, the study team applied other psychometric analysis methods to examine item functioning and differential item functioning across student subgroups. These analyses provided evidence of whether the KOT rating categories for each item were ordered correctly and whether there were any potential biases in how teachers rated student subgroups. Finally, the study team examined the proportion of the variation in the KOT's domain scores and item ratings at the classroom level to explore the extent to which domain scores and item ratings provide information about individual student abilities.

The study identified valid and reliable approaches for scoring KOT item ratings, although not based on the developer's intended six-domain structure. Additional development and validation work is still needed to assess benchmarks for school readiness, determine whether particular items are biased for student subgroups, and examine the sources of classroom-level variations in scores.

Key findings include:
- Construct validity analyses supported an overall general school readiness score and two domain scores: a cognitive school readiness domain score and a non-cognitive school readiness domain score. The analyses did not support the six domain scores identified by the developer.
- KOT domain scores in the one- and two-factor structures showed strong internal consistency.

- KOT domain scores were moderately correlated with scores from an established measure of early literacy skills. The KOT was consistent with the established measure in classifying students as ready for school (or not) for two-thirds of students. The correlation patterns support the conclusion that the KOT domains measure distinct dimensions of school readiness.
- Four items demonstrated potential bias in how they were used for particular student subgroups. Additional work is required to determine whether those items are truly unfair to certain student subgroups.
- Rating categories were distinct (that is, no category was redundant) and ordered appropriately (that is, teachers used higher rating categories for students with higher overall ability).
- Substantial classroom-level variation was found for KOT domain scores and item ratings. Such variation is not uncommon among observational measures or indirect assessments, but it raises questions about the extent to which scores measure students' true abilities.

Contents

Tables

Why this study?

Children's knowledge, skills, and experience when they enter school vary widely (Aikens, Kopack Klein, Tarullo, & West, 2013; Duncan & Magnuson, 2005; Janus & Duku, 2007; Larson, Russ, Nelson, Olson, & Halfon, 2015; Reardon & Portilla, 2016; Shonkoff & Phillips, 2000). Measuring individual children's knowledge and skills in various developmental domains when they enter kindergarten can help teachers individualize instruction in elementary school, can inform efforts to reduce school readiness gaps, and can help evaluate outcomes of children's early learning experiences before kindergarten within the context of a rigorous evaluation design. Many states have developed (or are using existing) kindergarten entry assessments to measure and document children's knowledge and skills systematically when they enter school.

New Mexico has developed a multidimensional observational measure of students' knowledge and skills at kindergarten entry—the New Mexico Kindergarten Observation Tool (KOT). Its primary purpose is to inform instruction, so that kindergarten teachers can use the information about their students' knowledge and skills from the KOT to inform their curricular and pedagogical decisions. Stakeholders also are interested in using data from the KOT for other purposes, such as assessing students' readiness for school statewide and identifying disparities in students' readiness for school across the state.

This study responds to interest in evidence for the Kindergarten Observation Tool's (KOT) validity and reliability as well as in guidance for refining the KOT

The current study responds to the New Mexico Prekindergarten Research Alliance's[1] and the New Mexico Public Education Department's interest in evidence for the KOT's validity and reliability as well as in guidance for further refining the KOT. The findings can inform decisions about appropriate applications of data from the KOT, including how to produce summary scores.

The New Mexico Public Education Department contracted with WestEd to develop the KOT from the existing Prekindergarten Observation Assessment Tool and in alignment with the New Mexico Content Standards and Common Core State Standards. After an early version was piloted in fall 2014 with a small sample of students, the KOT was revised to include 24 indicators across six intended domains:
- Physical development, health, and well-being.
- Literacy.
- Numeracy.
- Scientific conceptual understanding.
- Self, family, and community.
- Approaches to learning.

The 2015 KOT included scoring rubrics for 26 items designed to measure the 24 indicators (see appendix A for the rubric). In fall 2015 WestEd and Central Regional Educational Cooperative 5 conducted a field test of the KOT in 45 of the 89 school districts in New Mexico. This study uses data from that field test.

Training teachers to use the Kindergarten Observation Tool

For the fall 2015 field test, teachers received training on administering and scoring the KOT through a train-the-trainer model. First, WestEd and Central Regional Educational Cooperative 5 provided a two-day training in spring 2015 to professional development

providers, including school district staff, representing all participating school districts and all regional educational cooperatives with participating schools. The training covered the content and administration of the KOT as well as how to train teachers to use it.

In summer 2015 the local professional development providers trained kindergarten teachers who had been recruited in participating districts to administer the KOT. Local professional development providers could use a one-day format designed for returning teachers or a two-day format designed for new teachers. The training focused on the content of each indicator, how to observe students, how to use the scoring rubric, how to submit scores to the New Mexico Public Education Department data system, and how to use data from the KOT to inform instruction. Following the training, teachers were asked to practice scoring by using videos and to self-report whether they understood how to administer the KOT and score students or needed additional training.

Administration procedures for the Kindergarten Observation Tool

To administer the KOT, teachers observed their students in a variety of classroom activities during the first 30 instructional days of kindergarten. After collecting sufficient evidence through observation, teachers scored each item by using the scoring rubric at some point during the 30-day observation period or during a scoring period after the observation period (but drawing only from their observations during the 30-day observation period). Students with many absences or late registrations and some students with Individualized Education Program accommodations (about 5 percent of students with an Individualized Education Program) and medical issues were exempt from KOT ratings, and teachers were asked to mark "exempt" under these circumstances. If teachers felt that they had not had an opportunity to observe a particular area of knowledge or skill for a student during the observation window, they were asked to leave the item blank.

To administer the KOT, teachers observed their students in a variety of classroom activities during the first 30 instructional days of kindergarten

Teachers entered the resulting scores for each student into the New Mexico Public Education Department's secure database, the KOT Application. After all KOT item ratings were entered, teachers could generate a student summary report that provided individual student ratings for each item and a classroom summary report that provided classroom-level frequencies at each rating for each item to inform instruction. Teachers could also generate a family–teacher report to share and discuss with parents.

What the study examined

Six research questions guided this study:

1. What domains does the KOT measure?
   a. How do the latent constructs measured align with the kindergarten readiness domains that the KOT was designed to measure?
   b. Are there differences in latent constructs measured by student gender, English learner status, eligibility for the federal school lunch program, special education status, or race/ethnicity?
   c. What is the most reliable way to score the latent constructs measured?

2. To what extent do items in each domain measure the same underlying construct?
   a. What is the internal consistency of each latent construct?
   b. Would the internal consistency and scoring of the latent constructs be improved by removing any items?

3. To what extent does the KOT measure the same underlying constructs as other accepted measures of kindergarten readiness?

4. Do any of the items exhibit potential bias for student subgroups?
   a. Do any of the items exhibit potential bias by student gender?
   b. Do any of the items exhibit potential bias by student English learner status?
   c. Do any of the items exhibit potential bias by student eligibility for the federal school lunch program?
   d. Do any of the items exhibit potential bias by special education status?
   e. Do any of the items exhibit potential bias by student race/ethnicity?

5. Do teachers use the rating categories for each item as intended?

6. To what extent do ratings provide information about individual student abilities?

The study team used data from the fall 2015 field test of the KOT, data from the fall 2015 administration of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next, and demographic data from the New Mexico Public Education Department's Student Teacher Accountability Reporting System. Analytic methods included exploratory factor analyses, multiple-group confirmatory factor analyses, differential item functioning analyses, Rasch modeling, and variance component analyses. See box 1 for a summary of data and methods and appendix B for more details.

The study team used data from the fall 2015 field test of the KOT, data from the fall 2015 administration of the Dynamic Indicators of Basic Early Literacy Skills Next, and demographic data as well as exploratory factor analyses, multiple-group confirmatory factor analyses, differential item functioning analyses, Rasch modeling, and variance component analyses

Box 1. Data and methods

Data

In fall 2015 WestEd and Central Regional Educational Cooperative 5 conducted a field test of the Kindergarten Observation Tool (KOT) in 45 of the 89 school districts in New Mexico. The analytic sample included data from 5,259 students (with 132,971 observed items) across 340 teachers and 112 schools who participated in the field test. The subsample for research question 3 included 3,257 students in the field test who also had a score on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next, an assessment used to measure the acquisition of early literacy skills, administered by their classroom teacher at the beginning of their kindergarten year in fall 2015. The study team also obtained student demographic data from the New Mexico Public Education Department through the Student Teacher Accountability Reporting System (see table).

The analytic sample had a lower percentage of students who were eligible for the federal school lunch program (64 percent) than did all districts that participated in the field test (74 percent) and the statewide population (75 percent; see table). For most demographic measures the subsample for research question 3 mirrored the full analytic sample, but the subsample for research question 3 had a higher proportion of American Indian/Alaska Native students (24 percent) than the full analytic sample (17 percent) and a smaller proportion of Hispanic students (47 percent compared with 52 percent).

(continued)

Box 1. Data and methods *(continued)*

Table. Demographic characteristics of New Mexico kindergarten student population, kindergarten students in school districts that participated in the 2015 Kindergarten Observation Tool field test, the analytic sample, and the subsample for research question 3, fall 2015

Characteristic	Percent of New Mexico kindergarten student population (*N* = 27,567)	Percent of kindergarten students in participating school districts (*N* = 5,305)	Percent of students in analytic sample (*N* = 5,259)	Percent of students in subsample for research question 3[a] (*N* = 3,257)
Female	48	48	49	49
English learner student	19	18	15	16
Living in poverty	40	36	26	27
Eligible for the federal school lunch program	75	74	64	63
In special education	12	10	11	9
Race/ethnicity[b]				
Hispanic	60	56	52	47
White	24	24	25	23
American Indian/Alaska Native	11	14	17	24
Black	2	2	1	1
Asian	1	1	1	1
Native Hawaiian/Pacific Islander	< 1	< 1	< 1	< 1
Two or more races/ethnicities	2	3	4	3

a. Students in the analytic sample who had Dynamic Indicators of Basic Early Literacy Skills Next scores.

b. Racial/ethnic groups are mutually exclusive.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and demographic data from the New Mexico Public Education Department.

Methods

To identify the latent constructs that the KOT measured, the study team conducted exploratory and confirmatory factor analyses with the 2015 field test data (research question 1). First, an exploratory factor analysis uncovered the number of distinct domains or groups of items that best describe the data. Without specifying the domains beforehand, exploratory factor analysis uncovers domains on the basis of how the items relate to one another. Confirmatory factor analysis was then used to test the fit of the developer's intended domain structure to the data. The study team selected the final factor structure on the basis of five criteria: the hyperplane count (Yates, 1988), the number of nonsalient loadings below a threshold of 0.4 (Stevens, 2009), the number of double loaders (wherein a smaller number suggests better fit), the closeness to a simple structure (wherein the simpler the structure, the better; Fabrigar, Wegener, MacCallum, & Strahan, 1999), and the meaningfulness of each factor (Gadermann, Guhn, & Zumbo, 2012).

Multiple-group confirmatory factor analyses and differential item functioning analyses also were conducted to determine whether the domain structure was invariant across student subgroups and to identify items that functioned differently for subgroups (research questions 1 and 4). To examine the extent to which KOT items related to one another, the study team examined the internal consistency of the three domains identified by means of factor analyses

(continued)

Box 1. Data and methods *(continued)*

(research question 2). The study team used psychometric analysis methods to generate scores for each validated domain that were supported by factor analyses and to determine reliable approaches for scoring.

To examine the concurrent relationship between KOT scores and another accepted measure of kindergarten readiness, the study team estimated correlations between students' DIBELS Next score and their ratings on each of the 26 KOT items (research question 3). In addition, the study team estimated the relationship between students' scores on each of the three domains of school readiness identified by the construct validity analyses (that is, the average of a student's ratings across all items in the domain) and their DIBELS Next composite scaled score. Finally, the team compared benchmarks of kindergarten readiness between the KOT and DIBELS Next to determine whether New Mexico's current school readiness benchmarks on the KOT reflected school readiness as measured by the DIBELS Next benchmark.

The study team also used Rasch modeling to examine how the rating categories for each item were ordered on the estimated parameters. This analysis determines whether behaviors described for each rating category represented a higher level skill or were more difficult to achieve than behaviors described for the ratings below it (research question 5). Variance component analyses were conducted to determine the proportion of variation in KOT scores at the classroom level (research question 6). Four-level unconditional models (student, classroom, school, and district) were fit for each item and for the three domains to examine the classroom-level, within-school, and within-district variation. Variation in scores above the student level reduces the certainty of the extent to which scores measure students' true abilities.

Additional methodological details are provided in appendix B.

The study found insufficient support for the developer's intended six-domain structure for the KOT, but it did find support for a one-factor general school readiness structure and a two-factor structure with a cognitive school readiness domain and a noncognitive school readiness domain

What the study found

This section discusses the seven main findings of the study.

Construct validity analyses supported a one-factor structure for the Kindergarten Observation Tool to generate an overall general school readiness score and a two-factor structure for it to generate a cognitive school readiness domain score and a noncognitive school readiness domain score

The study found insufficient support for the developer's intended six-domain structure for the KOT. On the basis of fit statistics, the developer's intended structure did not fit the data as well as other similarly complex models did (for example, a six-domain structure identified from exploratory factor analysis), indicating that the developer's intended structure was not the optimal factor solution. Moreover, three domains in the developer's intended structure included only two items; typically, more than two items are needed for valid and reliable measurement of a latent construct.

The study found support for a one-factor general school readiness structure and a two-factor structure with a cognitive school readiness domain and a noncognitive school readiness domain (see figure C1 in appendix C for a scree plot and table C1 in appendix C for factor loadings). In the one-factor structure the general school readiness domain score included 25 of the 26 items—coordination and strength (item 1.1) was excluded. In the two-factor structure the cognitive school readiness domain included 16 items (from the literacy, numeracy, and scientific conceptual understanding domains of the developer's

intended structure), and the noncognitive school readiness domain included 8 items (from the literacy; self, family, and community; and approaches to learning domains of the developer's intended structure; table 1). These factor structures did not vary by gender, English learner status, special education status, eligibility for the federal school lunch program, or race/ethnicity (see table C2 in appendix C).

The two items from the physical development, health, and well-being domain of the developer's intended structure did not contribute clearly to either domain in the two-factor structure and were thus not included in either the cognitive or noncognitive school readiness domain.

Both the one- and two-factor structures showed strong internal consistency for each domain of the Kindergarten Observation Tool

In the one-factor structure the general school readiness domain had an estimated reliability of .97 (as measured with the ordinal version of Cronbach's alpha). In the two-factor structure the cognitive school readiness domain had an estimated reliability of .96, and the noncognitive school readiness domain had an estimated reliability of .94. Pearson reliabilities also were high: .94 for the general school readiness domain in the one-factor structure and .90 for the cognitive

In the two-factor structure the cognitive school readiness domain included 16 items, and the noncognitive school readiness domain included 8 items

Table 1. New Mexico Kindergarten Observation Tool items and factor loadings by validated two-factor structure domains, fall 2015

Cognitive school readiness domain		Noncognitive school readiness domain	
Item	Factor loading	Item	Factor loading
Number words (Numeracy 9.3a)	.88	Self-control (Self, family, and community 18.1)	.86
Alphabet knowledge and word/letter recognition (Literacy 7.4)	.84	Cares for possessions (Self, family, and community 19.1)	.83
Letter-sounds, beginning sound (Literacy 5.3b)	.82	Guidance and support (Self, family, and community 21.2)	.80
One-to-one correspondence, number relationships (Numeracy 9.1)	.81	Independence (Approaches to learning 24.2)	.78
Numerals (Numeracy 9.3b)	.78	Social problem solving (Self, family, and community 20.2)	.77
Vocabulary (Literacy 5.4)	.73	Focus (Approaches to learning 27.1)	.76
Story comprehension (Literacy 7.2)	.72	Plays and interacts (Self, family, and community 20.1)	.74
Measurement (Numeracy 11.3)	.71	Follows directions (Literacy 5.2)	.52
Book conventions (Literacy 7.3)	.69		
Rhyme (Literacy 5.3a)	.68		
Book enjoyment (Literacy 7.1)	.67		
Conversational ability (Literacy 6.1)	.66		
Writing (Literacy 8.3)	.66		
Sorting (Numeracy 12.1)	.64		
Investigations (Scientific conceptual understanding 14.1)	.55		
Earth science (Scientific conceptual understanding 16.1)	.55		

Note: In each domain items are sorted in descending order of factor loadings. Text in parentheses is the domain in the developer's intended structure to which the item belonged and the item number. $N = 5,259$ students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

school readiness domain and .87 for the noncognitive school readiness domain in the two-factor structure. In addition, reliability of the general school readiness domain improved when coordination and strength (item 1.1) was excluded (see table C3 in appendix C).

Kindergarten Observation Tool domain scores were moderately correlated with scores from an established measure of early literacy skills and concurred with that measure's scores in recognizing students as demonstrating foundational knowledge and skills at kindergarten entry for approximately two-thirds of students

Scores for all three KOT domains were significantly correlated with DIBELS Next beginning of year scores and represented moderate convergence (that is, correlations were between .40 and .69; Di Iorio, 2005). The Spearman rank-order correlation coefficient was .64 between the general school readiness domain score and the DIBELS Next score, .67 between the cognitive school readiness domain score and the DIBELS Next score, and .46 between the noncognitive school readiness domain score and the DIBELS Next score (figure 1).

The correlation between the cognitive domain score and the DIBELS Next score was significantly different from the correlation between the noncognitive domain score and the DIBELS Next score (on the basis of a Wald test; see appendix B). Given that the DIBELS Next measures early literacy skills, this pattern aligns with expectations because the general and cognitive school readiness domains include literacy items, but the noncognitive school readiness domain does not. The pattern also supports the notion that the cognitive and noncognitive school readiness domains measure different (although related) underlying constructs.

The Spearman rank-order correlation coefficient was .64 between the general school readiness domain score and the DIBELS Next score, .67 between the cognitive school readiness domain score and the DIBELS Next score, and .46 between the noncognitive school readiness domain score and the DIBELS Next score

Figure 1. New Mexico Kindergarten Observation Tool domain scores and Dynamic Indicators of Basic Early Literacy Skills Next scores were moderately correlated, fall 2015

Correlation between Kindergarten Observation Tool domain score and Dynamic Indicators of Basic Early Literacy Skills Next score (Spearman rank-order correlation coefficient)
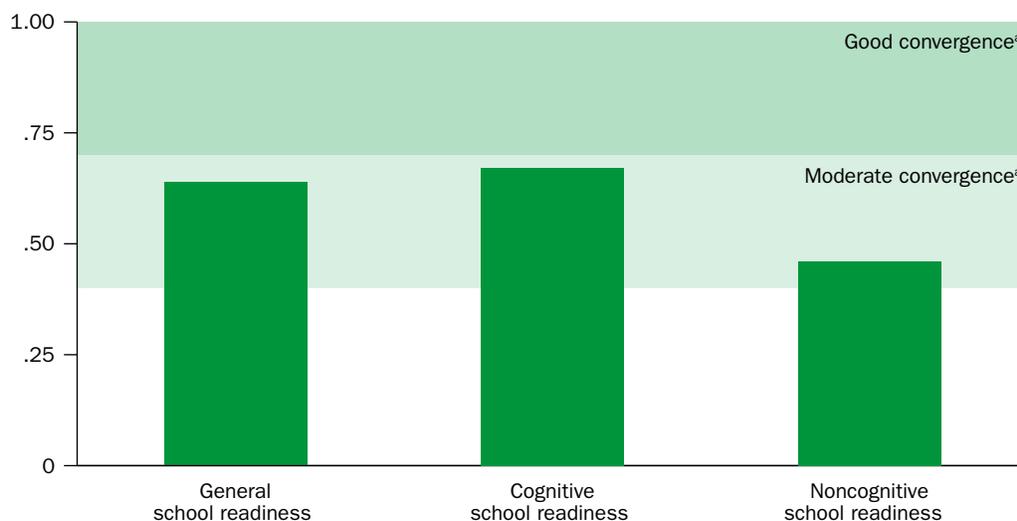
Note: All correlation coefficients were statistically significant at the $p < .01$ level after adjustment for multiple comparisons using the Bonferroni adjustment. $N = 3,257$ students.

a. Based on Di Iorio (2005).

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and data on the 2015 Dynamic Indicators of Basic Early Literacy Skills Next from the New Mexico Public Education Department.

None of the ratings for the 26 KOT items was strongly correlated with the DIBELS Next score (that is, no correlations were .70 or higher; Di Iorio, 2005), but ratings for 18 items (14 of them from the cognitive school readiness domain) were moderately correlated with the DIBELS Next score. Of the eight items for which scores were weakly correlated with the DIBELS Next score, two were from the cognitive school readiness domain. This pattern further supports the conclusion that the cognitive school readiness domain measures early literacy skills more closely than the noncognitive school readiness domain and thus that the two domains measure different underlying constructs (see table C4 in appendix C for correlation coefficients between each KOT item rating and DIBELS Next score).[2]

Classification of students as demonstrating foundational knowledge and skills at kindergarten entry on the basis of the KOT benchmark for school readiness (scoring 2.0 or higher in a domain)[3] and on the basis of the DIBELS Next benchmark for making progress in reading and being likely to achieve subsequent reading benchmarks (a composite score of 26 or higher) concurred for two-thirds of students. The proportion of students for whom classification as ready or not ready for school on the basis of the KOT and on the basis of DIBELS Next concurred was 70 percent for the KOT general school readiness domain, 73 percent for the cognitive school readiness domain, and 62 percent for the noncognitive school readiness domain (see table C5 in appendix C).

The KOT benchmark identified students who were above the benchmark more accurately than it identified students who were below the benchmark, on the basis of concurrence with the DIBELS Next benchmark. For example, the percentage of students classified as below the benchmark on both the KOT general school readiness domain and the DIBELS Next was 60 percent, and the percentage of students classified as above those benchmarks was 84 percent. This suggests that the KOT benchmark of 2.0 does not identify students who may need additional instructional support. There also were substantial differences across student subgroups in concurring classifications by KOT domains (see tables C6 and C7 in appendix C).

The KOT benchmark identified students who were above the benchmark more accurately than it identified students who were below the benchmark, on the basis of concurrence with the DIBELS Next benchmark. This suggests that the KOT benchmark of 2.0 does not identify students who may need additional instructional support

Four of twenty-six Kindergarten Observation Tool items demonstrated differential item functioning for student subgroups, which signals potential item bias

Differential item functioning occurs when two subgroups of students with equal ability, as measured by domain score, differ in the likelihood of having an individual item positively scored by the teacher. Differential item functioning signals potential item bias—the possibility that an item does not measure student ability accurately for a particular subgroup. However, differential item functioning analyses cannot confirm whether the item is functioning differently because of measure or rater bias versus real differences in discrete competencies or skills for a student subgroup.

Four of 26 KOT items demonstrated differential item functioning for one or more subgroup contrasts, meaning that the log odds ratio was 0.64 or greater (Zwick, Thayer, & Lewis, 1999). In the general school readiness domain, American Indian/Alaska Native students generally received higher ratings on fine motor skills (item 2.1) than did Hispanic students and White students who had equivalent overall general school readiness (table 2). In other words, teachers rated American Indian/Alaska Native students as demonstrating greater manual coordination in writing and crafting than Hispanic and White students who demonstrated similar overall ability across all skills measured by the KOT (see tables C8–C10 in appendix C for log odds ratios of each differential item functioning contrast

Table 2. New Mexico Kindergarten Observation Tool items that demonstrated differential item functioning, fall 2015

Item	Domain	Subgroup comparison	Favored subgroup
Fine motor skills (2.1)	General school readiness	Hispanic students versus American Indian/Alaska Native students	American Indian/Alaska Native students
		White students versus American Indian/Alaska Native students	American Indian/Alaska Native students
Cares for possessions (19.1)	General school readiness	White students versus American Indian/Alaska Native students	American Indian/Alaska Native students
Conversational ability (6.1)	Cognitive school readiness	Special education status	Students who are not in special education
Social problem solving (20.2)	Noncognitive school readiness	White students versus American Indian/Alaska Native students	White students

Note: N = 5,259 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Four of 26 KOT items demonstrated differential item functioning for one or more subgroup contrasts: fine motor skills, cares for possessions, conversational ability, and social problem solving

examined). In the general school readiness domain, American Indian/Alaska Native students also generally received higher ratings on cares for possessions (item 19.1) than did White students who had equivalent overall general school readiness.

In the cognitive school readiness domain, students who were in special education generally received lower ratings on conversational ability (item 6.1) than did students who were not in special education who had equivalent overall cognitive school readiness. In other words, teachers rated students who were not in special education as conversing more effectively than students who were in special education who demonstrated similar overall ability across all skills measured in the cognitive school readiness domain.

In the noncognitive school readiness domain, White students generally received higher ratings on social problem solving (item 20.2) than did American Indian/Alaska Native students when overall noncognitive school readiness competencies were held constant. In other words, teachers rated White students as demonstrating greater social-emotional skills in terms of negotiation, compromise, and discussion to resolve conflicts than American Indian students who demonstrated similar overall ability across all skills measured in the noncognitive school readiness domain.

Additional analyses, such as a review by a panel of child development and cultural experts, would be needed to determine whether the differences indicate bias or reflect true differences in abilities between the subgroups for these items.

Kindergarten Observation Tool rating categories were clearly distinguished from one another and ordered appropriately

The rating categories for each KOT item were functioning appropriately. The categories were (see the rubric in appendix A for descriptions):
- Making progress for 4s [four-year-olds].
- Accomplished for 4s [four-year-olds].
- Making progress for K [kindergarteners].
- Accomplished for K [kindergarteners].
- Making progress for grade 1.

First, item characteristic curves of each item indicated that no rating category was redundant (see figure 2, where none of the curve peaks for the example item rating category is subsumed within another curve). Each rating category was used to differentiate students with different levels of ability.

Second, students with higher overall ability scores typically received higher ratings on individual items (see figure 2, where the sequence of the curve peaks—with each curve representing a rating category—is ordered as expected, with peaks for higher rating categories aligned with higher difficulty levels). Teachers used the five skill rating categories of the KOT in the intended order of difficulty (see table C11 in appendix C for threshold parameter estimates—cutoff scores between adjacent rating category—for each item.)

Item characteristic curves of each item indicated that no rating category was redundant. Each rating category was used to differentiate students with different levels of ability

Figure 2. New Mexico Kindergarten Observation Tool rating categories differentiated students with different ability levels, fall 2015

Category probability
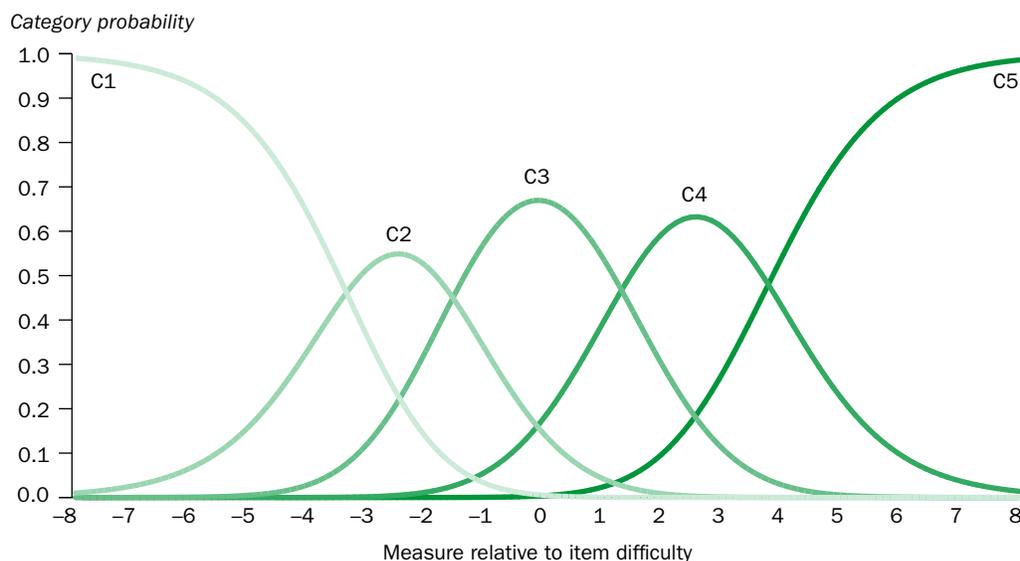
Note: This figure illustrates a typical item characteristic curve for the Kindergarten Observation Tool items. The *x*-axis represents student ability, and the *y*-axis represents the probability of a student receiving a particular rating. Darker colors represent higher rating categories. For example, if a student's ability is approximately average (0 on the *x*-axis), the student mostly likely will receive a rating of 3 (the medium-colored curve) for this item. As the ability increases, the probability of receiving a rating higher than 3 increases, and the probability of receiving a rating of 3 or below decreases. The figure indicates that all rating categories for the item are functioning appropriately—no category is redundant, and students with higher ability scores usually receive a higher rating on this particular item. *N* = 5,259 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Substantial classroom-level variation in Kindergarten Observation Tool domain scores and item ratings reduces confidence that the scores and ratings measure individual students' knowledge and skills exclusively

Approximately a third of the variation in KOT domain scores was at the classroom level (36 percent for the general school readiness domain, 36 percent for the cognitive school readiness domain, and 35 percent for the noncognitive school readiness domain). Among the same students the classroom-level variation on the DIBELS Next was smaller—approximately 12 percent—suggesting that nonrandom sorting of students with similar ability was not the sole source of classroom-level variation in the KOT scores.[4]

Classroom-level variation for KOT items ranged from 25 percent to 61 percent (figure 3; see also table C12 in appendix C). For four items—measurement (item 11.3), sorting (item 12.1), investigations (item 14.1), and earth science (item 16.1)—classroom-level variation was 50 percent or greater and was greater than variation in domain scores (the average score of all items in the domain) by more than 10 percentage points. These four items represent early math and science competencies, which have received limited attention in early childhood education settings compared with language and literacy instruction (Clements & Sarama, 2014; Early et al., 2010). Some kindergarten teachers may have less preparation to support student development in these areas, which could explain rater error. However, additional investigation would be necessary to differentiate the specific causes of variation above the student level.

Classroom-level variation in student outcomes is common among observational, indirect assessment measures (Hoyt & Kerns, 1999; Mashburn & Henry, 2004; Waterman, McDermott, Fantuzzo, & Gadsden, 2012). In other multisite studies of kindergarten students, classroom-level variation is usually less than 5 percent for measures administered

For four items— measurement, sorting, investigations, and earth science— classroom-level variation was 50 percent or greater and was greater than variation in domain scores (the average score of all items in the domain) by more than 10 percentage points

Figure 3. All New Mexico Kindergarten Observation Tool items showed variation at the classroom level, fall 2015
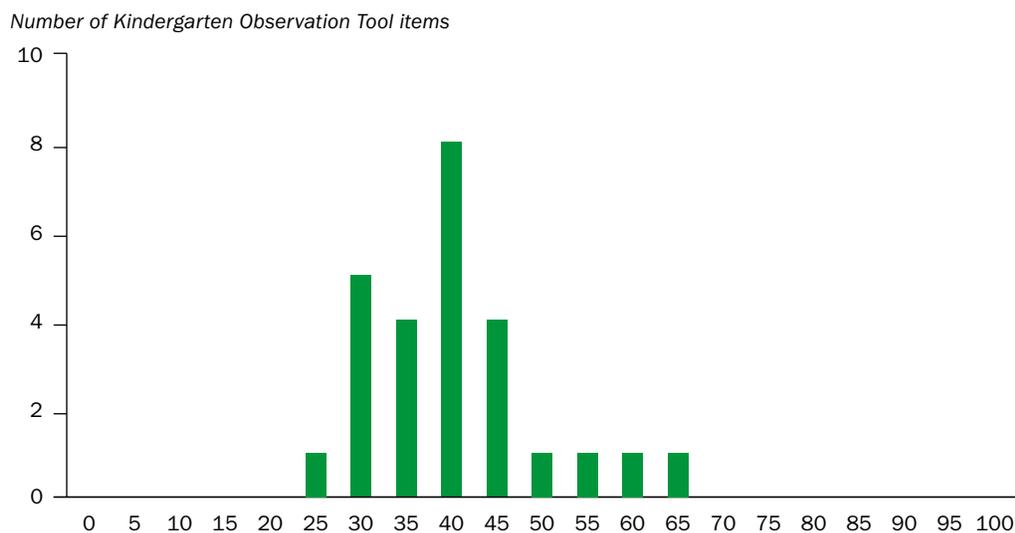
Number of Kindergarten Observation Tool items

Note: *N* = 5,259 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

by extramural assessors but 28 percent for a teacher-observation measure (specifically, the Preschool Child Observation Record) and 31 percent for a standardized direct assessment (specifically, the DIBELS Next; Waterman et al., 2012). Such variation can happen when students with similar levels of school readiness or academic ability are clustered by school or classroom. But an observational measure such as the KOT requires many subjective judgments and thus can also reflect the sensitivity of the measures to the rater, which is problematic from a measurement point of view. The greater the share of variation that is explained by the classroom or school, the lower the confidence that the KOT scores reliably measure students' true abilities.

Implications of the study findings

The findings provide preliminary empirical support for the use of the validated KOT domains by kindergarten teachers and principals to better understand individual students' knowledge and skills at kindergarten entry.

Although the study found no support for generating scores on the basis of the developer's intended six-domain structure, it found support for generating scores on the basis of general, cognitive, and noncognitive school readiness domains. The support extended to the use of these domains across student subgroups.

Although the benchmark for school readiness on the DIBELS Next partly concurred with the benchmark for school readiness on the KOT, roughly a third of students were identified inaccurately as school ready or not school ready using the benchmark on the KOT's general school readiness domain. Future investigation is required to identify the best benchmark for each KOT domain to identify students as school ready.

Four items functioned differently for some student subgroups. Further investigation, such as a review by a panel of child development and cultural experts familiar with the backgrounds and outcomes of New Mexico kindergarten students, would be needed to determine whether the differential item functioning reflected rater bias or a true difference in the skills measured by those items across the subgroups (Ercikan et al., 2010).

Variation in scores above the student level reduces the certainty of the extent to which scores measure students' true abilities. Classroom-level variation of 35–36 percent was found for KOT domain scores. That variation could be due to true differences in student ability between classrooms (for example, students with higher or lower ability clustered together), or it could reflect rater error. Comparison with DIBELS Next data suggests that only 12 percent of the KOT's classroom-level variation could be due to nonrandom sorting of students with similar ability into the same classroom. To better identify sources of classroom-level variation, future validation efforts could place multiple raters in a random subset of classrooms, so that rater error could be quantified or ruled out. Rater error can be minimized by modifying items to increase clarity and objectivity and by training and providing ongoing support to raters to improve their objectivity.

Next steps for New Mexico could include replicating construct validity analyses with future administrations of the KOT using a statewide sample, consulting with a content expert review panel to further explore item bias for students in special education and students in racial/ethnic subgroups, and further investigating the sources of classroom-level variation

Next steps for New Mexico could include replicating construct validity analyses using a statewide sample, consulting with a content expert review panel to further explore item bias, and further investigating the sources of classroom-level variation

through examination by multiple raters in a classroom or through analyses with concurrent independent measures of students' readiness for school. In addition, New Mexico could engage in a standard-setting process with a panel of content experts and local stakeholders to identify appropriate benchmark scores of school readiness for each domain.

Limitations of the study

The findings from this study are based on a large sample of students in districts that participated in the 2015 KOT field test and may not generalize to other districts or years. In the field test most teachers were in their first year of using the instrument, and teachers in their second year of using it may demonstrate more or less fidelity to the measurement protocol—for example, they may become more adept at using the instrument after additional training prior to their second year or they may deviate from instructions about how to use the measure as time passes from the initial training.

The findings from this study are based on a large sample of students in districts that participated in the 2015 KOT field test and may not generalize to other districts or years

The proportion of students eligible for the federal school lunch program was lower in the 2015 KOT field test sample than in the statewide population, even though a higher proportion might have been expected to be eligible because the field test focused on schools that offered publicly funded prekindergarten and were thus more likely to be in Title I school zones and have more eligible students. Overall, teachers in districts that participated in the 2015 KOT field test were less likely to complete the KOT for students who were eligible for the federal school lunch program than for students who were not eligible. Educator credentials and instructional supports are often lower in districts serving students eligible for the federal school lunch program (Hanushek, Kain, & Rivkin, 2004; Johnson, Kardos, Kauffman, Liu, & Donaldson, 2004), and educators in these districts may use the KOT differently from how educators in other New Mexico districts use it.

The results are based on analyses of the 2015 version of the KOT. The New Mexico Public Education Department modified the KOT for the 2016 administration by increasing the number of rating categories. Analysis of subsequent KOT data is needed to confirm the validity and reliability of the revised instrument.

Finally, students' DIBELS Next scores were not linked to classroom teachers in the same way that students' KOT scores were linked to classrooms. Because both assessments were completed at the same time of year, estimates of classroom-level variation in DIBELS Next scores were based on the assumption that the same teacher completed both the DIBELS Next assessment and the KOT for each student.

Appendix A. New Mexico Kindergarten Observation Tool essential indicator rubrics

This appendix includes the 2015 New Mexico Kindergarten Observation Tool.

2015

Kindergarten Essential Indicator Rubrics

Field Test

NEW MEXICO
Kindergarten
Observation Tool

Table of Contents

New Mexico Kindergarten Observation Tool Essential Indicator Rubrics

Table of Contents
(continued)

New Mexico Kindergarten Observation Tool Essential Indicator Rubrics

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 1 > Physical Development, Health, and Well Being

Outcome 1 > The child uses gross motor control independently, including balance, spatial awareness, and stability.

Indicator 1.1 > Exhibits body coordination and strength in activities such as climbing stairs with alternating feet, marching, running, jumping, hopping, dancing, riding tricycles and scooters.

Aligned New Mexico Kindergarten Standards:

Physical Education Standards

Content Standard 1: Demonstrates competency in many movement forms and proficiency in a few movement forms. Students will:

    K-4 Benchmark 1: Demonstrate competency in selected motor skills.

        K-2.1. travel in a variety of locomotor patterns (i.e., hop, skip, jump, gallop, slide, etc.) using mature form;

        K-2.2. demonstrate skills of chasing, fleeing and dodging to avoid others; and

        K-2.3. demonstrate smooth transitions between sequential motor skills (i.e., running into a jump).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Climbs on age-appropriate playground equipment with minimal adult assistance, walks, runs, jumps, marches, and hops.	Climbs on age-appropriate playground equipment independently, walks, runs, jumps, marches, hops, and gallops.	Demonstrates gross motor control by transitioning smoothly between movements (e.g., running into a jump) and attempting to skip independently.	Exhibits gross motor coordination and strength in a variety of activities and movements, including skipping using mature form.	Consistently exhibits gross motor coordination and strength in the age-appropriate range of activities and movements in play and complex games.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 1 > Physical Development, Health, and Well Being

Outcome 2 > The child independently uses fine motor skills.

Indicator 2.1 > Is developing manual coordination to use cutting and writing tools and demonstrate self-help skills such as buttoning and zipping.

Rubric 2.1a > Develops manual coordination to use cutting and writing tools.

Aligned New Mexico Kindergarten Standards:
Art Content Standards for Visual Arts
Standard 1: Learn and develop skills and meet technical demands unique to dance, music, theatre/drama and visual arts.
 K-5 Benchmark 1B: Explore and develop skills using art materials, tools and techniques
 K-1.1. Use a variety of art materials and related skills.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Rubric 2.1a				
Uses writing and crafting tools (e.g., pencils and scissors) with adult guidance and support.	Uses writing and crafting tools with minimal support, but may use an incorrect grip.	Demonstrates fine motor control in using writing and crafting tools independently using correct grip.	Exhibits fine motor coordination in using a variety of writing and crafting tools independently so that work products have detail.	Consistently exhibits fine motor coordination and skill in using a variety of writing and crafting tools to create intricately detailed work products.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 5 > The child demonstrates development and expansion of listening skills.

Indicator 5.2 > Follows increasingly complex directions.

Aligned New Mexico Kindergarten Standards:
N/A

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Follows simple directions that involve one or two steps with adult modeling and support (e.g., "Sit in your chair," or, "After you finish your drawing, put it in your folder").	Follows simple directions that involve two steps in a series of unrelated sequences of action with some prompting (e.g., "After you put your paper in your folder, walk to line up by the door").	Follows directions that involve two or more steps in a series of unrelated sequences of action most of the time, but may require minimal prompting (e.g., "Listen to the story and then draw a picture showing an important event in the story. After you have finished your drawing, put it in your folder").	Follows multi-step directions in a series of unrelated sequences of action independently, without prompting between steps.	Retains multi-step directions for activities, discussions, or projects over an extended period of time and follows through with them independently (e.g., follows the steps of the writing process over the period of a multi-day assignment without the need for prompting between each step).

3

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 5 > The child demonstrates development and expansion of listening skills.

Indicator 5.3 > Hears and discriminates the sounds of language in words to develop phonological awareness.

Rubric 5.3a > Recognizes rhyming sounds in spoken language.

Rubric 5.3b > Knows and applies letter-sound correspondence and beginning sound-recognition skills.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RF.K.2 Demonstrate understanding of spoken words, syllables, and sounds (phonemes).

　　RF.K.2.A Recognize and produce rhyming words.

　　RF.K.2.D Isolate and pronounce the initial, medial vowel, and final sounds (phonemes) in three-phoneme (consonant-vowel-consonant, or CVC) words. (*This does not include CVCs ending with /l/, /r/, or /x/.)

　　RF.K.2.E Add or substitute individual sounds (phonemes) in simple, one-syllable words to make new words.

RF.K.3 Know and apply grade-level phonics and word analysis skills in decoding words.

　　RF.K.3.A Demonstrate basic knowledge of one-to-one letter-sound correspondences by producing the primary sound or many of the most frequent sounds for each consonant.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Rubric 5.3a				
Recognizes rhyming sounds in familiar songs or stories.	Identifies words that rhyme when presented with pairs of words.	Identifies words that rhyme and do not rhyme when presented with a small set of words.	Consistently distinguishes words that rhyme from those that do not rhyme. Produces a rhyming word when presented with a simple consonant-vowel-consonant word.	Recognizes rhyming patterns in texts and is able to make predictions about what words might come next in shared reading activities involving rhymes. Produces rhymes in writing activities.
Rubric 5.3b				
Makes some letter-sound associations (e.g., recognizes words that start with the same letter/sound as own name).	Makes simple letter-sound associations with beginning consonants with adult support.	Makes simple letter-sound associations with beginning and ending consonants.	Makes letter-sound associations with medial (vowel) sounds in CVC words. Adds or substitutes individual sounds in CVC words to create new words.	Makes letter-sound associations in words with 4 or more letters. Adds or substitutes individual sounds, digraphs, and blends in 4-letter words to create new words.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 5 > The child demonstrates development and expansion of listening skills.

Indicator 5.4 > Demonstrates understanding of new vocabulary introduced in conversations, activities, stories, or books.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RL.K.4 Ask and answer questions about unknown words in a text.

RI.K.4 With prompting and support, ask and answer questions about unknown words in a text.

L.K.4 Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on kindergarten reading and content.

L.K.5 With guidance and support from adults, explore word relationships and nuances in word meanings.

     L.K.5.D Distinguish shades of meaning among verbs describing the same general action (e.g., *walk, march, strut, prance*) by acting out the meanings.

L.K.6 Use words and phrases acquired through conversations, reading and being read to, and responding to texts.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Demonstrates understanding of vocabulary that includes basic category names and related words (e.g., toys: car/truck; colors: red/blue).	Demonstrates understanding of vocabulary that includes basic concepts with related words (e.g., comparison: less/more/same) and descriptors (i.e., adjectives and adverbs).	Demonstrates understanding of vocabulary that includes specialized areas of interest (e.g., vocabulary related to a unit of study at school).	Demonstrates understanding of vocabulary and phrases that distinguish shades of meaning among verbs describing the same type of action (e.g., write/draw, walk/run).	With guidance and support, uses context clues and/or applies knowledge of affixes, roots, and word relationships in order to access higher-level vocabulary.

Note: Students may "demonstrate understanding of vocabulary" by exhibiting the skills listed in the kindergarten standards.

5

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 6 > The child communicates experiences, ideas, and feelings through speaking.

Indicator 6.1 > Converses effectively in his or her home language, English, or Sign language for a variety of purposes relating to real experiences and different audiences.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

SL.K.1 Participate in collaborative conversations with diverse partners about *kindergarten topics and texts* with peers and adults in small and larger groups.

SL.K.1.A Follow agreed-upon rules for discussions (e.g., listening to others and taking turns speaking about the topics and texts under discussion).

SL.K.1.B Continue a conversation through multiple exchanges.

SL.K.3 Ask and answer questions in order to seek help, get information, or clarify something that is not understood.

SL.K.4 Describe familiar people, places, things, and events and, with prompting and support, provide additional detail.

SL.K.6 Speak audibly and express thoughts, feelings, and ideas clearly.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Uses five- to six-word sentences to express ideas.	Uses complex questions and/or statements of seven or more words to present and get information.	Uses two or three connected sentences to express ideas and reply with relevant information to questions and comments of others.	Converses effectively through five or more exchanges. Expresses ideas clearly and completely, using complete sentences. Uses multiple sentences to articulate an idea.	Converses effectively through five or more on-topic exchanges. Expresses ideas clearly and completely, using detail and elaboration to communicate ideas in complete sentences.

6

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 7 > The child engages in activities that promote the acquisition of emergent reading skills.

Indicator 7.1 > Demonstrates an interest and enjoyment in books, listening to stories read aloud, and/or looking at books using illustrations or familiar text.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RL.K.2 With prompting and support, retell familiar stories, including key details.

RL.K.7 With prompting and support, describe the relationship between illustrations and the story in which they appear (e.g., what moment in a story an illustration depicts).

RL.K.10 Actively engage in group reading activities with purpose and understanding.

RI.K.2 With prompting and support, identify the main topic and retell key details of a text.

RI.K.7 With prompting and support, describe the relationship between illustrations and the text in which they appear (e.g., what person, place, thing, or idea in the text an illustration depicts).

RI.K.10 Actively engage in group reading activities with purpose and understanding.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Demonstrates interest in an adult reading a story/text and chooses familiar books to look at independently.	Actively attends to stories/texts during a read aloud and begins to show interest in retelling familiar stories/texts, using the illustrations as a guide.	With prompting and support, retells familiar stories and/or key details in a text, using the illustrations as a guide (may not be accurate). Begins to recognize familiar words in texts that correspond to the illustrations.	With prompting and support, retells familiar stories and/or key details in a text with accuracy, using the illustrations and familiar words as guides.	Engages in retelling and discussing familiar stories in the context of reading and writing activities.

7

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 7 > The child engages in activities that promote the acquisition of emergent reading skills.

Indicator 7.2 > Demonstrates comprehension of a story read aloud by asking relevant questions or making pertinent comments.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RL.K.1 With prompting and support, ask and answer questions about key details in a text.
RL.K.10 Actively engage in group reading activities with purpose and understanding.
RI.K.1 With prompting and support, ask and answer questions about key details in a text.
RI.K.10 Actively engage in group reading activities with purpose and understanding.
SL.K.2 Confirm understanding of a text read aloud or information presented orally or through other media by asking and answering questions about key details and requesting clarification if something is not understood.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Asks and answers basic factual questions about a story during a read aloud (e.g., What is happening on this page?). Makes comments that relate to the story, but may sometimes get off topic.	Asks and answers factual questions about a story during a read aloud (e.g., What is the main character's name? What happened first?) and begins to ask and answer inferential questions (e.g., What might happen next?). Makes comments that are relevant to the story and stays on topic.	Asks and answers factual and inferential questions about a story during a read aloud (e.g., How is the main character feeling and what happened to make him feel that way?). Makes comments that demonstrate a sense of story (e.g., identifying beginning, middle, and end; naming characters; discussing key details of plot).	Asks and answers factual, inferential, and critical thinking questions about a story during a read aloud (e.g., Based on what you know about the main character, what do you think he is going to do next? What in the story makes you think that?). Makes comments that demonstrate critical thinking related to the story.	Asks and answers factual, inferential, and critical thinking questions about a story during a read aloud, extending such practice to independent reading activities. Responds to prompts that demonstrate understanding of stories in writing.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 7 > The child engages in activities that promote the acquisition of emergent reading skills.

Indicator 7.3 > Progresses in understanding and using conventions of reading (including holding book upright, identifying front and back, turning pages correctly, and recognizing that print proceeds from left to right).

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RI.K.5 Identify the front cover, back cover, and title page of a book.

RF.K.1 Demonstrate understanding of the organization and basic features of print.

　　RF.K.1.A Follow words from left to right, top to bottom, and page by page.

　　RF.K.1.C Understand that words are separated by spaces in print.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Holds a book upright and attempts to turn pages. Identifies the first page to begin reading (i.e., first page of text).	Holds and handles a book correctly (turns pages independently). Identifies the front cover, back cover, and title of a book. Distinguishes pictures from print.	Identifies the first word on the first page as the place to begin reading. Demonstrates some understanding of directionality in a text (i.e., follows words in a book from left to right, top to bottom, and page by page), but may skip lines or miss pages when attempting to read independently.	Identifies book elements with ease and follows rules related to directionality proficiently when reading appropriately leveled texts. Demonstrates an understanding of one-to-one correspondence in a text (e.g., through pointing to the words while reading).	Demonstrates understanding of the organization and basic features of print, and begins to recognize more distinguishing features of printed words and sentences (e.g., punctuation).

9

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 7 > **The child engages in activities that promote the acquisition of emergent reading skills.**

Indicator 7.4 > Progresses in understanding of alphabet knowledge and word recognition skills.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

RF.K.1 Demonstrate understanding of the organization and basic features of print.

    RF.K.1.D Recognize and name all upper- and lowercase letters of the alphabet.

RF.K.3 Know and apply grade-level phonics and word analysis skills in decoding words.

    RF.K.3.C Read common high-frequency words by sight (e.g., *the, of, to, you, she, my, is, are, do, does*).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Names and identifies some uppercase letters (e.g., 3-5 letters) with personal significance (e.g., letters in own name).	Names and identifies many uppercase letters (e.g., at least 12 uppercase and 12 lowercase letters).	Names and identifies most uppercase letters (e.g., at least 20 uppercase letters and 20 lowercase letters).	Names and identifies all uppercase and lowercase letters (including variations of a and g; a/ɑ, g/ɡ) and recognizes familiar combinations of letters in order to read common high-frequency words (e.g., 15-35 words).	Recognizes a variety of high-frequency words with regular and irregular sound-symbol correspondence (e.g., reads more than 100 words by sight).

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 2 > Literacy

Outcome 8 > The child engages in activities that promote the acquisition of emergent writing skills.

Indicator 8.3 > Increasingly attempts to represent meaningful words and print in the environment using the early stages of writing.

Aligned New Mexico Kindergarten Standards:

English Language Arts Common Core State Standards

L.K.1 Demonstrate command of the conventions of standard English grammar and usage when writing or speaking.

L.K.1.A Print many upper- and lowercase letters.

L.K.2 Demonstrate command of the conventions of standard English capitalization, punctuation, and spelling when writing.

L.K.2.D Spell simple words phonetically, drawing on knowledge of sound-letter relationships.

W.K.1 Use a combination of drawing, dictating, and writing to compose opinion pieces in which they tell a reader the topic or the name of the book they are writing about and state an opinion or preference about the topic or book (e.g., *My favorite book is...*).

W.K.2 Use a combination of drawing, dictating, and writing to compose informative/explanatory texts in which they name what they are writing about and supply some information about the topic.

W.K.3 Use a combination of drawing, dictating, and writing to narrate a single event or several loosely linked events, tell about the events in the order in which they occurred, and provide a reaction to what happened.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Uses drawing to represent ideas on paper, though ideas may be unclear and require dictation. May attempt to form letters that are familiar (e.g., letters in name).	Uses drawing to represent ideas on paper with some level of clarity, sometimes relying on dictation to clarify content. Forms letters in first and last name to sign drawings and other written work.	Uses a combination of drawing, dictating, and writing to represent ideas on paper. Uses labeling that demonstrates some knowledge of phonics (e.g., draws family and labels each member with the letter that corresponds with the first sound in their names).	Uses a combination of drawing, dictating, and writing (relying mostly on independent writing skills) to represent ideas on paper in some detail. Uses labeling and simple phonetically spelled words to articulate meaning.	Writes to convey meaning. Spells simple words phonetically and begins to apply grade-level phonics skills (e.g., spells words with consonant digraphs and uses conventions for representing long vowel sounds).

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 3 > Numeracy

Outcome 9 > The child understands numbers, ways of representing numbers, and relationships between quantities and numerals.

Indicator 9.1 > Uses numbers and counting as means for solving problems and determining quantity.

Aligned New Mexico Kindergarten Standards:

Mathematics Common Core State Standards

K.CC.B.4 Understand the relationship between numbers and quantities; connect counting to cardinality.

> K.CC.B.4.A When counting objects, say the number names in the standard order, pairing each object with one and only one number name and each number name with one and only one object.

> K.CC.B.4.B Understand that the last number name said tells the number of objects counted. The number of objects is the same regardless of their arrangement or the order in which they were counted.

> K.CC.B.4.C Understand that each successive number name refers to a quantity that is one larger.

K.CC.B.5 Count to answer "how many?" questions about as many as 20 things arranged in a line, a rectangular array, or a circle, or as many as 10 things in a scattered configuration; given a number from 1–20, count out that many objects.

K.OA.A.1 Represent addition and subtraction with objects, fingers, mental images, drawings, sounds (e.g., claps), acting out situations, verbal explanations, expressions, or equations.

K.OA.A.2 Solve addition and subtraction word problems, and add and subtract within 10, e.g., by using objects or drawings to represent the problem.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Demonstrates one-to-one correspondence (e.g., pairs each object with one and only one number name and each number name with one and only one object) in counting 1–10 objects in a group when arranged in a line.	Demonstrates one-to-one correspondence in counting 1–15 objects in a group when arranged in a line, a rectangular array, or a circle. Recognizes that each successive number name refers to a quantity that is one larger.	When given a number from 1–20, counts out that many objects and begins to solve problems involving joining and combining using small quantities of objects (i.e., totals of up to 5).	Solves problems involving joining, separating, and combining using small quantities of objects (i.e., totals of up to 10).	Relates counting to addition and subtraction by counting on (i.e., by counting on 2 to add 2; e.g., student counts 5 objects, 2 objects are added to the group, and student adds 2 by counting on to 6, 7 rather than starting the count over at 1).

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 3 > Numeracy

Outcome 9 > The child understands numbers, ways of representing numbers, and relationships between quantities and numerals.

Indicator 9.3 > Progresses in understanding of number words and numeral recognition skills.

Rubric 9.3a > States number words in sequence when counting.

Rubric 9.3b > Names and identifies written numerals.

Aligned New Mexico Kindergarten Standards:

Mathematics Common Core State Standards

K.CC.A.1 Count to 100 by ones and by tens.

K.CC.A.2 Count forward beginning from a given number within the known sequence (instead of having to begin at 1).

K.CC.A.3 Write numbers from 0 to 20. Represent a number of objects with a written numeral 0-20 (with 0 representing a count of no objects).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Rubric 9.3a				
States number words in sequence from 1 to 10 when counting by ones.	States number words in sequence from 1 to 30 when counting by ones.	States number words in sequence from 1 to 100 when counting by ones.	States number words in sequence to 100 when counting by tens.	States number words in sequence to 100 by twos or fives.
Rubric 9.3b				
Recognizes numerals distinctly from letters or other symbols in print, calling them "numbers."	Names and identifies a few written numerals with personal significance (e.g., numeral representing own age).	Names and identifies written numerals from 0 to 10.	Names, identifies, and writes numerals from 0 to 20.	Identifies place value of digits in written numerals and writes numerals from 0 to 100.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 3 > Numeracy

Outcome 11 > The child demonstrates an understanding of non-standard units to measure and make comparisons.

Indicator 11.3 > Demonstrates emerging knowledge of measurement.

Aligned New Mexico Kindergarten Standards:

Mathematics Common Core State Standards

K.MD.A.1 Describe measurable attributes of objects, such as length or weight. Describe several measurable attributes of a single object.

K.MD.A.2 Directly compare two objects with a measurable attribute in common, to see which object has "more of"/"less of" the attribute, and describe the difference. *For example, directly compare the heights of two children and describe one child as taller/shorter.*

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Demonstrates emerging understanding of measurement by describing attributes of singular objects (uses simple words to describe size and length).	Describes measurable attributes of singular objects (size, length, weight, and capacity) with accuracy.	Compares two objects based on a measurable attribute and explains how they are different (e.g., bigger/smaller, taller/shorter, heavier/lighter, more full/less full).	Orders three or more objects by a measurable attribute (e.g., smallest to biggest).	Uses standard and non-standard measurement tools to compare sets of objects based on measurable attributes.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 3 > Numeracy

Outcome 12 > The child demonstrates the ability to investigate, organize, and create representations.

Indicator 12.1 > Sorts, classifies, and groups materials by one or more characteristics.

Aligned New Mexico Kindergarten Standards:

Mathematics Common Core State Standards

K.MD.B.3 Classify objects into given categories; count the numbers of objects in each category and sort the categories by count.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Begins to sort and classify objects based on one characteristic (e.g., color).	Determines a classification scheme for a collection of objects that creates a group for every item and tells about the groups.	Sorts and classifies objects into groups by one characteristic and compares the number of objects in the groups using comparison vocabulary (e.g., more/less).	Sorts and classifies a group of objects by more than one characteristic (i.e., is able to re-sort and re-classify a group based on different characteristics). Counts the number of objects in each category and sorts the categories by count.	Sorts and classifies a group of objects by more than one characteristic into multiple categories. Counts the number of objects in each category and sorts the categories by count.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 5 > Scientific Conceptual Understandings

Outcome 14 > The child uses the scientific method to investigate the physical and natural worlds and to hypothesize and make predictions.

Indicator 14.1 > Uses senses to investigate characteristics and behaviors in the physical and natural worlds and begins to form explanations of observations and explorations.

Aligned New Mexico Kindergarten Standards:

Science Standards

Standard I: Understand the processes of scientific investigations and use inquiry and scientific ways of observing, experimenting, predicting and validating to think critically.
    K-4 Benchmark I: Use scientific methods to observe, collect, record, analyze, predict, interpret, and determine reasonableness of data.
        K.1. Use observation and questioning skills in science inquiry (e.g., What happens when something is pushed or pulled?).
        K.2. Ask and answer questions about surroundings and share findings with classmates.
        K.3. Record observations and data with pictures, numbers, and/or symbols.
    K-4 Benchmark II: Use scientific thinking and knowledge and communicate findings.
        K.1. Communicate observations and answer questions about surroundings.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Demonstrates an interest in the surrounding world, using senses to observe and explore surroundings.	Uses two or more senses (e.g., both sight and smell or both hearing and touch) to explore the world and makes one or more detailed comments describing sensory experiences.	Participates in small hands-on multisensory experiments with adult guidance and uses observation and questioning skills to investigate and draw conclusions.	Participates in a variety of hands-on multisensory experiments with adult guidance. Uses observation and questioning skills in order to draw conclusions. Demonstrates an ability to record and analyze data through drawing and writing.	Participates in a variety of hands-on multisensory experiments that require high-level observational skills (e.g., explores cause and effect relationships). Demonstrates an ability to record and analyze data through writing, charting, and graphing.

16

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 5 > Scientific Conceptual Understandings

Outcome 16 > **The child acquires scientific knowledge related to earth science.**

Indicator 16.1 > Investigates, compares, and contrasts seasonal and weather changes in the immediate environment.

Aligned New Mexico Kindergarten Standards:

Science Standards

Standard III (Earth and Space Science): Understand the structure of Earth, the solar system, and the universe, the interconnections among them, and the processes and interactions of Earth's systems.

　　K-4 Benchmark II: Know the structure and formation of Earth and its atmosphere and the processes that shape them.

　　　　K.1. Observe that changes in weather occur from day to day and season to season.

　　　　K.2. Observe that the sun warms the land and water, and they warm the air.

Social Studies Standards

Content Standard II: Students understand how physical, natural, and cultural processes influence where people live, the ways in which people live, and how societies interact with one another and their environments.

　　K-4 Benchmark II-C: Be familiar with aspects of human behavior and man-made and natural environments in order to recognize their impact on the past and present.

　　　　K.2. Describe the natural characteristics of places (e.g., landforms, bodies of water, natural resources, and weather).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Observes how weather can change within a day or from day to day (e.g., "It was raining this morning, and now it is not raining," or, "It was colder yesterday than it is today").	Observes patterns regarding weather over an extended period of time (e.g., "It has been windy all week") and begins to understand how weather relates to seasons (e.g., understanding that snow comes in winter).	Observes, records, and describes patterns regarding weather and the effects on the immediate environment (e.g., understanding that rain over a period of days may cause flooding).	Investigates ways in which weather variables (e.g., temperature, precipitation, wind) affect us or cause changes to Earth's features (e.g., stream has greater water flow after snow melts).	Identifies and describes different climates and how weather affects climate (e.g., desert, arctic, rainforest).

17

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 6 > Self, Family, and Community

Outcome 18 > The child develops self control.

Indicator 18.1 > Adapts behavior to fit different situations (for example, accepts transitions, follows daily routines, and/or incorporates cultural expectations).

Aligned New Mexico Kindergarten Standards:

Social Studies Standards

Content Standard III: Students understand the ideals, rights, and responsibilities of citizenship and understand the content and history of the founding documents of the United States with particular emphasis on the United States and New Mexico constitutions and how governments function at local, state, tribal, and national levels.

      K-4 Benchmark III-D: Understand rights and responsibilities of "good citizenship" as members of a family, school and community.

            K.2. Explain what is meant by "good citizenship," to include:

                  a. taking turns and sharing

                  b. taking responsibility for own actions, assignments, and personal belongings within the classroom and respecting the property of others.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Follows basic routines for pre-K (e.g., transitioning between activities with one-step directions) and complies with basic expectations for behavior (e.g., "Keep your hands to yourself"), but may require frequent adult prompting and support.	Follows basic routines for pre-K and K (e.g., transitioning between activities with one- or two-step directions) and complies with basic expectations for behavior (e.g., "Stay in your chair"), but sometimes requires adult prompting.	Follows complicated routines (e.g., activities that include multiple steps) and high-level expectations for behavior ("Treat your peers with respect, and show responsibility for your learning") with minimal prompting.	Internalizes classroom routines and behaves in socially acceptable ways without the need for prompting.	Serves as a model for behavior and contributes to classroom culture in positive ways (e.g., providing support to peers who may be struggling in a learning activity).

18

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 6 > Self, Family, and Community

Outcome 19 > The child demonstrates personal responsibility.

Indicator 19.1 > Cares for personal and group possessions.

Aligned New Mexico Kindergarten Standards:

Social Studies Standards

Content Standard III: Students understand the ideals, rights, and responsibilities of citizenship and understand the content and history of the founding documents of the United States with particular emphasis on the United States and New Mexico constitutions and how governments function at local, state, tribal, and national levels.

    K-4 Benchmark III-D: Understand rights and responsibilities of "good citizenship" as members of a family, school and community.

        K.2. Explain what is meant by "good citizenship," to include:

            a. taking turns and sharing

            b. taking responsibility for own actions, assignments, and personal belongings within the classroom and respecting the property of others.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Places personal items in designated space without assistance (may need reminding), and participates in cleanup time with some independence and some adult help.	Places personal items in designated space without assistance, and participates in cleanup time independently (without adult help) almost every day.	Routinely demonstrates responsibility in caring for personal and group possessions with minimal prompting, and may engage peers to assist with care of classroom environment.	Internalizes expectations around caring for personal and group possessions, demonstrating a consistently high level of responsibility by exercising reasonable care and returning found items to their proper places or owners.	Practices citizenship in all areas of the school, demonstrating a consistently high level of responsibility that extends beyond what they are personally responsible for (e.g., picks up garbage on the playground that they did not put there).

19

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 6 > Self, Family, and Community

Outcome 20 > The child works cooperatively with other children and adults.

Indicator 20.1 > Plays and interacts with various children, sharing experiences and ideas with others.

Aligned New Mexico Kindergarten Standards:

Physical Education Standards

Content Standard 5: Demonstrates responsible personal and social behavior in physical activity settings. Students will:

    K-4 Benchmark 2: work cooperatively and productively with a partner or small group:

        K-2.1. invite a peer to take his turn at a piece of apparatus before repeating turn; and

        K-2.2. assist partner by sharing observations about skill performance during practice.

Content Standard 6: Demonstrates understanding and respect for differences among people in physical activity settings. Students will:

    K-4 Benchmark 2: recognize the talents that individuals with differences can bring to group activities:

        K-2.1. work productively with a variety of partners.

    K-4 Benchmark 3: experience differences and similarities among people of different backgrounds by participating in activities of national, cultural and ethnic origins:

        K-2.1. accept all playmates without regard to personal differences (i.e., age, race, ethnicity, gender, ability level, etc.).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Interacts with other children, sharing objects, conversation, and ideas to cooperate in play activities.	Develops or extends themes in cooperative work and play activities.	Takes turns being a leader and group member in cooperative play and work.	Participates in cooperative play and work projects as leader or group member (e.g., accepts direction from peer when in "member" role and provides positive direction when in "leader" role).	Demonstrates flexibility and maturity in interactions with other children. Coordinates roles effectively, considering each group member's individual strengths, and shows leadership in activities when appropriate.

20

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 6 > Self, Family, and Community

Outcome 20 > The child works cooperatively with other children and adults.

Indicator 20.2 > Uses and accepts negotiation, compromise, and discussion to resolve conflicts.

Aligned New Mexico Kindergarten Standards:

Physical Education Standards

Content Standard 5: Demonstrates responsible personal and social behavior in physical activity settings. Students will:

    K-4 Benchmark 3: recognize the influence of peer pressure and identify ways of resolving conflict:

        K-2.1. demonstrate the elements of socially acceptable conflict resolution; and

        K-2.2. demonstrate effective communication skills.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Generates own ideas of appropriate ways to handle conflict with guidance from teacher and comes to an agreeable solution.	Tries to work through conflicts with peers in appropriate ways (may or may not need adult help).	Frequently initiates and completes conflict resolution successfully, with minimal adult assistance.	Independently negotiates, compromises, and discusses conflict with success on a regular basis.	Models positive ways to resolve conflict for peers, stepping in to support others when appropriate.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 6 > Self, Family, and Community

Outcome 21 > The child develops relationships of mutual trust and respect with others.

Indicator 21.2 > Accepts guidance and direction from classroom and school personnel and seeks their support when needed.

Aligned New Mexico Kindergarten Standards:

Social Studies Standards

Content Standard III: Students understand the ideals, rights, and responsibilities of citizenship and understand the content and history of the founding documents of the United States with particular emphasis on the United States and New Mexico constitutions and how governments function at local, state, tribal, and national levels.

K-4 Benchmark III-A: Know the fundamental purposes, concepts, structures, and functions of local, state, tribal, and national governments.

K.1. Identify authority figures and describe their roles (e.g., parents, teachers, principal, superintendent, police, public officials).

Health Standards

Content Standard 1: Students will comprehend concepts related to health promotion and disease prevention. Students will:

K-4 Benchmark 4: describe how physical, social and emotional environments influence personal health:

K.3. know how to access help (e.g., dial 911 in an emergency, trusted adult).

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Accepts guidance and support from classroom personnel, but may not actively seek support when needed (e.g., may abandon an activity when frustrated without asking for help).	Accepts guidance and direction from classroom personnel, but compliance with their directions may vary depending on the student's perception of their authority (e.g., may follow all directions from the primary classroom teacher but behave differently with another supervising adult (e.g., a substitute)).	Accepts guidance and support from classroom and school personnel, appropriately recognizing their authority. Actively seeks support from adults when needed, but may not identify the appropriate authority figure to help in every situation or articulate a need clearly.	Accepts guidance and support from all classroom and school personnel. Identifies the appropriate authority figure for support most of the time and articulates a need with some level of clarity.	Accepts guidance and direction from all classroom and school personnel. Consistently identifies the appropriate authority figure for support and articulates a need clearly, only after attempting to try something independently.

22

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 7 > Approaches to Learning

Outcome 24 > The child takes initiative.

Indicator 24.2 > Develops increasing independence during activities, routines, and play.

Aligned New Mexico Kindergarten Standards:

Physical Education Standards

Content Standard 5: Demonstrates responsible personal and social behavior in physical activity settings. Students will:

K-4 Benchmark 4: work independently and on-task for short periods of time:

K-2.1. demonstrate independent work habits during short-term activity.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Requires adult guidance and support during activities, routines, and play, but shows interest in trying things independently.	Self-initiates activities and play and shows increasing independence in routines, calling on adults when help is needed.	Demonstrates independence during activities, routines, and play. Attempts to assist peers or asks for peer support during activities, routines, and play before calling on adults for help.	Maintains independence during activities, routines, and play over extended periods of time. Works collaboratively with peers to overcome problems, calling on adults only when necessary.	Sets a goal, follows through, and maintains concentration during learning activities, routines, and play alone or with others.

New Mexico Kindergarten Observation Tool Essential Indicators

Domain 7 > Approaches to Learning

Outcome 27 > The child displays persistence and pursues challenges.

Indicator 27.1 > Focuses and completes a variety of tasks, activities, projects, and experiences.

Aligned New Mexico Kindergarten Standards:

Physical Education Standards
Content Standard 5: Demonstrates responsible personal and social behavior in physical activity settings. Students will:
K-4 Benchmark 4: work independently and on-task for short periods of time:
K-2.1. demonstrate independent work habits during short-term activity.

Health Standards
Content Standard 6: Students will demonstrate the ability to use goal-setting and decision-making skills to enhance health. Students will:
K-4 Benchmark 1: demonstrate the ability to apply a decision-making process to health issues and problems:
K .1. list steps in the decision-making process.

4-Year-Old Rubric		Kindergarten Rubric		Grade 1 Rubric
Making Progress for 4s	**Accomplished for 4s (First Steps for K)**	**Making Progress for K**	**Accomplished for K (First Steps for Grade 1)**	**Making Progress for Grade 1**
Maintains focus for 5 to 10 minutes and attempts to solve problems that arise, but may be easily distracted.	Maintains focus for 10–15 minutes at a time and attempts to complete tasks and activities, but may get frustrated or distracted at times and abandon progress.	Maintains focus for 15–25 minutes at a time and works to complete tasks and activities, ignoring most distractions, but may call on adults for support.	Maintains focus for 25–40 minutes at a time and persists in completing tasks and activities independently regardless of distractions, only calling on adults for support after attempting more than one strategy.	Maintains focus for more than 40 minutes at a time and persists in completing complicated tasks and activities independently, continuing to persevere regardless of distractions and through multiple attempted strategies.

24

Appendix B. Data and methodology

This appendix describes the data sources and details the study methodology.

Data

The data used for the study from the 2015 field test of the New Mexico Kindergarten Observation Tool (KOT) and the fall 2015 Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next assessment were provided by the New Mexico Public Education Department. Student demographic data were obtained from the New Mexico Public Education Department through the Student Teacher Accountability Reporting System (STARS). The two datasets were merged using a student identification variable.

A total of 134,653 KOT item ratings were available for 5,305 students. After merging the KOT data with the STARS data, problematic observations at the rating level were excluded for the following reasons:

- One student had two identical ratings with the same observation identification number (52 observations were dropped, no students were dropped).
- Ratings were invalid, such as –1, which represented a student being exempt from that item (620 observations were dropped, eight students were dropped). Exemptions were given for students with many absences, late registrations, Individualized Education Program accommodations, and medical issues.
- One student had two identical ratings but different observation identification numbers associated with the ratings (54 observations were dropped, no students were dropped).
- STARS data varied for the same student—for example, a student had two sets of KOT ratings but different STARS demographic characteristics for each set of KOT ratings (79 observations were dropped, three students were dropped).
- Gender did not match between the KOT and STARS data (205 observations were dropped, eight students were dropped).
- Birthdates were far outside the range of kindergarten ages (younger than age 4 or older than age 8) or for which birthdates did not match between the KOT and STARS data (621 observations were dropped, 25 students were dropped).
- The STARS grade level was not kindergarten (51 observations were dropped, two students were dropped).

After deletion, the final analytic file included 132,971 observed items for 5,259 students.

Methodology

This section provides details on the psychometric analysis methods used to address the six research questions.

Research question 1 (What domains does the KOT measure?). To identify the latent constructs that the KOT measures, the study team conducted both exploratory factor analysis and confirmatory factor analysis. Because all the items are on a five-point ordinal scale, the study team conducted exploratory factor analysis on the basis of a polychoric correlation matrix and used full information maximum likelihood estimation with Huber-White covariance adjustment (specifically, the robust maximum likelihood estimation

option in Mplus) to account for missing data and nonnormality. The full sample of 5,259 students was divided into two random subsamples of equal size. The study team first conducted exploratory factor analysis with the first random subsample to determine the optimal factor structure and then conducted confirmatory factor analysis with the second random subsample to test the model fit. The confirmatory factor analysis results served as validation for the empirical latent constructs identified by the exploratory factor analysis (Thompson, 2004).

For the exploratory factor analysis, squared multiple correlations were used as the initial communality estimates. The scree test of eigenvalues was used to provide information on the number of potential underlying factors. Geomin, oblimin, and promax rotations were tried to obtain the optimal factor structure. The study team selected the final factor structure on the basis of five criteria: the hyperplane count (Yates, 1988), the number of nonsalient loadings below a threshold of 0.4 (Stevens, 2009), the number of double loaders (wherein a smaller number suggests better fit), the closeness to a simple structure (wherein the simpler the structure, the better; Fabrigar et al., 1999), and the meaningfulness of each factor (Gadermann et al., 2012).

After determining the optimal factor structure from the exploratory factor analysis with the first random subsample, the study team performed confirmatory factor analysis with the second random subsample. For the confirmatory factor analysis goodness-of-fit statistics such as comparative fit index and root mean square error of approximation were used to determine model fit. According to Kline (2005), a comparative fit index of more than .9 and a root mean square error of approximation of less than .08 (with an upper limit of less than .1) signal acceptable fit. A comparative fit index of more than .95 and a root mean square error of approximation of less than .05 suggest close fit (Hu & Bentler, 1999).

The study team then tested whether the same factor structure was observed among students in different subgroups. Multiple-group confirmatory factor analyses were conducted to determine whether factor invariance was achieved across subgroups. Configural invariance, metric invariance, and scalar invariance were examined in sequence. A stronger case of invariance is established when there is no sizable difference in the model fit statistics (Muthén & Asparouhov, 2002), indicated by either the nonsignificance of the chi-square difference test (Marsh & Grayson, 1994) or the minimal shift in comparative fit index and root mean square error of approximation (change in comparative fit index of less than .01 and change in root mean square error of approximation of less than .015; Cheung & Rensvold, 2002).

Finally, the study team examined reliability for retained factors by using two different approaches: an internal consistency approach in which Cronbach's alpha was calculated (see details under research question 2) and a Rasch modeling approach in which a person separation index was calculated. Reliability numbers higher than .9 indicate that a factor has high reliability, and numbers higher than .7 suggest that a factor has acceptable reliability (Kline, 2013). Factors with more items are more likely to have higher reliability. A person separation index higher than 2 indicates satisfactory model fit (Andrich, 1982). Rasch scores also were calculated for the identified factors. The Rasch score has two advantages over the simple average of items. First, the Rasch scores are on an interval scale, whereas the raw scores are on an ordinal scale. Second, the Rasch scores are less skewed (and therefore closer to a normal distribution) than the raw scores. The study team

examined the correlation between Rasch scores and average scores and computed reliabilities for each validated domain.

Research question 2 (To what extent do items in each domain measure the same underlying construct?). To examine the extent to which KOT items related to one another, the study team examined the internal consistency of the three KOT domains identified by factor analyses and further assessed whether internal consistency could be improved by removing any items from the domains.

The study team first examined the internal consistency for each domain obtained from the item-correlation matrix in the prior step. The study team also examined the item–total correlations for each item. Next, the study team ran internal consistency analyses to assess the unidimensionality of each domain as measured by the ordinal version of Cronbach's alpha (Zumbo, Gadermann, & Zeisser, 2007).

To assess whether the internal consistency of the latent constructs that the KOT measured would be improved by removing any items, the study team calculated the internal consistency (Cronbach's alpha) for each domain after removing each item and flagged items that would lead to a higher internal consistency when removed. The study team also used the Rasch model to generate fit statistics for each item, thus flagging the items that were not well aligned with the underlying domain (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008).

Research question 3 (To what extent does the KOT measure the same underlying constructs as other accepted measures of kindergarten readiness?). To examine the extent to which the KOT measures the same underlying constructs as other accepted measures of kindergarten readiness, the study team estimated Spearman rank order correlations between ratings for each of the 26 KOT items and DIBELS Next beginning of year composite scores and between scores for each of the three domains of school readiness identified by the factor analyses (general school readiness, cognitive school readiness, and noncognitive school readiness) and DIBELS Next scores. DIBELS Next is a set of procedures and short (one-minute) fluency measures for regularly monitoring the development of early literacy and early reading skills.[5] DIBELS Next includes five indicators: phonemic awareness, alphabetic principle, accuracy and fluency with connected text, reading comprehension, and vocabulary. DIBELS was administered to students during the first five days of the K–3 Plus program[6] and during the first 10 days of the regular school year kindergarten program. It is considered the students' beginning of year assessment. The DIBELS Next score represents the sum of scores on two measures: first sound fluency and letter naming fluency (University of Oregon Center on Teaching and Learning, 2012).

KOT item ratings (which are ordinal and range from 1 to 5), KOT domain scores, and DIBELS Next scores were transformed into rank-order measures.[7] A small percentage of students with DIBELS Next scores had missing scores on some KOT item ratings. Missing data were handled using maximum likelihood missing values estimation that permitted the retention of all students with both KOT and DIBELS Next scores.

Wald tests[8] were used to evaluate whether correlation coefficients from the structural equation model between each KOT domain score and the DIBELS Next score were statistically indistinguishable. Rejection of the null hypothesis (that the correlation coefficients

are equivalent) indicates that the terms being evaluated in the Wald test are statistically significantly different from zero.

To evaluate the convergent validity of each KOT and DIBELS Next score, the correlation coefficients were transformed into discrete, ordinal categories using the taxonomy provided by Di Iorio (2005). The thresholds and rankings for each measure of convergent validity are:
- Correlation coefficient ≥ .70: good convergence
- Correlation coefficient ≥ .40 and < .70: moderate convergence
- Correlation coefficient < .40: weak convergence

To determine the alignment between KOT and DIBELS benchmarks for identifying students as school ready, the KOT mean domain scores and the DIBELS Next scaled score were used to create binary benchmark indicators for each measure. The preliminary readiness standard for the KOT provided by New Mexico Public Education Department program staff and the guidelines provided by the technical documentation for DIBELS Next were used.[9] For the KOT a score of 2.0 indicates that the student is accomplished for a four-year-old and is making first steps for a kindergartner. For DIBELS Next a score of 26 was the benchmark threshold. Scores in the analytic sample ranged from 0 to 109. A score at or above 26 means that the student is making progress in reading and is likely to achieve subsequent reading benchmarks with appropriate and effective curriculum and instruction (Good & Kominski, 2011).

The study team then computed the classification accuracy of the benchmark for each KOT domain by summing the number of students with a score above both the KOT and DIBELS Next benchmark and the number of students with a score below both benchmarks. The study team also calculated the sensitivity (true positive rate) and the specificity (true negative rate) of the KOT benchmarks. The sensitivity measure assesses how well the benchmark for each KOT domain accurately identified students who were classified as not kindergarten ready (based on the DIBELS Next benchmark), and the specificity measure assesses how well the benchmark for each KOT domain accurately identified students who were classified as kindergarten ready (again, on the basis of the DIBELS Next benchmark). Put another way, the sensitivity measure conveys the likelihood that a student who is classified by DIBELS Next as not yet demonstrating foundational kindergarten entry knowledge and skills is classified the same way by the KOT. The specificity measure provides the proportion of students who are classified by DIBELS Next demonstrating foundational kindergarten entry knowledge and skills who are classified the same way by the KOT.

Table B1 provides a schematic of the components used to compute this measure. The overall classification accuracy was calculated using the formula $((A+D)/(A+B+C+D))*100$. That is, the sum of the number of students below the benchmark for both measures (cell A) and the number of students above the benchmark for both measures (cell D) was divided by the total number of students with nonmissing KOT domain scores and DIBELS scores. Sensitivity was calculated using the formula $(A/(A+C))*100$. Specificity was calculated using the formula $(D/(D+B))*100$.

Table B1. School readiness classification schematic for the New Mexico Kindergarten Observation Tool and Dynamic Indicators of Basic Early Literacy Skills Next, fall 2015

Kindergarten Observation Tool school readiness domain benchmark	Dynamic Indicators of Basic Early Literacy Skills Next benchmark	
	Below standard	At or above standard
Below standard	A	B
At or above standard	C	D

Source: Authors' creation.

Research question 4 (Do any of the items exhibit potential bias for student subgroups?). To examine whether KOT items functioned differently for different subgroups, differential item functioning analyses were conducted by gender, English learner status, eligibility for the federal school lunch program, special education status, and race/ethnicity. Differential step functioning analysis was then conducted for each item flagged by the differential item functioning analyses to determine at what step the largest differential effect occurred. Differential step functioning calculates whether the probability of attaining a rating on an item is statistically equivalent across subgroups after adjustment for kindergarten readiness as determined by all other items (Wright & Stone, 1979). The study team used log odds ratios to screen the differential step functioning, which did not require an adequate model fit assumption according to item-response-theory methods (Cohen, Kim, & Baker, 1993; Liu & Agresti, 1996; Penfield & Algina, 2003). The threshold used for these analyses was a log odds ratio of 0.64 (Zwick et al., 1999). Log odds ratios below 0.43 indicate negligible differential item functioning, log odds ratios of 0.43–0.64 indicate small to moderate differential item functioning, and log odds ratios above 0.64 indicate moderate to large differential item functioning. An absolute value of 0.64 or more for the log odds ratio was the criterion for potential bias in an item to avoid a high rate of type I error (Penfield, Alvarez, & Lee, 2009; Zieky, 1993). The analyses used complete cases only.

Research question 5 (Do teachers use the rating categories for each item as intended?). The study team examined the performance of each item from an item-response-theory perspective to understand how the rating categories were ordered on the estimated parameters —that is, whether behaviors described for each rating category represented a higher level skill or were more difficult to achieve than behaviors described for the rating categories below it. For each dimension the study team fit a partial credit polytomous Rasch model to the items. The Rasch model estimates the threshold parameters corresponding to the jump represented by two adjacent skill points (van der Linden & Hambleton, 1997) on the same scale as the latent trait. The study team used joint maximum likelihood estimation to account for missing data. Items would have been flagged as problematic if, for example, the threshold parameter for the jump from a rating of 1 to a rating of 2 were higher than the threshold parameter for the jump from a rating of 2 to a rating of 3 (Andrich, 2010; Bond & Fox, 2007).

Research question 6 (To what extent do ratings provide information about individual student abilities?). To examine the proportion of variation in scores on each KOT item that can be attributed to the classroom rather than to students, the school, or the district, the study team fit a four-level unconditional model (student, classroom, school, and district) for each indicator and for the three domains. The team then calculated the variance

partition coefficient (Goldstein, Browne, & Rasbash, 2002) or the proportion of the variation explained by the classroom level of the model.

The variance partition coefficient reflects both classroom-level clustering (similarity of students in the same classroom, relative to students in other classrooms) and rater effects (measurement error due to differences in ratings completed by different raters, such as differences in leniency or stringency of the rater). Because of this, the value of the classroom-level variance partition coefficient for each item is not a mere reflection of true classroom differences or interrater reliability in the data. The size of variance partition coefficient was examined across all items, and items for which classroom variation accounted for more than 50 percent of total variation were flagged for further investigation.

Appendix C. Supplemental tables and figures

This appendix includes supplemental tables and figures of study results.

Figure C1. Scree plot from exploratory factor analyses, New Mexico Kindergarten Observation Tool, fall 2015
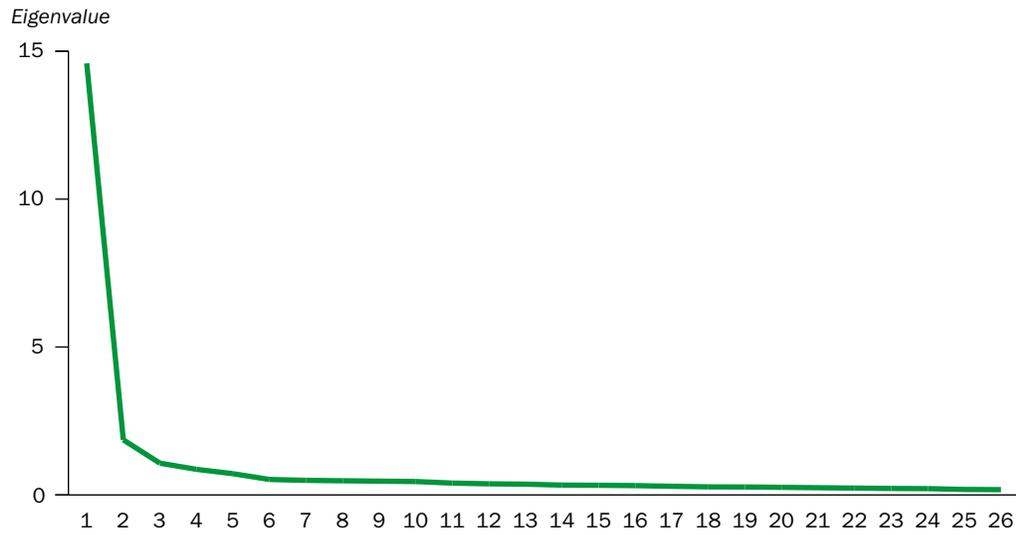
Eigenvalue

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C1. Factor loadings for each item in the New Mexico Kindergarten Observation Tool, by domain, fall 2015

Developer's intended domain and item	One factor structure	Two factor structure	
	General school readiness domain	Cognitive school readiness domain	Noncognitive school readiness domain
Physical development, health, and well-being			
Coordination and strength (1.1)	.52	—	—
Fine motor skills (2.1)	.67	—	—
Literacy			
Follows directions (5.2)	.78	—	.52
Rhyme (5.3a)	.74	.68	—
Letter-sounds, beginning sound (5.3b)	.79	.82	—
Vocabulary (5.4)	.85	.73	—
Conversational ability (6.1)	.79	.66	—
Book enjoyment (7.1)	.80	.67	—
Story comprehension (7.2)	.81	.72	—
Book conventions (7.3)	.76	.69	—
Alphabet knowledge and word/letter recognition (7.4)	.77	.84	—
Writing (8.3)	.77	.66	—
Numeracy			
One-to-one correspondence, number relationships (9.1)	.78	.81	—
Number words (9.3a)	.73	.88	—
Numerals (9.3b)	.72	.78	—
Measurement (11.3)	.72	.71	—
Sorting (12.1)	.75	.64	—
Scientific conceptual understanding			
Investigations (14.1)	.69	.55	—
Earth science (16.1)	.69	.55	—
Self, family, and community			
Self-control (18.1)	.76	—	.86
Cares for possessions (19.1)	.75	—	.83
Plays and interacts (20.1)	.77	—	.74
Social problem solving (20.2)	.73	—	.77
Guidance and support (21.2)	.71	—	.80
Approaches to learning			
Independence (24.2)	.80	—	.78
Focus (27.1)	.76	—	.76

— indicates that the item did not load saliently on the domain.

Note: Numbers in parentheses are the Kindergarten Observation Tool developer's original item numbers. $N = 5,259$ students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C2. Summary of factor invariance results for the New Mexico Kindergarten Observation Tool, fall 2015

Grouping variable	Configural invariance established	Chi square test non significant	Shift in comparative fit index ≤ .01	Shift in root mean square error of approximation ≤ .015	Strong factor invariance achieved
Gender	Yes	No	Yes	Yes	Yes
English learner status	Yes	No	Yes	Yes	Yes
Eligibility for the federal school lunch program	Yes	No	Yes	Yes	Yes
Special education status	Yes	No	Yes	Yes	Yes
Race/ethnicity	Yes	No	Yes	Yes	Yes

Note: N = 5,259 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C3. Item-level internal consistency analysis and Rasch analysis summary for the New Mexico Kindergarten Observation Tool, fall 2015

Developer's intended domain and item	One factor structure	Two factor structure	
	General school readiness domain	Cognitive school readiness domain	Noncognitive school readiness domain
Coordination and strength (1.1)	a	na	na
Fine motor skills (2.1)	√	na	na
Follows directions (5.2)	√	na	√
Rhyme (5.3a)	√	√	na
Letter-sounds, beginning sound (5.3b)	√	√	na
Vocabulary (5.4)	√	√	na
Conversational ability (6.1)	√	√	na
Book enjoyment (7.1)	√	√	na
Story comprehension (7.2)	√	√	na
Book conventions (7.3)	√	√	na
Alphabet knowledge and word/letter recognition (7.4)	√	√	na
Writing (8.3)	√	√	na
One-to-one correspondence, number relationships (9.1)	√	√	na
Number words (9.3a)	√	√	na
Numerals (9.3b)	√	√	na
Measurement (11.3)	√	√	na
Sorting (12.1)	√	√	na
Investigations (14.1)	√	√	na
Earth science (16.1)	√	√	na
Self-control (18.1)	√	na	√
Cares for possessions (19.1)	√	na	√
Plays and interacts (20.1)	√	na	√
Social problem solving (20.2)	√	na	√
Guidance and support (21.2)	√	na	√
Independence (24.2)	√	na	√
Focus (27.1)	√	na	√

√ indicates that the item has a salient factor leading and performed well under both analysis approaches. na indicates that the item was not part of the domain.

Note: Numbers in parentheses are the Kindergarten Observation Tool developer's original item numbers. Pairwise deletion was used when calculating correlations between any pair of items; however, all students contributed to the calculation of alpha. For Rasch analyses, Winsteps uses joint maximum likelihood estimation, which does not require complete data from each student to produce estimates. $N = 5{,}259$ students.

a. Item was identified as problematic under the Cronbach's alpha analysis and under Rasch analysis.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C4. Spearman rank-ordered correlation coefficients between New Mexico Kindergarten Observation Tool items and Dynamic Indicators of Basic Early Literacy Skills Next beginning of year composite scores, fall 2015

One-factor structure	Two-factor structure	Item	Correlation coefficient[a]	Convergent validity[b]
General school readiness domain	Cognitive school readiness domain	Rhyme (5.3a)	.480	Moderate
		Letter-sounds, beginning sound (5.3b)	.620	Moderate
		Vocabulary (5.4)	.508	Moderate
		Conversational ability (6.1)	.458	Moderate
		Book enjoyment (7.1)	.449	Moderate
		Story comprehension (7.2)	.483	Moderate
		Book conventions (7.3)	.488	Moderate
		Alphabet knowledge and word/letter recognition (7.4)	.689	Moderate
		Writing (8.3)	.520	Moderate
		One-to-one correspondence, number relationships (9.1)	.541	Moderate
		Number words (9.3a)	.539	Moderate
		Numerals (9.3b)	.597	Moderate
		Measurement (11.3)	.427	Moderate
		Sorting (12.1)	.402	Moderate
		Investigations (14.1)	.370	Weak
		Earth science (16.1)	.365	Weak
	Noncognitive school readiness domain	Follows directions (5.2)	.446	Moderate
		Self-control (18.1)	.346	Weak
		Cares for possessions (19.1)	.280	Weak
		Plays and interacts (20.1)	.324	Weak
		Social problem solving (20.2)	.336	Weak
		Guidance and support (21.2)	.344	Weak
		Independence (24.2)	.416	Moderate
		Focus (27.1)	.401	Moderate
	Not included	Fine motor skills (2.1)	.400	Moderate
Not included	Not included	Coordination and strength (1.1)	.222	Weak

Note: Numbers in parentheses are the Kindergarten Observation Tool developer's original item numbers. All correlations were statistically significant at the $p < .01$ level after significance level was adjusted using the Bonferroni multiple comparisons correction. $N = 3,257$ students.

a. Represents Spearman rank-order coefficients between each Kindergarten Observation Tool item and the Dynamic Indicators of Basic Early Literacy Skills Next beginning of year composite score.

b. Provides the convergent and validity strength of each Kindergarten Observation Tool and Dynamic Indicators of Basic Early Literacy Skills Next beginning of year composite score association measure.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and data on the 2015 Dynamic Indicators of Basic Early Literacy Skills Next from the New Mexico Public Education Department.

Table C5. Percentage of students for whom student classification for school readiness on the basis of the New Mexico Kindergarten Observation Tool benchmark and on the basis of the Dynamic Indicators of Basic Early Literacy Skills Next benchmark concur, by student subgroup, fall 2015

Student subgroup	One-factor structure	Two-factor structure	
	General school readiness domain	Cognitive school readiness domain	Noncognitive school readiness domain
All students	69.5	73.4	61.5
Gender			
Female	68.2	72.9	59.2
Male	70.8	73.8	63.7
English learner status			
English learner student	73.9	78.9	60.9
Non–English learner student	68.7	72.3	61.6
Eligibility for the federal school lunch program			
Eligible	69.2	73.4	59.5
Not eligible	70.2	73.3	64.9
Special education status			
In special education	74.4	76.0	60.9
Not in special education	69.0	73.1	66.8
Race/ethnicity			
Hispanic	66.0	71.1	57.4
White	72.4	73.2	68.2
American Indian/Alaska Native	71.9	77.1	61.1

Note: *N* = 3,257 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and data on the 2015 Dynamic Indicators of Basic Early Literacy Skills Next from the New Mexico Public Education Department.

Table C6. Sensitivity: Percentage of students below both the Kindergarten Observation Tool and the Dynamic Indicators of Basic Early Literacy Skills Next benchmarks for school readiness, by student subgroup, fall 2015

Student subgroup	One factor structure	Two factor structure	
	General school readiness domain	Cognitive school readiness domain	Noncognitive school readiness domain
All students	59.7	67.4	46.6
Gender			
Female	55.8	66.2	38.5
Male	63.1	68.5	53.8
English learner status			
English learner student	55.9	63.8	43.8
Non–English learner student	73.8	81.2	57.0
Eligibility for the federal school lunch program			
Eligible	48.1	55.3	39.4
Not eligible	64.5	72.5	49.6
Special education status			
In special education	57.4	65.6	43.7
Not in special education	78.1	82.2	69.7
Race/ethnicity			
Hispanic	58.5	66.7	44.5
White	48.3	52.9	41.0
American Indian/Alaska Native	68.1	77.5	52.7

Note: *N* = 3,257 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and data on the 2015 Dynamic Indicators of Basic Early Literacy Skills Next from the New Mexico Public Education Department.

Table C7. Specificity: Percentage of students above both the Kindergarten Observation Tool and Dynamic Indicators of Basic Early Literacy Skills Next benchmarks for school readiness, by student subgroup, fall 2015

Student subgroup	One-factor structure — General school readiness domain	Two-factor structure — Cognitive school readiness domain	Two-factor structure — Noncognitive school readiness domain
All students	83.2	82.0	84.0
Gender			
Female	87.0	81.9	85.0
Male	79.1	82.2	82.9
English learner status			
English learner student	84.1	83.2	84.9
Non–English learner student	74.2	71.1	74.4
Eligibility for the federal school lunch program			
Eligible	87.9	89.5	90.1
Not eligible	78.9	75.3	78.4
Special education status			
In special education	84.7	83.5	85.2
Not in special education	59.3	59.8	64.6
Race/ethnicity			
Hispanic	79.1	78.5	78.8
White	90.0	89.5	91.7
American Indian/Alaska Native	81.8	76.1	81.4

Note: *N* = 3,257 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test and data on the 2015 Dynamic Indicators of Basic Early Literacy Skills Next from the New Mexico Public Education Department.

Table C8. Differential item functioning summary for the New Mexico Kindergarten Observation Tool general school readiness domain, fall 2015

Item	Subgroup contrast (log odds estimates)						
	Female students	English learner students	Students eligible for the federal school lunch program	Students in special education	Hispanic students (White students as reference)	American Indian students (Hispanic students as reference)	American Indian students (White students as reference)
Fine motor skills (2.1)	–0.44	–0.42	–0.34	–0.08	0.16	–0.68[†]	–0.82[†]
Follows directions (5.2)	–0.26	–0.14	–0.27	0.27	0.31	0.14	–0.15
Rhyme (5.3a)	0.11	0.31	0.13	0.03	–0.17	0.09	0.26
Letter-sounds, beginning sound (5.3b)	0.20	–0.13	–0.08	–0.13	–0.10	–0.25	–0.15
Vocabulary (5.4)	0.33	0.46	0.34	0.10	–0.13	0.04	0.14
Conversational ability (6.1)	0.09	0.20	0.13	0.57	–0.05	0.49	0.57
Book enjoyment (7.1)	–0.02	–0.04	–0.08	0.06	0.16	0.14	0.04
Story comprehension (7.2)	0.18	0.31	0.02	0.42	–0.09	0.28	0.35
Book conventions (7.3)	0.15	0.12	0.12	0.03	–0.18	–0.24	0.06
Alphabet knowledge and word/letter recognition (7.4)	0.24	0.34	0.22	–0.47	–0.35	–0.35	0.05
Writing (8.3)	–0.09	0.00	0.25	–0.30	–0.18	–0.20	–0.07
One-to-one correspondence, number relationships (9.1)	0.44	0.13	–0.01	–0.25	0.05	0.13	0.06
Number words (9.3a)	0.40	0.49	0.44	–0.02	–0.31	0.20	0.49
Numerals (9.3b)	0.51	0.21	0.07	–0.34	–0.21	–0.14	0.06
Measurement (11.3)	0.32	0.22	0.39	–0.03	–0.04	0.34	0.40
Sorting (12.1)	0.11	–0.04	0.06	–0.11	–0.07	0.15	0.22
Investigations (14.1)	0.24	0.09	0.41	0.00	–0.13	0.19	0.37
Earth science (16.1)	0.24	0.10	0.34	0.11	–0.06	0.03	0.16
Self-control (18.1)	–0.59	–0.38	–0.35	0.00	0.24	–0.14	–0.46
Cares for possessions (19.1)	–0.44	–0.58	–0.39	–0.01	0.34	–0.23	–0.67[†]
Plays and interacts (20.1)	–0.18	–0.07	–0.33	–0.03	0.30	0.07	–0.24
Social problem solving (20.2)	–0.17	0.03	0.18	0.03	0.08	0.45	0.39
Guidance and support (21.2)	–0.29	–0.22	–0.41	0.04	0.03	–0.25	–0.35
Independence (24.2)	–0.21	–0.16	–0.36	0.02	0.20	–0.14	–0.33
Focus (27.1)	–0.57	–0.36	–0.28	0.18	0.13	0.03	–0.19

† Log odds ratios larger than 0.64; the associated items were flagged for differential item functioning.

Note: Numbers in parentheses are Kindergarten Observation Tool item numbers. Each column represents a set of differential item functioning analyses. The column header represents the focal group, and the rest of the students in the sample form the reference group. Negative numbers indicate potential bias favoring the focal group, and positive numbers indicate potential bias favoring the reference group. $N = 5,259$ students for all analyses except the comparison of Hispanic students with White students ($N = 4,017$ students), the comparison of American Indian students with Hispanic students ($N = 3,613$ students), and the comparison of American Indian students with White students ($N = 2,199$ students).

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C9. Differential item functioning summary for the New Mexico Kindergarten Observation Tool for the cognitive school readiness domain, fall 2015

Item	Subgroup contrast (log odds estimates)						
	Female students	English learner students	Students eligible for the federal school lunch program	Students in special education	Hispanic students (White students as reference)	American Indian students (Hispanic students as reference)	American Indian students (White students as reference)
Rhyme (5.3a)	–0.09	0.19	0.02	0.02	–0.11	0.08	0.14
Letter-sounds, beginning sound (5.3b)	–0.04	–0.35	–0.28	–0.11	0.07	–0.27	–0.41
Vocabulary (5.4)	0.03	0.28	0.12	0.13	–0.01	–0.02	–0.05
Conversational ability (6.1)	–0.16	0.04	–0.04	0.65[†]	0.06	0.46	0.36
Book enjoyment (7.1)	–0.27	–0.23	–0.27	0.11	0.28	0.06	–0.21
Story comprehension (7.2)	–0.08	0.11	–0.19	0.47	0.02	0.16	0.20
Book conventions (7.3)	–0.09	–0.04	–0.07	0.04	–0.08	–0.35	–0.19
Alphabet knowledge and word/letter recognition (7.4)	0.05	0.16	0.03	–0.53	–0.25	–0.47	–0.20
Writing (8.3)	–0.34	–0.11	0.10	–0.27	–0.10	–0.28	–0.18
One-to-one correspondence, number relationships (9.1)	0.23	–0.06	–0.22	–0.21	0.21	0.11	–0.10
Number words (9.3a)	0.25	0.34	0.30	–0.03	–0.20	0.10	0.29
Numerals (9.3b)	0.36	0.04	–0.11	–0.34	–0.07	–0.21	–0.19
Measurement (11.3)	0.12	0.04	0.22	–0.04	0.07	0.27	0.23
Sorting (12.1)	–0.09	–0.17	–0.09	–0.05	0.07	0.18	0.07
Investigations (14.1)	0.06	–0.05	0.26	0.04	–0.01	0.15	0.20
Earth science (16.1)	0.08	–0.03	0.21	0.17	0.06	0.00	0.01

† Log odds ratios larger than 0.64; the associated items were flagged for differential item functioning.

Note: Numbers in parentheses are Kindergarten Observation Tool item numbers. Each column represents a set of differential item functioning analyses. The column header represents the focal group, and the rest of the students in the sample form the reference group. Negative numbers indicate potential bias favoring the focal group, and positive numbers indicate potential bias favoring the reference group. $N = 5,259$ students for all analyses except the comparison of Hispanic students with White students ($N = 4,017$ students), the comparison of American Indian students with Hispanic students ($N = 3,613$ students), and the comparison of American Indian students with White students ($N = 2,199$ students).

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C10. Differential item functioning summary for the New Mexico Kindergarten Observation Tool noncognitive school readiness domain, fall 2015

Item	Subgroup contrast (log odds estimates)						
	Female students	English learner students	Students eligible for the federal school lunch program	Students in special education	Hispanic students (White students as reference)	American Indian students (Hispanic students as reference)	American Indian students (White students as reference)
Follows directions (5.2)	0.06	0.19	0.12	0.37	−0.01	0.20	0.21
Self-control (18.1)	−0.30	−0.23	−0.12	−0.18	0.11	−0.20	−0.33
Cares for possessions (19.1)	−0.13	−0.49	−0.18	−0.17	0.24	−0.29	−0.56
Plays and interacts (20.1)	0.18	0.19	−0.01	−0.05	0.08	0.12	0.09
Social problem solving (20.2)	0.17	0.29	0.49	−0.04	−0.10	0.54	0.68[†]
Guidance and support (21.2)	0.06	−0.02	−0.21	−0.13	−0.17	−0.29	−0.14
Independence (24.2)	0.21	0.16	−0.04	−0.02	−0.06	−0.09	−0.04
Focus (27.1)	−0.28	−0.13	−0.05	0.18	−0.05	0.04	0.08

† Log odds ratios larger than 0.64; the associated items were flagged for differential item functioning.

Note: Numbers in parentheses are Kindergarten Observation Tool item numbers. Each column represents a set of differential item functioning analyses. The column header represents the focal group, and the rest of the students in the sample form the reference group. Negative numbers indicate potential bias favoring the focal group, and positive numbers indicate potential bias favoring the reference group. N = 5,259 students for all analyses except the comparison of Hispanic students with White students (N = 4,017 students), the comparison of American Indian students with Hispanic students (N = 3,613 students), and the comparison of American Indian students with White students (N = 2,199 students).

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C11. Rasch analysis results for the New Mexico Kindergarten Observation Tool, fall 2015

Item	One-factor structure	Two-factor structure	
	General school readiness domain	Cognitive school readiness domain	Noncognitive school readiness domain
Coordination and strength (1.1)	—	—	—
Fine motor skills (2.1)	–3.4, –1.6, 1.3, 3.8	—	—
Follows directions (5.2)	–3.1, –1.2, 1.1, 3.2	—	–4.3, –1.6, 1.4, 4.5
Rhyme (5.3a)	–2.4, –1.3, 0.6, 3.1	–2.8, –1.5, 0.7, 3.5	—
Letter-sounds, beginning sound (5.3b)	–2.9, –1.2, 0.9, 3.2	–3.4, –1.3, 1.1, 3.7	—
Vocabulary (5.4)	–3.1, –1.3, 1.0, 3.4	–3.6, –1.4, 1.1, 3.9	—
Conversational ability (6.1)	–2.6, –1.4, 0.9, 3.1	–3.0, –1.5, 1.0, 3.5	—
Book enjoyment (7.1)	–3.3, –1.1, 0.9, 3.5	–3.8, –1.3, 1.1, 4.1	—
Story comprehension (7.2)	–3.2, –1.4, 0.8, 3.6	–3.7, –1.5, 0.9, 4.3	—
Book conventions (7.3)	–3.2, –1.3, 1.4, 3.1	–3.7, –1.5, 1.5, 3.6	—
Alphabet knowledge and word/letter recognition (7.4)	–2.2, –1.6, 1.0, 2.8	–2.6, –1.8, 1.2, 3.2	—
Writing (8.3)	–3.1, –1.0, 1.2, 2.9	–3.6, –1.2, 1.3, 3.4	—
One-to-one correspondence, number relationships (9.1)	–2.7, –1.2, 1.2, 2.7	–3.1, –1.3, 1.3, 3.1	—
Number words (9.3a)	–2.7, –0.7, 0.7, 2.7	–3.1, –0.9, 0.8, 3.2	—
Numerals (9.3b)	–3.4, –1.8, 1.0, 4.2	–3.9, –2.0, 1.1, 4.8	—
Measurement (11.3)	–2.5, –1.6, 0.8, 3.4	–3.0, –1.8, 0.9, 3.9	—
Sorting (12.1)	–3.0, –1.6, 0.9, 3.7	–3.5, –1.7, 1.0, 4.2	—
Investigations (14.1)	–4.3, –1.9, 1.4, 4.8	–4.9, –2.1, 1.5, 5.6	—
Earth science (16.1)	–3.6, –1.1, 1.8, 2.9	–4.2, –1.3, 1.9, 3.6	—
Self-control (18.1)	–3.0, –0.9, 0.9, 2.9	—	–4.1, –1.3, 1.2, 4.1
Cares for possessions (19.1)	–3.3, –1.3, 1.0, 3.7	—	–4.5, –1.8, 1.3, 5.1
Plays and interacts (20.1)	–3.3, –1.3, 1.2, 3.4	—	–4.5, –1.7, 1.5, 4.7
Social problem solving (20.2)	–3.1, –0.9, 1.2, 2.7	—	–4.3, –1.4, 1.5, 4.1
Guidance and support (21.2)	–3.0, –1.5, 1.0, 3.5	—	–4.1, –1.9, 1.2, 4.8
Independence (24.2)	–3.3, –1.3, 1.1, 3.5	—	–4.5, –1.8, 1.4, 4.9
Focus (27.1)	–1.9, –0.5, 2.4, NR	—	–2.6, –0.6, 3.2, NR

— is not included in that domain. NR is no students rated at that level.

Note: Numbers in parentheses are Kindergarten Observation Tool item numbers. Values in the table are threshold parameter estimates obtained by Rasch analysis at different rating categories—from left to right and separated by commas—the jump at rating 2, at rating 3, at rating 4, and at rating 5. The numbers in each cell always ascend from left to right, indicating there was no reversed ordering of the rating categories. $N = 5{,}259$ students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Table C12. Item-level classroom-level variation for the New Mexico Kindergarten Observation Tool, fall 2015

Item	Percent of variation at classroom level
Coordination and strength (1.1)	42
Fine motor skills (2.1)	26
Follows directions (5.2)	32
Rhyme (5.3a)	36
Letter-sounds, beginning sound (5.3b)	30
Vocabulary (5.4)	38
Conversational ability (6.1)	36
Book enjoyment (7.1)	34
Story comprehension (7.2)	38
Book conventions (7.3)	42
Alphabet knowledge and word/letter recognition (7.4)	25
Writing (8.3)	37
One-to-one correspondence, number relationships (9.1)	37
Number words (9.3a)	35
Numerals (9.3b)	30
Measurement (11.3)	61
Sorting (12.1)	50
Investigations (14.1)	51
Earth science (16.1)	58
Self-control (18.1)	29
Cares for possessions (19.1)	40
Plays and interacts (20.1)	41
Social problem solving (20.2)	45
Guidance and support (21.2)	36
Independence (24.2)	33
Focus (27.1)	29

Note: N = 5,259 students.

Source: Authors' analysis of data from the 2015 Kindergarten Observation Tool field test.

Notes

1. Alliance member organizations are the Children's World Child Development Center; the New Mexico Association for the Education of Young Children; the New Mexico Early Childhood Development Partnership; the New Mexico Public Education Department; New Mexico Voices for Children; the State of New Mexico Children, Youth and Families Department; STG International Inc.; the University of New Mexico CE Early Childhood Services Center; the Thornburg Foundation; and the United Way of Santa Fe County.

2. Correlations of scores on items in the literacy domain of the developer's intended structure for the Kindergarten Observation Tool with DIBELS Next score were not consistently stronger than correlations between scores on items in the numeracy domain of the developer's intended structure and DIBELS Next score. The correlation coefficient ranged from .48 to .69 between scores on items in the literacy domain of the developer's intended structure and DIBELS Next score and from .40 to .60 between scores on items in the numeracy domain and DIBELS Next score.

3. This benchmark was established by the New Mexico Public Education Department prior to the fall 2015 administration of the Kindergarten Observation Tool.

4. Because the DIBELS Next data provided by the New Mexico Public Education Department did not include a teacher or classroom ID associated with the DIBELS Next administration, the study team used the staff ID connected to students' Kindergarten Observation Tool administration to link students to classrooms.

5. Previous studies have explored relationships between the DIBELS Next and other measures. Rouse and Fantuzzo (2006) found evidence of concurrent and predictive validity between kindergarten students' scores on three DIBELS subscales and reading assessments, teacher-reported measures of students' reading ability, and nationally administered standardized assessments. DIBELS Next has also shown moderate to moderately high concurrent validity with other measures, including the Woodcock-Johnson psycho-educational battery, the Test of Early Reading Ability Reading Quotient, and the Terra Nova Reading subtest (Fien et al., 2008). Studies have explored other psychometric properties of DIBELS Next. Carlson et al. (2010) reported acceptable reliability coefficients across DIBELS Next subtests, and support for the content validity of DIBELS Next has been provided by reading experts and demonstrated in reviews and meta-analyses of the research literature (for example, Ehri, Nunes, Stahl, & Willows, 2001; Riedel, 2007).

6. The K–3 Plus program funds additional education time for students in grades K–3 in public schools in which 80 percent or more of students are eligible for the federal school lunch program and in public schools that received a D or F grade in the previous year.

7. Stata's egen rank function was used to perform the transformation. Tied observations (for example, observations with the same Kindergarten Observation Tool score) are assigned the average rank among the ties.

8. This Wald test procedure was performed using the postestimation commands (estat stdize: test) available after fitting the structural equation model for evaluating the equality of standardized coefficients.

9. All analyses in this section used pairwise deletion to remove cases with missing Kindergarten Observation Tool domain scores.

References

Aikens, N., Kopack Klein, A., Tarullo, L., & West, J. (2013). *Getting ready for kindergarten: Children's progress during Head Start (FACES 2009 report)*. OPRE report No. 2013-21a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika, 47*(1), 105–113.

Andrich, D. (2010). Understanding the response structure and process in the polytomous Rasch model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 123–152). New York, NY: Routledge.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences,* 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

Carlson, J. F., Römhild, A., McCormick, C., Chin, K., Geisinger, K. F., Shaw, L., & Foley, B. P. (2010). *Evaluating reading tests for the state of Florida: Dynamic Indicators of Basic Early Literacy Skills, 6th edition (DIBELS) and Florida Assessments for Instruction in Reading (FAIR)*. Lincoln, NE: Buros Center for Testing.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. http://eric.ed.gov/?id=EJ651463

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach,* 2nd ed. New York, NY: Routledge.

Cohen, A. S., Kim, S.-H., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17*(4), 335–350. http://eric.ed.gov/?id=EJ485766

Di Iorio, C. K. (2005). *Measurement in health behavior: Methods for research and evaluation*. San Francisco, CA: Jossey-Bass.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*(1), 35–54. https://eric.ed.gov/?id=EJ795842

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., et al. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 25*(2), 177–193. https://eric.ed.gov/?id=EJ874840

Ehri, L., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393–447.

Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice, 29*(2), 24–35. http://eric.ed.gov/?id=EJ886732

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Fien, H., Baker, S. K., Smolkowski, K., Smith, J. L. M., Kame'enui, E. J., & Beck, C. T. (2008). Using nonsense word fluency to predict reading proficiency in kindergarten through second grade for English learners and native speakers. *School Psychology Review 37*(3), 391–408. http://eric.ed.gov/?id=EJ817301

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1–13. http://eric.ed.gov/?id=EJ977577

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics, 1*(4), 223–231.

Good, R. H., & Kominski, R. A. (2011). *Dynamic Indicators of Basic Literacy Skills: DIBELS Next assessment manual.* Eugene, OR: Dynamic Measurement Group.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources, 39*(2), 326–354. http://eric.ed.gov/?id=EJ746489

Hoyt, W. T. & Kerns, M. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*(4), 403–424.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. http://eric.ed.gov/?id=EJ576570

Janus, M., & Duku, E. (2007). The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn. *Early Education and Development, 18*(3), 375–403. http://eric.ed.gov/?id=EJ780658

Johnson, S. M., Kardos, S. M., Kauffman, D., Liu, E., & Donaldson, M. L. (2004). The support gap: New teachers' early experiences in high-income and low-income schools. *Education Policy Analysis Archives, 12*(61), 1–25. https://eric.ed.gov/?id=EJ853526

Kline, R. B. (2005). *Principles and practice of structural equation modeling,* 2nd ed. New York, NY: Guilford Press.

Kline, R. B. (2013). *Handbook of psychological testing,* 2nd ed. New York, NY: Routledge.

Larson, K., Russ, S. A., Nelson, B. B., Olson, L. M., & Halfon, N. (2015). Cognitive ability at kindergarten entry and socioeconomic status. *Pediatrics 135*(2), e440–e448.

Liu, I.-M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics, 52*(4), 1223–1234.

Marsh, H. W., & Grayson, D. (1994). Longitudinal confirmatory factor analysis: Common, time-specific, item-specific, and residual-error components of variance. *Structural Equation Modeling, 1*(2), 116–145.

Mashburn, A. J., & Henry, G. T. (2004). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice, 23*(4), 16–30.

Muthén, B., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus.* Los Angeles, CA: University of California, Los Angeles. Retrieved January 27, 2016, from http://www.statmodel.com/download/webnotes/CatMGLong.pdf.

Penfield, R. D., & Algina, J. (2003). Applying the Liu-Agresti estimator of the cumulative common odds ratio to DIF detection in polytomous items. *Journal of Educational Measurement, 40*(4), 353–370.

Penfield, R. D., Alvarez, K., & Lee, O. (2009). Using a taxonomy of differential step functioning form to improve the interpretation of DIF in polytomous items: An illustration. *Applied Measurement in Education, 22*(1), 61–78. http://eric.ed.gov/?id=EJ822892

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open, 2*(3), 1–18.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*(4), 546–567. http://eric.ed.gov/?id=EJ776733

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the dynamic indicators for basic early literacy skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*(3), 341–355. http://eric.ed.gov/?id=EJ788273

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academies Press. http://eric.ed.gov/?id=ED446866

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*(1), 33–44.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences,* 5th ed. New York, NY: Routledge.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

University of Oregon Center on Teaching and Learning. (2012). *DIBELS Next recommended benchmark goals*. Eugene, OR: University of Oregon.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York, NY: Springer.

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly, 27*(1), 46–54. http://eric.ed.gov/?id=EJ947506

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press. http://eric.ed.gov/?id=ED436552

Yates, A. (1988). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY: State University of New York Press.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Lawrence Erlbaum.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21–29.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1–28. https://eric.ed.gov/?id=EJ590611

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research