CEMETR-2013-05
DECEMBER 2013

CEME Technical Report

The Center for Educational Measurement and Evaluation

Technical Manual (2nd Edition) for the *Teaching Strategies GOLD*™ Assessment System

Richard Lambert
Do-Hong Kim
Diane Burts

RICHARD LAMBERT
CHUANG WANG
MARK D'AMICO
SERIES EDITORS

A PUBLICATION OF THE CENTER FOR EDUCATIONAL MEASUREMENT AND EVALUATION

CEME
The Center for Educational Measurement and Evaluation
College of Education
UNC CHARLOTTE

Technical Manual for the *Teaching Strategies GOLD™* Assessment System (2nd edition)

Richard G. Lambert
Do-Hong Kim
Center for Educational Measurement and Evaluation
UNC Charlotte

Diane C. Burts
Louisiana State University

November 2013

The measurement properties of any new assessment instrument should be rigorously examined and the results made available to stakeholders. The first version of the technical manual for the *Teaching Strategies GOLD™* Assessment System (Lambert, Kim, Taylor, & McGee, 2010) presented initial reporting of reliability and validity evidence based on the information the measure provides to teachers of young children. The manual contained evidence concerning the dimensions measured by the assessment system and their interrelationships. The results outlined the measurement model used to create scale scores for each dimension. The report also contained a variety of strong statistical evidences concerning the fit of the data provided by the assessment system to the measurement model. Strong reliability evidence was presented from both classical and modern indexes of internal consistency, along with the results of a study of inter-rater reliability. Norm tables for each scale score were provided based on three month age bands spanning ages 6 to 71 months.

Four studies were conducted using the two initial national norm samples (n_1=10,963; n_2=33,612). These samples were diverse and included children of different backgrounds, races, ethnicities, and special needs who were enrolled in varied educational programs across the United States. Sample sizes varied according to study research

questions. A total of 4,580 teachers who were new *GOLD*® users gathered the data on children in their classrooms. Following is a summary of the four studies.

Study1 explored the (a) factorial structure of the *GOLD*®, (b) indexes of reliability, and (c) inter-rater reliability. Findings suggested that the *GOLD*® measures six separate domains as intended. Inter-rater reliability between a master trainer and teachers was high. Reliability coefficients for all three checkpoints were also high. Results of longitudinal invariance CFA indicated the constructs were equivalent across time implying that the interpretations of changes in children's development and learning obtained from the measure are valid (Lambert, Kim, & Burts, 2012).

Study 2 examined the validity of the *GOLD*® for use with English language learners (ELLs) and for children with disabilities. Data from three-, four-, and five -year-olds with complete item responses from the fall, winter, and spring were analyzed according to each child's primary language or disability status. The majority of items in the *GOLD*® displayed little or no Differential Item Functioning (DIF) with the exception of one item, "uses conventional grammar" (Kim, Lambert, & Burts, 2013).

Study 3 investigated child and classroom composition characteristics associated with teacher ratings and the variability between raters when controlling for these characteristics. Three-level growth curve modeling indicated that teacher ratings were associated in expected directions. Children with disabilities began the year behind their typically developing peers and grew more slowly throughout the year. Girls demonstrated advantages in some areas over boys. ELLs were rated lower at the beginning of the year but exhibited somewhat faster growth rates than native English-speakers. Differences in rater effects (i.e., how teachers used the *GOLD*® to rate the children in their classrooms) ranged

from 16% to 25%, which is considerably lower than reported in some studies (Lambert, Kim, & Burts, 2013).

Study 4 was designed to develop interval level scale scores that could be used to track children's development and learning across the entire age range of the *GOLD*®. The study examined: (a) dimensionality, (b) rating scale effectiveness, (c) hierarchy of item difficulties, and (d) the relationship of scale scores to child age. Results indicated that each subscale measures only one underlying latent construct. The rating structure functioned effectively with the exceptions that ratings at the lowest and highest ends of the scale were somewhat less reliable and in-between ratings were less distinct. Overall, items formed theoretically expected hierarchies such that items which were less difficult for children were rated by teachers as less difficult. Correlations of developmental scale scores with child age were moderate (Kim, Lambert, & Burts, in press).

At the time the initial manual was produced, the assessment system was relatively new and many of the teachers had been using the system for only one year. Since the last report, many more states and programs have adopted the assessment system, much more training has taken place, and more research has been conducted to support the use of the system.

Since *Teaching Strategies GOLD*® was released in the Fall of 2010, the number of teachers using the tool has grown to more than 45,000, with over a million children with portfolios. All teachers have access to training through the online courses, as well as Interrater reliability. In addition to the free training, thousands of teachers are trained each year, using face-to-face training, to ensure their knowledge of how to use the tool. Teaching Strategies GOLD is widely used in all states for Pre-k assessment. Additionally, Teaching

Strategies has 22 state-level agreements for Pre-k assessment and 12 state-level agreements for Kindergarten assessment."

Given the widespread use of the *GOLD*®, greater availability of teacher training, and much more sophisticated and experienced use of the system, the purpose of this manual is to provide an updated set of evidences based on a current nationally representative norm sample that reflects how the *GOLD*® is currently used. To meet this purpose, this manual provides updated reliability and validity evidence based on both classical and Item Response Theory based measurement models. A new set of norm tables is also provided that cover three month age bands for children ages birth through 71 months. For each age band, expected scores for the fall, winter, and spring assessments, age specific standard errors of measurement, and expected growth from fall to spring are provided.

Norm Sample

A total population of 934,073 children had skills rated using the *Teaching Strategies GOLD™* assessment system for the 2012-2013 academic year. These children received educational services from centers and school-based sites across all 50 states, the District of Columbia, and Puerto Rico. This total population of children was divided into the following six age or grade strata according to information provided by the teacher: birth to 11 months of age, 12 to 23 months of age, 24 to 35 months of age, 3 year old pre-kindergarten, 4 year old pre-kindergarten, and kindergarten. Within each age strata, the dataset was reduced to a subset of children with rating scale data across the fall, winter, and spring assessment checkpoints, complete age in months data at each checkpoint, and the date of assessment for each checkpoint.

Random samples were then selected from among the qualifying children within each of the six age strata in order to create six samples of 3,000 children, each of which is nationally representative with respect to race and ethnicity. The teachers collected information about the race and ethnicity of each child and entered this information into the online component of the assessment system. The questions they answered about each child were the same as those used by the U.S. Census Bureau. Given that Hispanic identity is an ethnicity, not a racial grouping, and given the importance of representing children of Hispanic ethnicity in the norm sample, the race and ethnicity variables were combined into the following seven ethnic subgroups: 1.) White, not Hispanic, 2.) African-American, not Hispanic, 3.) Native American, not Hispanic, 4.) Asian, not Hispanic, 5.) Hawaiian / Pacific Islander, not Hispanic, 6.) multiracial, not Hispanic, and 7.) Hispanic. The norm sample was created by combining the six samples for a total of 18,000 children. The sampling procedure was conducted to match the U.S. Census Bureau 2009 estimates for children ages birth to 5 years 11 months with respect to the seven ethnic subgroups.

The resulting norm sample contained children from all fifty states, the District of Columbia, and Puerto Rico. The children in the norm sample received preschool services in a variety of settings including Head Start, Early Head Start, private childcare, programs based in school systems, school based sites with Title I funding, state pre-kindergarten programs, programs for military families, university-based programs, early intervention programs, and other programs for children with special needs. As shown in Table 1, the total population of children rated using the *Teaching Strategies GOLD™* assessment system closely approximates the U.S. Census Bureau estimates for the U.S. population of children under the age of 6. White children are represented in numbers very close to the population

estimates (52.1% Census estimates vs. 51.1% norm sample). A similar pattern was found for children with African-American (13.6% Census estimates vs. 14.9% norm sample), Native American (0.9% Census estimates vs. 0.8% norm sample), Asian (4.5% Census estimates vs. 3.6% norm sample), Pacific Islander (0.2% Census estimates vs. 0.1% norm sample), multi-racial (3.2% Census estimates vs. 3.9% norm sample), and Hispanic ethnic background (25.5% Census estimates vs. 25.6% norm sample). Since the ethnic subgroups percentages in the norm sample were close to those in the Census estimates, unweighted data was used for all analyses.

The 18,000 children in the norm sample are 51.4% male and 48.6% female. Typically developing children comprised 91.0% of the norm sample. Children with an IEP (5.4%) or IFSP (3.6%) comprised the remaining 9.0% of the sample. Children from economically disadvantaged backgrounds who qualify for free or reduced lunch comprised 30.4% of the sample. Children from homes where the primary language spoken is English comprise 79.6% of the sample while the remaining 20.4% is made up of children from Spanish speaking homes (15.4%) and homes where other languages are spoken (5.0%).

Rasch Analyses

Rasch scaling, the one parameter IRT model, was used to create ability estimates for each child on each construct and to examine the measurement properties of the information provided by each item. Data were analyzed using the Rasch Rating Scale Model (RSM; Andrich, 1978), with Winsteps software (Linacre, 2012). A separate Rasch analysis was conducted for each of the six domains of development.

The RSM and Partial Credit Model (PCM; Masters, 1982) are the two most widely used Rasch model for polytomous response data. The RSM, rather than the PCM, was

chosen because the items share the same rating scale structure (i.e., use of the same number of rating scale categories and labels across items). In cases where each item has its own rating scale structure, the PCM would be the appropriate model to apply. The decision was also based on preliminary analysis results showing that the RSM yielded better fit of the data to the model. For example, when applying the PCM, there were less than 10 observations in the highest category labeled Level 9. Too few observations in a category can lead to inaccurate and misleading results.

Dimensionality

Rasch modeling assumes what is called unidimensionality, meaning that the items in question measure one and only one underlying latent construct. The unidimensionality of each scale was evaluated by using Mean Square (MNSQ) item fit statistic and Rasch Principal Components Analysis of Residuals (PCAR). The MNSQ fit values between 0.6 and 1.4 are considered reasonable for rating scale items (Bond & Fox, 2007). For PCAR, a variance of greater than 50% explained by measures is considered good, supporting for scale unidimensionality. If a secondary dimension has an eigenvalue of smaller than 3 and accounts for less than 5% of the unexplained variance, unidimensionality is considered plausible (Linacre, 2012).

Social Emotional Scale (9 items)

The principal components analysis of the residuals (PCA) showed that for the Social Emotional scale, the Rasch dimension explained 84.8% of the variance in the data, with its eigenvalue of 50.2. The first contrast (the largest secondary dimension) had an eigenvalue of 2.3 and accounted only for 3.9% of the unexplained variance. The fit statistics for all of

the Social Emotional items were well within acceptable limits: the infit MNSQ ranged from 0.81 to 1.24; the outfit MNSQ ranged from 0.81 to 1.19.

Physical Scale (5 items)

The PCA showed that for the Physical scale, the Rasch dimension explained 88.3% of the variance in the data, with its eigenvalue of 37.6. The first contrast (the largest secondary dimension) had an eigenvalue of 1.8 and accounted only for 4.1% of the unexplained variance. The fit statistics for all of the Physical items were well within acceptable limits: the infit MNSQ ranged from 0.80 to 1.48; the outfit MNSQ ranged from 0.82 to 1.47.

Language Scale (8 items)

The PCA showed that for the Oral Language scale, the Rasch dimension explained 89.2% of the variance in the data, with its eigenvalue of 66.1. The first contrast (the largest secondary dimension) had an eigenvalue of 2.0 and accounted only for 2.7% of the unexplained variance. The fit statistics for all of the Language items were well within acceptable limits: the infit MNSQ ranged from 0.76 to 1.11; the outfit MNSQ ranged from 0.81 to 1.14.

Cognitive Scale (10 items)

The PCA showed that for the Cognitive scale, the Rasch dimension explained 87.8% of the variance in the data, with its eigenvalue of 71.8. The first contrast (the largest secondary dimension) had an eigenvalue of 2.2 and accounted only for 2.6% of the unexplained variance. The fit statistics for all of the Cognitive items were well within acceptable limits: the infit MNSQ ranged from 0.81 to 1.29; the outfit MNSQ ranged from 0.80 to 1.23.

Literacy Scale (12 items)

The PCA showed that the Rasch dimension explained 80.9% of the variance in the data, with its eigenvalue of 50.9. The first contrast (the largest secondary dimension) had an eigenvalue of 2.0 and accounted for 3.2% of the unexplained variance. All Literacy items except one (item 16a) exhibited good fit to the unidimensional Rasch model: the infit MNSQ ranged from 0.72 to 1.90; the outfit MNSQ ranged from 0.62 to 1.48. Item 16a slightly beyond the 1.4 cutoff (infit MNSQ = 1.90; outfit MNSQ = 1.48).

Mathematics Scale (7 items)

The PCA showed that the Rasch dimension explained 82.8% of the variance in the data, with its eigenvalue of 33.6. The first contrast (the largest secondary dimension) had an eigenvalue of 1.8 and accounted for 4.3% of the unexplained variance. All Mathematics items except one (item 20c) exhibited good fit to the unidimensional Rasch model: the infit MNSQ ranged from 0.67 to 1.72; the outfit MNSQ ranged from 0.66 to 1.52. Item 20c slightly beyond the 1.4 cutoff (infit MNSQ = 1.72; outfit MNSQ = 1.52).

In summary, with very few exceptions, these model fit statistics suggest that the data does in fact fit the Rasch rating scale model very well. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Rating Category Effectiveness

The items are measured on a 10-point scale labeled 0 through 9. The use of rating scale categories was examined, which can provide information about whether teachers utilize the instrument in the manner in which it was intended. It is recommended that each rating category has a minimum of 10 observations. The average of the ability estimates for all persons in the sample who chose that particular response category was examined (Bond & Fox, 2007). Average measure score should advance monotonically with rating scale

category values. Thresholds (also called step calibrations) are the difficulties estimated for choosing one response category over another (Bond & Fox, 2007). Thresholds should also increase monotonically with rating scale category. The magnitudes of the distances between adjacent category thresholds should be large enough so that each step defines a distinct position and each category has a distinct peak in the probability curve graph (Bond & Fox, 2007).

For all six scales, the average measure increased with the category level and the thresholds advanced with the categories. An examination of the Rasch category probability curves indicated that for some of the scales, some of the in-between categories (i.e., categories 1, 3, 5, and 7) were not quite as distinct and seemed somewhat redundant with adjacent categories. It is important to note that compared to initial reporting on rating category effectiveness, the rating category structure appeared to function better. This result may suggest that teachers can more clearly distinguish between an "Indicator" level and an "In-between" level. This may be a result of the expanded teacher training and the expansion of the adoption of the assessment system since the last report.

Item Difficulty Measures

For all six scales, the item location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children.

Social Emotional Scale

The item pertaining to a child's ability to balance needs and rights of self and others (3a) was found to be the most difficult item, whereas the item pertaining to a child's ability to form relationships with adults (2a) was estimated as the easiest item.

Physical Scale

The item pertaining to a child's ability to use writing tools (7b) was found to be the most difficult item, whereas the item pertaining to a child's ability to walk (4) was estimated as the easiest item. The item pertaining to a child's ability to use their fingers and hands (7a) was also rated as approximately as easy as the item pertaining to walking.

Language Scale

The item pertaining to a child's ability to use an expanding expressive vocabulary (9d) was found to be the most difficult item, whereas the item pertaining to a child's ability to ability to speak clearly (9b) was estimated as the easiest item.

Cognitive Scale

The item pertaining to a child's use of classification skills (13) was found to be the most difficult item, whereas the item pertaining to a child's ability to attend and engage (11a) was estimated as the easiest item.

Literacy Scale

The item pertaining to a child's use of letter-sound knowledge (16b) was found to be the most difficult item, whereas the item pertaining to a child's use and appreciation of books (17a) was estimated as the easiest item.

Mathematics Scale

The item pertaining to a child's ability of connecting numerals with quantities (20c) was found to be the most difficult item, whereas the items pertaining to a child's ability to explore and describe spatial relationships and shapes (21a and 21b) were estimated as the two easiest items.

Taken as a whole the item difficulty statistics indicate that the test developers were very successful in creating measures that offer a developmental pathway of sequential milestones that agree with developmental theory.

Reliability

Reliability was evaluated using person separation index, item separation index, person reliability, and item reliability provided by Winsteps. The person separation index, an estimate of the adjusted person standard deviation divided by the average measurement error, indicates how well the instrument can discriminate persons on each of the constructs. The item separation index indicates an estimate in standard error units of the spread or separation of items along the measurement constructs. Reliability separation indexes greater than 2 are considered adequate (Bond & Fox, 2007). High person or item reliability means that there is a high probability of replicating the same separation of persons or items across measurements. Specifically, person separation reliability estimates the replicability of person placement across other items measuring the same construct. Similarly, item separation reliability estimates the replicability of item placement along the construct development pathway if the same items were given to another sample with similar ability levels. The person reliability provided by Winsteps is equivalent to the traditional test reliability whereas the item reliability has no traditional equivalent. Low values in person and item reliability may indicate a narrow range of person or item measures. It may also indicate that the number of items or the sample size under study is too small for stable estimates (Linacre, 2009).

Social Emotional Scale

Based on the Rasch reliability indexes, the scale appear to be highly reliable, as evidenced by person separation indexes of 6.26, person reliabilities of .98, item separation indexes of 89.19, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .975, indicating high internal consistency reliability.

Physical Scale

Based on the Rasch reliability indexes, the scale appear to be highly reliable, as evidenced by person separation indexes of 5.50, person reliabilities of .97, item separation indexes of 55.22, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .97, indicating high internal consistency reliability.

Language Scale

Based on the Rasch reliability indexes, the scale appears to be highly reliable, as evidenced by person separation indexes of 7.09, person reliabilities of .98, item separation indexes of 80.86, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability.

Cognitive Scale

Based on the Rasch reliability indexes, the scale appear to be highly reliable, as evidenced by person separation indexes of 7.63, person reliabilities of .98, item separation indexes of 65.73, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability.

Literacy Scale

Based on the Rasch reliability indexes, the scale appear to be highly reliable, as evidenced by person separation indexes of 4.90, person reliabilities of .96, item separation indexes of

69.73, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability.

Mathematics Scale

Based on the Rasch reliability indexes, the scale appear to be highly reliable, as evidenced by person separation indexes of 4.92, person reliabilities of .96, item separation indexes of 44.93, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability.

In addition, we examined the internal consistency reliability across each of the three assessment checkpoints and each of the six age grade strata. As shown in Table 2, the Cronbach's alpha values are very consistent for each of the scale scores across all of these conditions. For the Cognitive scale scores, the values ranged from .910 to .988, all within the acceptable range. A similar pattern emerged for each of the other scale scores: Literacy (.855 - .988), Social Emotional (.904 - .982), Physical (.907 - .974), Language (.889 - .985), and Mathematics (.882 - .985). The relatively lowest values for each scale score, though still well within the acceptable range, were associated with the fall assessments of the youngest children. The highest values for each scale score were associated with the spring assessments of the oldest children. This finding suggests that teachers may become more consistent in the ratings as they get to know children across the academic year and collect more artifacts, evidences, and anecdotes of developmental progress. It may also suggest that teachers can be more consistent when rating older children for whom it is easier to document the relevant evidences to support ratings. However, it is noteworthy that these reliability values indicate acceptable and even in some cases high levels of internal consistency reliability, even associated with ratings of infants.

Scale Scores and Norm Tables

For the purpose of creating norm tables for the *Teaching Strategies GOLD™* assessment system, the children in the norm sample were divided into three month age bands. A total of 24 age bands were created with the youngest children falling into the 0-2 month category and the oldest children falling into the 69-71 month category. These strata were based on the ages of the children in months at the time of the first assessment. The fall assessment date for most programs took place in October of the academic year. Ratings were based on a portfolio of evidences that the teachers collected up to that date, including anecdotal records and artifacts. The children in the norm sample span the entire age range for which the assessment system is intended (birth through kindergarten).

The body of evidence to date from research studies and the Rasch modeling suggests that scale scores for each of the developmental domains outlined by the test developers would be appropriate. The scale scores were created by first creating interval level Rasch rating scale ability estimates. The ability estimates were then rescaled to conform to a distribution with a mean of 500 and standard deviation of 100. Values three or more standard deviations below the mean were given a value of 200 and values three or more standard deviations above the mean were given a value of 800. This scaling strategy is commonly used in educational and psychological testing.

For each scale score and three month age band, as shown in Tables 3 through 14, the mean, standard deviation, quartile boundaries, and standard error of measurement are reported. These results suggest that teachers are generally giving higher scores to older children and lower scores to younger children, while also discriminating between children of similar ages but differing rates of development as expected. They are also able to track

growth and development across the three assessment checkpoints. As can be seen across all six scale scores and across all 24 age bands, the mean scores for the age bands increase with age at a steady pace, enabling the tracking of developmental progress for children on an interval scale from year to year using the same measure. The quartile boundaries are also included to enable teachers to understand approximately where a child's falls relative to other children in the norm sample.

The norm tables include expected scores for fall, winter, and spring assessments and expected growth from fall to spring and clearly demonstrate that teachers can use the GOLD assessment system to track growth across the academic year for children of different ages. Children in the 0-35 month age bands can be expected to make about 50-60 scale score points of growth across the academic year. Children in the 36-71 month age bands can be expected to make about 70-100 scale score points of growth across the academic year.

It is important to note that information about these updated norm tables include children in the youngest two three month age bands (0-2 and 3-5). There was not sufficient data with acceptable reliability and validity from these subgroups to include in the first edition of this manual. Their inclusion in this manual is evidence that the expanded and ongoing training available to users and the expansion of the adoption of the assessment system has yielded useful data for even the youngest children.

Summary

Overall, the *Teaching Strategies GOLD™* assessment system appears to continue to yield highly reliable scores as indicated by both the classical and Rasch reliability statistics. The high reliability statistics were not only found for the overall norm sample, but extend

to each age cohort: birth to 1, 1 to 2, 2 to 3, three year old preschool, four year old prekindergarten, and kindergarten and across fall, winter, and spring assessments for all scales and age cohorts.

The results show strong statistical evidence that the items within each scale work very well together to measure a single underlying construct or domain of development. The items within each scale yield information that fits the statistical model that was used to develop the scoring strategy that is used to create the scale scores. The results further demonstrate evidence that the ratings can be successfully organized by developmental domain or latent construct generally as intended by the instrument development team. Analyses of the dimensionality of each scale score strongly suggest that the GOLD™ assessment system ratings measure six distinct domains of development and that each satisfies the Rasch model assumption of unidimensionality. The model fit statistics suggest that the data are a good fit for the Rasch rating scale model. These results also strongly suggest that teachers are able to make valid ratings of the developmental progress of children across the intended age range, from birth through 72 months.

There is also strong statistical evidence that teachers are able to use the rating scale to place children along a continuum of growth and development. When the items within each domain of development are arranged from the easier objectives for children to master to the most difficult objectives for children to master, the hierarchy that is created matches very well with what developmental theory indicates. Therefore, the range of item difficulties indicates that each section of the GOLD assessment can be used by teachers to help them understand the developmental trajectory that most children will follow.

References

Andrich, D. (1978). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*, 581-594.

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

Kim, D-H., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of *Teaching Strategies GOLD*® assessment tool for English language learners and children with disabilities. *Early Education and Development, 24*, 574-595.

Kim, D-H., Lambert, R. G., & Burts, D. C. (in press). Validating a developmental scale for young children using the Rasch Model: Applicability of the *Teaching Strategies GOLD*®assessment system. *Journal of Applied Measurement.*

Lambert, R. & Kim, D-H., Taylor, H., & McGee, J. (2010). *Technical manual for the Teaching Strategies GOLD assessment system.* Technical Report. Charlotte, N.C.: *Center for Educational Measurement and Evaluation*, University of North Carolina Charlotte.

Lambert, R. G., Kim, D-H., & Burts, D. C. (2012). [The measurement properties of the *Teaching Strategies GOLD*® assessment system]. Unpublished raw data.

Lambert, R. G., Kim, D-H., & Burts, D. C. (2013). Using teacher ratings to track the growth and development of young children using the *Teaching Strategies GOLD*® assessment system. *Journal of Psychoeducational Assessment*. doi:0734282913485214

Linacre, J. M. (2012). Winsteps (Version 3.75.1) [Computer Software]. Chicago, IL: Winsteps.com.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Table 1
Norm Sample by Ethnic Subgroup.

Racial and Ethnic Subgroup	2009 Census Bureau Estimates	Norm Sample
White, not Hispanic	52.1%	51.1%
African American, not Hispanic	13.6%	14.9%
Native American, not Hispanic	0.9%	0.8%
Asian, not Hispanic	4.5%	3.6%
Hawaiian or Pacific Islander, not Hispanic	0.2%	0.1%
Multirace, not Hispanic	3.2%	3.9%
Hispanic	25.5%	25.6%

Table 2
Cronbach's alpha by scale score, age group, and time of year

Scale	Age Group	Fall	Winter	Spring
Cognitive	Birth to 1 year	.910	.929	.941
	1 to 2 years	.932	.941	.948
	2 to 3 years	.950	.953	.958
	Pre-kindergraten - 4 year olds	.954	.954	.962
	Pre-kindergraten - 3 year olds	.962	.965	.970
	Kindergarten	.962	.961	.967
	All age groups	.984	.986	.988
Literacy	Birth to 1 year	.855	.874	.911
	1 to 2 years	.889	.918	.934
	2 to 3 years	.923	.933	.940
	Pre-kindergraten - 4 year olds	.954	.957	.962
	Pre-kindergraten - 3 year olds	.936	.942	.953
	Kindergarten	.949	.947	.958
	All age groups	.979	.985	.988
Social Emotional	Birth to 1 year	.904	.920	.931
	1 to 2 years	.917	.923	.934
	2 to 3 years	.929	.932	.938
	Pre-kindergraten - 4 year olds	.947	.947	.956
	Pre-kindergraten - 3 year olds	.939	.937	.943
	Kindergarten	.946	.941	.941
	All age groups	.975	.979	.982
Physical	Birth to 1 year	.907	.916	.923
	1 to 2 years	.911	.913	.916
	2 to 3 years	.910	.912	.920
	Pre-kindergraten - 4 year olds	.916	.916	.922
	Pre-kindergraten - 3 year olds	.903	.897	.906
	Kindergarten	.882	.963	.975
	All age groups	.966	.971	.974
Language	Birth to 1 year	.889	.908	.924
	1 to 2 years	.928	.933	.942
	2 to 3 years	.950	.952	.954
	Pre-kindergraten - 4 year olds	.959	.961	.965
	Pre-kindergraten - 3 year olds	.953	.950	.955
	Kindergarten	.946	.971	.979
	All age groups	.981	.984	.985
Mathematics	Birth to 1 year	.882	.895	.914
	1 to 2 years	.906	.917	.921
	2 to 3 years	.922	.896	.908
	Pre-kindergraten - 4 year olds	.917	.918	.941
	Pre-kindergraten - 3 year olds	.921	.922	.937
	Kindergarten	.937	.940	.948
	All age groups	.975	.981	.985

Table 3
Cognitive Scale Score norms by 3 month age band- ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	362.97	388.87	411.31	48.34
		SD	39.27	36.62	38.58	28.53
		25th	331	365	382	30
		50th	365	387	402	46
		75th	387	412	438	66
		SEM	13	11	11	
3-5	823	Mean	367.59	392.12	415.60	48.01
		SD	34.41	33.79	38.37	32.30
		25th	351	371	392	28
		50th	365	387	407	46
		75th	387	407	432	65
		SEM	13	11	11	
6-8	598	Mean	376.90	402.96	431.29	54.39
		SD	32.76	34.03	38.23	33.10
		25th	365	382	402	35
		50th	377	397	427	51
		75th	392	422	449	72
		SEM	12	11	12	
9-11	335	Mean	385.80	416.03	440.68	54.88
		SD	29.09	32.40	34.93	31.57
		25th	371	397	417	35
		50th	387	412	438	52
		75th	402	432	460	74
		SEM	11	11	12	
12-14	931	Mean	422.41	451.74	474.85	52.44
		SD	38.14	41.63	43.66	37.07
		25th	397	427	449	28
		50th	417	449	470	48
		75th	443	470	496	70
		SEM	12	12	12	
15-17	786	Mean	427.92	457.15	483.84	55.92
		SD	39.76	40.99	44.35	38.00
		25th	402	432	454	32
		50th	427	454	481	53
		75th	454	481	507	76
		SEM	12	12	12	
18-20	718	Mean	435.01	466.82	494.57	59.55
		SD	39.95	42.42	46.65	40.36
		25th	412	443	465	33
		50th	432	465	491	57
		75th	460	491	518	81
		SEM	12	12	12	
21-23	565	Mean	445.53	477.31	507.41	61.88
		SD	40.93	45.30	50.18	44.97
		25th	422	449	476	32
		50th	449	476	507	59
		75th	470	502	537	89
		SEM	12	12	12	
24-26	1095	Mean	479.85	509.30	535.10	55.25
		SD	48.70	49.58	51.89	41.47
		25th	449	476	502	32
		50th	476	507	530	53
		75th	507	537	563	77
		SEM	12	12	13	
27-29	828	Mean	481.19	512.08	541.52	60.33
		SD	49.27	50.08	54.82	47.03
		25th	449	481	507	32
		50th	476	507	537	58
		75th	507	537	576	85
		SEM	12	12	13	
30-32	700	Mean	495.94	526.64	558.01	62.07
		SD	51.14	54.38	56.85	45.76
		25th	460	491	524	33
		50th	494	524	557	60
		75th	524	557	587	90
		SEM	12	13	13	
33-35	377	Mean	502.69	539.74	575.38	72.69
		SD	55.15	57.03	57.89	49.12
		25th	470	505	543	41
		50th	496	537	570	68
		75th	537	576	604	106
		SEM	12	13	12	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 4
Cognitive Scale Score norms by 3 month age band- ages 36-71 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
36-38	892	Mean	515.73	557.47	592.18	76.44
		SD	61.70	64.28	70.37	45.93
		25th	476	518	550	48
		50th	518	557	587	74
		75th	550	593	625	102
		SEM	12	13	12	
39-41	761	Mean	525.07	572.98	610.27	85.20
		SD	61.30	62.68	69.87	53.02
		25th	491	537	570	52
		50th	524	570	604	78
		75th	557	609	650	113
		SEM	12	13	12	
42-44	798	Mean	535.91	585.09	623.79	87.88
		SD	66.39	68.01	72.45	55.39
		25th	496	543	582	57
		50th	537	582	614	83
		75th	570	614	665	117
		SEM	13	12	12	
45-47	549	Mean	545.27	592.76	630.94	85.67
		SD	64.69	62.70	68.74	49.94
		25th	507	550	587	54
		50th	537	587	620	80
		75th	582	625	673	117
		SEM	13	12	12	
48-50	792	Mean	565.18	620.23	668.29	103.12
		SD	61.70	61.88	66.10	54.79
		25th	526	587	625	66
		50th	570	620	673	101
		75th	604	658	710	134
		SEM	13	12	14	
51-53	826	Mean	571.12	628.19	680.70	109.58
		SD	60.26	58.19	65.87	59.75
		25th	537	593	637	72
		50th	576	631	681	105
		75th	609	665	728	140
		SEM	13	12	14	
54-56	773	Mean	575.72	636.00	690.71	114.99
		SD	59.03	57.05	61.04	60.12
		25th	543	598	650	74
		50th	582	637	697	113
		75th	609	673	733	146
		SEM	12	13	14	
57-59	609	Mean	581.74	639.30	692.47	110.74
		SD	61.42	61.21	64.77	59.05
		25th	543	604	650	71
		50th	587	637	697	112
		75th	620	681	739	143
		SEM	12	13	14	
60-62	811	Mean	600.70	655.34	704.43	107.32
		SD	80.39	74.79	81.31	69.13
		25th	557	620	665	61
		50th	609	665	716	102
		75th	658	703	767	144
		SEM	12	14	13	
63-65	545	Mean	609.64	669.04	721.57	114.76
		SD	83.70	76.74	80.06	68.64
		25th	570	631	689	72
		50th	620	673	739	113
		75th	665	722	783	146
		SEM	12	14	12	
66-68	463	Mean	620.97	678.00	731.12	111.41
		SD	74.50	71.16	73.99	60.50
		25th	576	637	697	73
		50th	631	681	747	113
		75th	665	728	791	147
		SEM	12	14	12	
69-71	408	Mean	626.46	684.16	736.22	117.78
		SD	75.87	74.75	77.68	64.49
		25th	589	654	716	79
		50th	637	697	756	118
		75th	673	733	791	150
		SEM	12	14	12	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 5
Literacy Scale Score norms by 3 month age band - ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	354.01	379.93	403.06	49.05
		SD	43.86	46.86	48.60	39.10
		25th	312	360	360	28
		50th	360	388	406	48
		75th	388	419	438	76
		SEM	39	28	24	
3-5	823	Mean	355.43	381.52	406.43	51.00
		SD	43.92	45.59	48.99	44.99
		25th	312	360	388	18
		50th	360	388	406	48
		75th	388	406	438	76
		SEM	39	28	24	
6-8	598	Mean	360.29	391.32	424.08	63.79
		SD	43.14	44.32	43.29	44.86
		25th	312	360	388	31
		50th	360	388	429	59
		75th	388	419	446	93
		SEM	39	28	21	
9-11	335	Mean	374.99	411.37	435.91	60.92
		SD	44.32	41.37	39.86	48.42
		25th	360	388	419	28
		50th	388	419	438	58
		75th	406	438	460	86
		SEM	28	24	18	
12-14	931	Mean	423.62	449.53	469.66	46.04
		SD	44.00	39.18	41.60	42.50
		25th	388	429	446	17
		50th	419	446	466	40
		75th	446	472	493	69
		SEM	21	17	15	
15-17	786	Mean	431.77	457.39	480.18	48.41
		SD	43.91	41.41	41.03	39.98
		25th	406	429	453	22
		50th	438	453	477	44
		75th	453	483	507	68
		SEM	19	16	14	
18-20	781	Mean	439.49	466.19	490.28	50.79
		SD	40.38	39.47	40.01	38.54
		25th	419	446	466	26
		50th	438	466	488	49
		75th	466	488	515	71
		SEM	18	15	14	
21-23	565	Mean	448.97	477.37	501.24	52.27
		SD	42.85	44.51	44.08	41.16
		25th	429	453	477	25
		50th	453	477	502	48
		75th	477	502	528	77
		SEM	17	14	13	
24-26	1095	Mean	478.77	503.12	522.37	43.60
		SD	44.57	41.59	42.54	37.00
		25th	453	477	497	21
		50th	483	502	520	41
		75th	507	528	548	62
		SEM	14	13	13	
27-29	828	Mean	480.62	506.91	528.97	48.35
		SD	42.27	38.85	42.50	39.46
		25th	453	483	502	23
		50th	483	507	528	44
		75th	507	532	555	67
		SEM	14	13	13	
30-32	700	Mean	491.58	517.51	541.78	50.20
		SD	43.75	44.48	44.13	37.10
		25th	472	493	520	27
		50th	493	520	536	48
		75th	520	544	569	70
		SEM	14	13	12	
33-35	377	Mean	497.53	527.48	550.41	52.88
		SD	47.10	44.29	41.95	38.35
		25th	472	502	528	30
		50th	493	528	548	51
		75th	528	551	572	76
		SEM	13	13	12	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 6
Literacy Scale Score norms by 3 month age band - ages 36-71 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
36-38	892	Mean	520.84	555.36	581.65	60.81
		SD	53.40	53.84	55.63	34.12
		25th	488	524	551	39
		50th	520	555	579	60
		75th	551	586	614	79
		SEM	13	13	11	
39-41	761	Mean	529.63	566.82	594.86	65.23
		SD	51.09	49.04	53.52	36.69
		25th	502	536	562	40
		50th	528	566	592	60
		75th	562	595	629	84
		SEM	13	11	11	
42-44	798	Mean	536.20	575.92	605.07	68.86
		SD	54.27	52.17	56.50	41.35
		25th	507	544	569	43
		50th	532	572	598	64
		75th	569	605	635	89
		SEM	12	11	11	
45-47	549	Mean	547.23	585.23	614.30	67.07
		SD	53.30	51.46	55.94	37.06
		25th	515	555	579	42
		50th	544	582	608	64
		75th	579	614	647	90
		SEM	12	11	11	
48-50	792	Mean	563.38	606.70	643.62	80.24
		SD	49.33	48.03	52.45	40.28
		25th	533	579	611	52
		50th	566	608	644	76
		75th	595	635	678	103
		SEM	12	11	11	
51-53	826	Mean	567.46	614.70	652.60	85.15
		SD	50.28	45.58	51.89	43.69
		25th	540	589	623	57
		50th	569	617	653	80
		75th	598	641	684	105
		SEM	11	11	11	
54-56	773	Mean	576.00	623.10	661.65	85.65
		SD	46.18	44.37	50.50	43.78
		25th	548	595	629	56
		50th	576	620	659	79
		75th	601	650	694	108
		SEM	11	11	11	
57-59	609	Mean	580.66	627.47	665.22	84.56
		SD	49.10	49.69	53.81	43.32
		25th	555	598	632	58
		50th	582	626	668	81
		75th	614	656	697	106
		SEM	11	11	11	
60-62	1018	Mean	608.45	662.99	706.95	99.70
		SD	62.04	53.98	57.69	52.81
		25th	572	632	668	64
		50th	611	665	714	93
		75th	650	697	750	131
		SEM	11	11	11	
63-65	751	Mean	616.18	673.72	722.65	108.77
		SD	65.69	63.78	63.45	50.15
		25th	576	641	694	75
		50th	620	681	737	107
		75th	659	717	767	138
		SEM	11	11	12	
66-68	630	Mean	624.87	685.07	732.37	105.26
		SD	64.48	57.86	53.45	50.66
		25th	589	656	710	73
		50th	629	690	743	107
		75th	665	725	767	135
		SEM	11	11	12	
69-71	578	Mean	632.74	695.36	740.27	110.31
		SD	59.94	54.42	51.93	50.67
		25th	598	670	721	76
		50th	632	700	755	108
		75th	675	733	767	142
		SEM	11	11	13	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 7
Social Emotional Scale Score norms by 3 month age band - ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	364.01	395.62	422.85	58.84
		SD	47.96	45.63	47.46	39.71
		25th	333	367	389	33
		50th	367	389	415	56
		75th	389	421	452	80
		SEM	17	16	15	
3-5	823	Mean	369.73	399.01	428.32	58.59
		SD	43.65	41.52	44.85	38.44
		25th	343	375	395	35
		50th	367	395	421	56
		75th	389	421	452	80
		SEM	17	15	15	
6-8	598	Mean	376.83	412.84	446.60	69.77
		SD	41.67	44.06	45.91	41.73
		25th	351	389	415	44
		50th	375	408	445	66
		75th	395	433	469	92
		SEM	16	15	15	
9-11	335	Mean	389.23	426.38	454.30	65.07
		SD	40.06	41.46	41.61	40.89
		25th	367	402	427	42
		50th	389	427	452	63
		75th	408	452	475	86
		SEM	16	15	15	
12-14	931	Mean	430.21	462.69	488.32	58.11
		SD	46.74	46.20	46.61	43.51
		25th	402	433	458	34
		50th	427	458	481	54
		75th	458	486	514	80
		SEM	15	15	14	
15-17	786	Mean	433.71	466.82	432.20	59.49
		SD	48.45	46.41	47.04	44.23
		25th	408	439	463	31
		50th	433	469	492	57
		75th	463	492	519	82
		SEM	15	15	15	
18-20	718	Mean	441.48	474.89	503.89	62.41
		SD	49.16	47.30	49	44.17
		25th	415	452	475	36
		50th	439	475	503	61
		75th	469	503	530	87
		SEM	15	15	14	
21-23	565	Mean	452.92	483.65	514.66	61.74
		SD	52.45	48.24	48.79	52.76
		25th	427	458	481	33
		50th	458	486	514	59
		75th	481	514	541	91
		SEM	15	14	14	
24-26	1095	Mean	483.33	510.74	535.85	52.53
		SD	51.73	49.25	50.92	44.07
		25th	452	481	508	57
		50th	481	514	535	50
		75th	514	535	557	77
		SEM				
27-29	828	Mean	481.97	512.15	540.02	58.05
		SD	51.56	48.45	53.59	50.35
		25th	452	482	508	30
		50th	481	514	535	54
		75th	514	541	568	82
		SEM	14	14	14	
30-32	700	Mean	495.70	524.70	552.71	57.01
		SD	54.35	55.54	54.53	47.72
		25th	463	492	524	28
		50th	497	524	546	55
		75th	524	552	574	85
		SEM	14	14	14	
33-35	377	Mean	504.52	539.02	570.15	65.63
		SD	64.72	57.02	57.06	54.93
		25th	469	503	535	33
		50th	508	535	563	66
		75th	541	574	602	93
		SEM	14	14	14	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 8
Social Emotional Scale Score norms by 3 month age band - ages 36-71 months

Age Band	n					
36-38	892	Mean	510.99	553.92	587.82	76.83
		SD	66.04	63.68	69.70	49.02
		25th	469	519	546	45
		50th	514	552	583	72
		75th	546	590	520	104
		SEM	14	14	14	
39-41	761	Mean	520.06	566.33	604.13	84.07
		SD	63.43	61.34	71.64	54.34
		25th	481	530	557	49
		50th	519	563	596	77
		75th	552	602	638	110
		SEM	14	14	15	
42-44	798	Mean	573.42	581.99	619.07	83.65
		SD	69.10	67.20	73.78	57.01
		25th	497	541	574	50
		50th	530	574	608	77
		75th	568	614	658	113
		SEM	14	14	15	
45-47	549	Mean	540.14	585.87	622.95	82.80
		SD	64.68	61.94	68.46	52.38
		25th	503	546	580	49
		50th	541	585	614	79
		75th	574	620	658	110
		SEM	14	14	15	
48-50	792	Mean	562.32	613.36	663.17	100.84
		SD	64.80	63.67	68.69	58.13
		25th	530	580	320	61
		50th	568	614	658	94
		75th	602	658	709	135
		SEM	14	15	16	
51-53	826	Mean	565.42	618.10	672.10	107.28
		SD	60.71	56.19	66.58	62.26
		25th	530	585	632	67
		50th	568	620	672	98
		75th	602	652	717	140
		SEM	14	15	16	
54-56	773	Mean	570.67	628.05	682.47	111.80
		SD	60.00	56.17	66.14	65.20
		25th	535	593	638	70
		50th	580	626	679	106
		75th	608	658	725	148
		SEM	14	15	16	
57-59	609	Mean	574.86	627.94	681.17	106.30
		SD	62.44	62.15	69.27	60.83
		25th	535	591	632	67
		50th	580	626	679	101
		75th	614	658	725	142
		SEM	14	15	16	
60-62	1018	Mean	596.81	652.70	702.82	107.75
		SD	76.43	67.81	73.72	66.03
		25th	546	614	656	63
		50th	608	652	709	103
		75th	645	701	760	148
		SEM	14	16	17	
63-65	751	Mean	606.61	665.73	719.21	115.56
		SD	80.07	77.02	79.18	65.02
		25th	562	626	679	74
		50th	614	672	735	115
		75th	658	717	795	455
		SEM	15	16	18	
66-68	630	Mean	615.57	675.39	727.04	113.13
		SD	75.90	75.00	73.59	60.26
		25th	574	632	686	73
		50th	626	679	746	111
		75th	665	725	800	155
		SEM	15	16	18	
69-71	578	Mean	624.22	682.88	735.54	120.32
		SD	81.74	74.54	72.53	63.44
		25th	585	632	701	75
		50th	632	686	760	120
		75th	679	725	800	162
		SEM	15	16	19	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 9
Physical Scale Score norms by 3 month age band - ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	352.39	388.58	416.21	63.83
		SD	50.17	46.60	48.65	37.49
		25th	324	361	380	37
		50th	353	380	415	61
		75th	380	415	451	87
		SEM	15	15	15	
3-5	823	Mean	361.33	392.81	423.38	62.05
		SD	43.73	42.03	46.29	42.55
		25th	335	370	389	37
		50th	361	380	415	61
		75th	380	415	451	81
		SEM	15	15	15	
6-8	598	Mean	377.13	411.15	442.02	64.89
		SD	37.88	40.68	44.07	40.74
		25th	353	380	415	43
		50th	370	407	432	62
		75th	389	424	461	83
		SEM	15	15	15	
9-11	335	Mean	391.33	425.41	452.52	61.19
		SD	38.28	41.17	44.20	40.09
		25th	370	407	424	37
		50th	389	424	451	62
		75th	407	441	471	81
		SEM	15	15	16	
12-14	931	Mean	433.60	462.08	484.79	51.19
		SD	47.60	48.89	49.94	45.99
		25th	407	432	461	26
		50th	432	461	471	46
		75th	461	482	514	72
		SEM	15	16	17	
15-17	786	Mean	439.99	468.20	493.40	53.41
		SD	49.62	49.86	52.19	46.74
		25th	415	441	461	27
		50th	441	461	482	52
		75th	461	494	524	81
		SEM	15	17	17	
18-20	718	Mean	446.84	474.70	502.44	55.60
		SD	49.95	49.18	51.82	44.71
		25th	415	451	471	30
		50th	441	471	494	53
		75th	471	504	534	80
		SEM	15	17	16	
21-23	565	Mean	456.30	483.85	513.73	57.43
		SD	53.61	53.67	53.15	53.97
		25th	424	451	482	30
		50th	461	482	514	54
		75th	482	514	543	83
		SEM	16	17	16	
24-26	1095	Mean	488.44	514.05	537.88	49.44
		SD	52.08	51.83	54.62	44.51
		25th	461	482	504	22
		50th	482	514	534	49
		75th	524	543	565	73
		SEM	17	16	16	
27-29	828	Mean	486.68	515.39	541.23	54.55
		SD	54.01	52.09	54.66	50.84
		25th	461	482	514	23
		50th	482	514	543	52
		75th	524	543	577	82
		SEM	17	16	16	
30-32	700	Mean	503.23	528.46	555.64	52.40
		SD	56.81	54.77	56.19	52.72
		25th	471	494	524	23
		50th	504	534	554	52
		75th	534	554	590	83
		SEM	16	16	17	
33-35	377	Mean	508.42	541.32	571.94	63.52
		SD	64.44	61.05	61.01	53.38
		25th	466	509	534	34
		50th	514	543	565	62
		75th	543	577	603	90
		SEM	16	16	18	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 10
Physical Scale Score norms by 3 month age band - ages 36-71 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
36-38	892	Mean	511.81	550.36	583.24	71.43
		SD	62.66	60.34	64.46	49.88
		25th	471	514	543	39
		50th	514	554	577	63
		75th	554	577	615	101
		SEM	16	17	18	
39-41	761	Mean	519.83	560.19	596.95	77.12
		SD	60.27	58.04	64.08	54.62
		25th	482	524	554	40
		50th	524	565	590	71
		75th	554	590	637	106
		SEM	16	18	18	
42-44	798	Mean	530.13	572.86	608.62	78.49
		SD	65.13	63.24	68.17	53.23
		25th	494	534	565	47
		50th	534	577	603	72
		75th	565	603	647	106
		SEM	16	18	18	
45-47	549	Mean	538.31	579.74	613.65	75.34
		SD	57.91	55.99	62.79	48.84
		25th	504	543	577	46
		50th	534	577	615	71
		75th	577	615	647	105
		SEM	16	18	17	
48-50	792	Mean	555.96	604.16	651.62	95.65
		SD	59.85	58.84	65.11	58.80
		25th	524	577	615	57
		50th	565	603	647	93
		75th	590	637	686	129
		SEM	17	18	17	
51-53	826	Mean	559.96	609.55	662.25	102.29
		SD	58.56	55.05	62.47	61.43
		25th	534	577	626	61
		50th	565	615	658	94
		75th	590	647	706	134
		SEM	18	17	17	
54-56	773	Mean	564.82	618.47	671.27	106.45
		SD	58.13	52.62	60.93	63.49
		25th	534	590	637	61
		50th	577	626	671	101
		75th	603	647	706	140
		SEM	18	17	19	
57-59	609	Mean	570.41	619.33	672.03	101.62
		SD	62.36	59.80	66.18	61.67
		25th	534	590	626	60
		50th	577	626	671	96
		75th	603	647	706	140
		SEM	18	17	19	
60-62	786	Mean	596.95	643.76	681.73	83.13
		SD	77.68	98.05	104.54	80.47
		25th	554	626	658	45
		50th	615	658	706	85
		75th	647	686	743	128
		SEM	18	17	21	
63-65	684	Mean	611.69	667.14	702.96	86.27
		SD	75.97	84.83	80.14	60.05
		25th	565	637	671	51
		50th	626	671	743	86
		75th	658	743	743	113
		SEM	17	19	26	
66-68	602	Mean	620.21	669.59	702.71	78.65
		SD	72.54	77.31	70.00	56.73
		25th	577	637	671	48
		50th	626	658	706	81
		75th	658	743	743	106
		SEM	17	19	26	
69-71	566	Mean	627.76	670.06	708.34	82.48
		SD	70.98	87.68	82.55	71.50
		25th	590	647	686	45
		50th	637	671	743	85
		75th	658	743	743	117
		SEM	17	19	26	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 11

Language Scale Score norms by 3 month age band - ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	367.99	393.28	413.44	45.45
		SD	38.82	35.01	35.38	27.66
		25th	349	373	392	28
		50th	373	392	408	42
		75th	392	413	433	62
		SEM	14	12	12	
3-5	823	Mean	370.83	395.43	417.22	46.39
		SD	35.60	31.33	32.90	31.32
		25th	349	380	397	25
		50th	373	397	413	44
		75th	392	408	433	61
		SEM	13	12	12	
6-8	598	Mean	378.41	404.35	428.59	50.18
		SD	36.43	33.30	34.07	34.03
		25th	366	386	408	31
		50th	380	403	428	46
		75th	397	418	443	65
		SEM	13	12	11	
9-11	335	Mean	388.47	414.49	436.58	48.10
		SD	32.47	28.66	29.59	33.11
		25th	373	403	418	26
		50th	392	418	433	44
		75th	408	433	457	65
		SEM	12	12	11	
12-14	931	Mean	421.05	446.81	470.18	49.12
		SD	38.20	37.85	40.78	34.92
		25th	403	423	447	28
		50th	423	443	466	44
		75th	443	466	487	66
		SEM	12	11	12	
15-17	786	Mean	428.26	454.79	480.91	52.65
		SD	37.94	38.17	42.38	35.24
		25th	408	433	457	33
		50th	428	452	476	48
		75th	452	476	505	70
		SEM	11	11	12	
18-20	718	Mean	436.72	463.32	491.48	54.76
		SD	39.10	40.72	45.32	36.22
		25th	413	443	466	31
		50th	436	461	487	53
		75th	461	483	517	73
		SEM	11	11	12	
21-23	565	Mean	446.67	473.24	504.23	57.56
		SD	42.54	42.72	50.70	41.87
		25th	418	447	471	29
		50th	447	471	499	54
		75th	476	499	535	80
		SEM	11	12	13	
24-26	1095	Mean	480.69	507.11	533.06	52.36
		SD	51.92	53.29	56.90	39.88
		25th	452	471	499	27
		50th	481	505	529	50
		75th	511	535	560	73
		SEM	12	13	13	
27-29	828	Mean	484.18	512.08	539.81	55.64
		SD	51.86	51.38	55.81	45.93
		25th	453	481	505	27
		50th	481	511	535	52
		75th	511	541	567	78
		SEM	12	13	13	
30-32	700	Mean	501.68	527.92	557.10	55.42
		SD	56.08	59.05	61.76	48.70
		25th	466	493	523	29
		50th	499	523	547	55
		75th	535	560	597	80
		SEM	13	13	13	
33-35	377	Mean	506.73	537.11	570.75	64.02
		SD	58.76	61.79	64.70	47.30
		25th	469	499	535	34
		50th	505	535	567	61
		75th	541	573	606	96
		SEM	13	13	14	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 12
Language Scale Score norms by 3 month age band - ages 36-71 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
36-38	892	Mean	509.73	546.17	580.96	71.24
		SD	65.84	68.69	76.74	46.22
		25th	466	505	535	42
		50th	511	541	573	67
		75th	547	581	625	95
		SEM	13	13	14	
39-41	761	Mean	523.08	564.69	602.95	79.87
		SD	63.16	67.81	76.76	53.07
		25th	481	523	547	46
		50th	523	567	597	73
		75th	560	606	654	110
		SEM	13	13	15	
42-44	798	Mean	533.66	576.94	617.28	83.62
		SD	70.39	72.18	79.71	55.81
		25th	493	529	560	48
		50th	529	573	615	78
		75th	573	615	663	113
		SEM	13	14	16	
45-47	549	Mean	544.76	588.69	625.59	80.83
		SD	67.18	68.81	76.58	51.89
		25th	505	541	573	48
		50th	541	581	615	74
		75th	588	635	679	109
		SEM	13	15	16	
48-50	792	Mean	560.96	612.11	660.74	99.78
		SD	66.77	69.89	74.67	56.35
		25th	517	567	615	62
		50th	560	615	663	95
		75th	606	663	711	134
		SEM	13	16	15	
51-53	826	Mean	566.91	620.05	674.61	107.71
		SD	64.08	64.55	72.38	60.81
		25th	529	579	625	68
		50th	573	625	679	104
		75th	606	663	720	144
		SEM	13	16	15	
54-56	773	Mean	574.43	630.80	686.17	111.74
		SD	64.57	66.26	70.34	61.12
		25th	535	588	645	68
		50th	573	635	695	108
		75th	615	671	730	148
		SEM	14	16	14	
57-59	609	Mean	580.16	632.05	689.83	109.67
		SD	67.49	67.68	74.27	63.18
		25th	535	588	645	68
		50th	588	635	695	105
		75th	625	679	742	149
		SEM	14	16	14	
60-62	1022	Mean	599.51	647.99	695.05	95.59
		SD	80.06	90.03	96.69	75.05
		25th	547	615	663	57
		50th	606	663	711	99
		75th	663	703	755	138
		SEM	15	16	15	
63-65	752	Mean	609.41	658.98	708.79	104.78
		SD	81.22	87.74	93.81	66.50
		25th	554	615	671	68
		50th	615	671	730	103
		75th	671	711	794	140
		SEM	16	15	15	
66-68	637	Mean	615.60	669.35	722.03	103.64
		SD	80.83	77.91	82.62	63.56
		25th	567	625	687	66
		50th	625	679	742	105
		75th	671	711	794	140
		SEM	16	15	16	
69-71	582	Mean	629.75	673.82	723.20	101.45
		SD	77.53	91.21	95.35	74.40
		25th	581	645	695	55
		50th	635	687	742	105
		75th	681	720	794	140
		SEM	16	15	16	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 13
Mathematics Scale Score norms by 3 month age band - ages 0-35 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
0-2	1244	Mean	352.57	368.32	388.44	35.87
		SD	28.32	41.98	53.31	46.80
		25th	343	343	343	0
		50th	343	343	343	0
		75th	343	381	441	76
		SEM	55	31	31	
3-5	823	Mean	351.72	365.03	392.42	40.71
		SD	27.06	40.78	53.15	47.77
		25th	343	343	343	0
		50th	343	343	381	37
		75th	343	381	441	79
		SEM	55	31	23	
6-8	598	Mean	351.09	373.03	411.15	60.06
		SD	27.35	42.19	50.22	46.47
		25th	343	343	343	0
		50th	343	343	419	61
		75th	343	404	450	98
		SEM	55	31	23	
9-11	335	Mean	359.92	395.34	424.84	64.93
		SD	34.82	45.94	47.39	48.61
		25th	343	343	404	23
		50th	343	404	441	61
		75th	343	441	459	107
		SEM	55	23	20	
12-14	931	Mean	415.27	446.84	473.54	58.28
		SD	50.85	47.44	45.79	26.00
		25th	381	418	450	26
		50th	419	450	475	52
		75th	450	475	504	88
		SEM	20	17	15	
15-17	786	Mean	426.41	457.87	484.59	58.18
		SD	49.62	46.09	44.68	43.60
		25th	381	441	459	30
		50th	436	459	490	53
		75th	459	490	511	86
		SEM	20	16	15	
18-20	718	Mean	438.31	469.22	496.75	58.44
		SD	44.10	40.88	40.60	41.89
		25th	419	450	475	30
		50th	450	475	497	54
		75th	467	490	524	84
		SEM	17	15	15	
21-23	565	Mean	449.17	479.32	506.22	57.05
		SD	48.25	47.70	44.99	41.64
		25th	425	459	483	30
		50th	459	490	511	52
		75th	483	511	536	83
		SEM	16	15	14	
24-26	1095	Mean	485.46	510.66	532.15	46.69
		SD	45.52	41.76	40.97	36.67
		25th	459	490	511	22
		50th	490	511	536	43
		75th	511	536	557	65
		SEM	15	14	13	
27-29	928	Mean	488.42	516.26	538.81	50.39
		SD	43.80	38.76	39.26	39.72
		25th	467	490	518	26
		50th	490	518	541	46
		75th	519	545	562	69
		SEM	15	14	13	
30-32	700	Mean	502.40	527.55	550.69	48.29
		SD	41.67	43.86	40.87	36.65
		25th	467	490	518	26
		50th	490	518	541	46
		75th	518	545	562	69
		SEM	14	13	12	
33-35	377	Mean	505.01	533.71	558.56	53.55
		SD	44.49	41.02	39.55	37.95
		25th	483	511	538	29
		50th	504	536	557	53
		75th	533	562	581	75
		SEM	14	13	12	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 14
Mathematics Scale Score norms by 3 month age band - ages 36-71 months

Age Band	n		Fall	Winter	Spring	Fall to Spring Growth
36-38	892	Mean	525.55	558.99	583.71	58.16
		SD	55.13	51.29	53.38	35.69
		25th	490	530	552	35
		50th	530	557	583	54
		75th	557	589	610	76
		SEM	13	12	13	
39-41	761	Mean	534.50	569.36	596.28	61.78
		SD	50.49	47.58	50.98	37.20
		25th	504	541	567	37
		50th	541	567	594	58
		75th	562	599	626	81
		SEM	13	12	13	
42-44	798	Mean	541.34	577.75	605.81	64.47
		SD	51.28	48.82	53.12	40.05
		25th	511	551	573	42
		50th	541	573	599	60
		75th	567	599	637	84
		SEM	13	13	12	
45-47	549	Mean	551.13	586.87	613.68	62.55
		SD	50.46	47.65	50.70	37.36
		25th	524	557	578	38
		50th	552	583	610	59
		75th	578	615	643	84
		SEM	12	13	12	
48-50	792	Mean	566.12	607.91	642.33	76.22
		SD	50.87	46.72	50.49	42.06
		25th	536	578	610	48
		50th	570	610	646	70
		75th	599	637	673	99
		SEM	12	12	13	
51-53	826	Mean	569.78	613.78	650.91	81.14
		SD	47.28	44.04	48.70	41.50
		25th	546	589	621	54
		50th	573	615	654	76
		75th	599	643	685	102
		SEM	12	12	13	
54-56	773	Mean	578.93	622.33	659.91	80.99
		SD	44.55	41.58	47.15	43.85
		25th	552	594	631	53
		50th	578	621	661	76
		75th	605	648	690	102
		SEM	13	13	13	
57-59	609	Mean	580.90	625.77	663.32	82.42
		SD	51.02	47.34	51.33	43.88
		25th	552	599	631	53
		50th	583	626	667	78
		75th	615	654	696	107
		SEM	13	13	13	
60-62	963	Mean	605.64	654.29	698.58	91.88
		SD	56.87	50.39	50.41	45.90
		25th	573	626	673	59
		50th	610	661	707	87
		75th	643	690	732	
		SEM	12	13	13	
63-65	688	Mean	610.82	660.42	708.95	98.37
		SD	62.10	58.26	59.12	45.13
		25th	578	637	685	69
		50th	615	670	719	97
		75th	654	696	749	127
		SEM	12	13	13	
66-68	568	Mean	617.98	671.30	717.03	94.84
		SD	58.71	50.02	50.07	47.77
		25th	585	50	50	48
		50th	621	679	725	88
		75th	654	701	749	127
		SEM	12	13	14	
69-71	529	Mean	624.56	673.03	724.53	102.51
		SD	51.53	48.32	43.20	47.80
		25th	594	651	701	65
		50th	626	679	732	100
		75th	661	707	760	136
		SEM	13	13	14	

Note. 25th, 50th, and 75th refer to percentiles within the three month age band.
SEM refers to the standard error of measurement.

Table 15

Percentages of children meeting expectations by time of year - Birth to 1 year

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	8.10%	1.50%	0.30%
	Meets Expectations	73.00%	52.80%	28.60%
	Exceeds Expectations	18.90%	45.70%	71.10%
Literacy	Below Expectations	0.00%	0.00%	0.00%
	Meets Expectations	84.50%	63.10%	39.20%
	Exceeds Expectations	15.50%	36.90%	60.80%
Social Emotional	Below Expectations	7.50%	1.30%	0.30%
	Meets Expectations	75.30%	55.00%	29.50%
	Exceeds Expectations	17.20%	43.70%	70.20%
Physical	Below Expectations	16.50%	2.80%	0.20%
	Meets Expectations	71.10%	60.20%	36.30%
	Exceeds Expectations	12.40%	37.00%	63.50%
Language	Below Expectations	16.60%	4.40%	1.30%
	Meets Expectations	79.90%	78.20%	57.60%
	Exceeds Expectations	3.50%	17.40%	41.20%
Mathematics	Below Expectations	0.00%	0.00%	0.00%
	Meets Expectations	86.70%	64.20%	40.90%
	Exceeds Expectations	13.30%	35.80%	59.10%

Table 16

Percentages of children meeting expectations by time of year - 1 to 2 years

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	13.20%	3.50%	1.10%
	Meets Expectations	75.60%	65.40%	44.90%
	Exceeds Expectations	11.20%	31.10%	54.00%
Literacy	Below Expectations	25.10%	9.10%	3.30%
	Meets Expectations	46.60%	40.00%	25.00%
	Exceeds Expectations	28.40%	50.90%	71.70%
Social Emotional	Below Expectations	17.70%	5.00%	1.80%
	Meets Expectations	74.50%	70.90%	53.50%
	Exceeds Expectations	7.80%	24.10%	44.70%
Physical	Below Expectations	22.90%	8.10%	2.80%
	Meets Expectations	57.50%	52.80%	37.30%
	Exceeds Expectations	19.60%	39.10%	59.90%
Language	Below Expectations	36.90%	14.40%	5.40%
	Meets Expectations	58.50%	71.20%	59.70%
	Exceeds Expectations	4.50%	14.40%	34.80%
Mathematics	Below Expectations	54.30%	29.70%	13.50%
	Meets Expectations	40.20%	53.70%	47.40%
	Exceeds Expectations	5.50%	16.60%	39.10%

Table 17

Percentages of children meeting expectations by time of year - 2 to 3 years

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	31.80%	12.60%	5.40%
	Meets Expectations	56.90%	61.80%	48.10%
	Exceeds Expectations	11.30%	25.60%	46.40%
Literacy	Below Expectations	36.60%	17.50%	7.60%
	Meets Expectations	51.80%	55.60%	47.50%
	Exceeds Expectations	11.60%	26.90%	44.90%
Social Emotional	Below Expectations	27.60%	10.70%	4.90%
	Meets Expectations	61.50%	66.40%	53.10%
	Exceeds Expectations	10.90%	22.90%	42.00%
Physical	Below Expectations	20.30%	9.00%	4.00%
	Meets Expectations	66.50%	63.80%	50.50%
	Exceeds Expectations	13.20%	27.30%	45.50%
Language	Below Expectations	33.60%	17.00%	8.10%
	Meets Expectations	56.20%	62.10%	55.50%
	Exceeds Expectations	10.20%	20.90%	36.50%
Mathematics	Below Expectations	38.90%	18.30%	7.80%
	Meets Expectations	56.60%	67.80%	61.80%
	Exceeds Expectations	4.50%	13.90%	30.40%

Table 18
Percentages of children meeting expectations by time of year - Preschool 3

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	44.80%	17.80%	7.50%
	Meets Expectations	49.00%	65.20%	57.30%
	Exceeds Expectations	6.20%	17.00%	35.20%
Literacy	Below Expectations	49.70%	19.30%	8.70%
	Meets Expectations	41.60%	58.40%	49.60%
	Exceeds Expectations	8.70%	22.30%	41.70%
Social Emotional	Below Expectations	42.30%	16.00%	7.40%
	Meets Expectations	48.80%	61.80%	50.60%
	Exceeds Expectations	8.90%	22.20%	42.00%
Physical	Below Expectations	37.20%	14.70%	6.40%
	Meets Expectations	56.90%	69.80%	60.80%
	Exceeds Expectations	5.90%	15.60%	32.80%
Language	Below Expectations	47.40%	24.30%	12.20%
	Meets Expectations	47.40%	61.40%	58.80%
	Exceeds Expectations	5.20%	14.40%	29.10%
Mathematics	Below Expectations	53.50%	25.00%	11.70%
	Meets Expectations	40.10%	58.20%	53.30%
	Exceeds Expectations	6.40%	16.80%	35.00%

Table 19

Percentages of children meeting expectations by time of year - Preschool 4

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	54.80%	19.40%	5.90%
	Meets Expectations	44.10%	73.70%	64.30%
	Exceeds Expectations	1.10%	6.90%	29.80%
Literacy	Below Expectations	44.90%	11.50%	3.80%
	Meets Expectations	53.90%	80.90%	68.40%
	Exceeds Expectations	1.20%	7.70%	27.80%
Social Emotional	Below Expectations	57.80%	22.50%	7.70%
	Meets Expectations	39.70%	66.80%	53.30%
	Exceeds Expectations	2.50%	10.70%	39.00%
Physical	Below Expectations	50.60%	19.30%	6.50%
	Meets Expectations	48.80%	77.00%	73.40%
	Exceeds Expectations	0.70%	3.70%	20.20%
Language	Below Expectations	52.10%	24.20%	9.50%
	Meets Expectations	46.70%	70.60%	65.40%
	Exceeds Expectations	1.20%	5.20%	25.10%
Mathematics	Below Expectations	72.50%	34.50%	14.50%
	Meets Expectations	26.80%	60.80%	61.60%
	Exceeds Expectations	0.70%	4.70%	23.90%

Table 20

Percentages of children meeting expectations by time of year - Kindergaten

Scale	Expectations	Fall	Winter	Spring
Cognitive	Below Expectations	77.70%	41.20%	19.50%
	Meets Expectations	22.30%	58.50%	77.80%
	Exceeds Expectations	0.10%	0.40%	2.70%
Literacy	Below Expectations	57.70%	16.50%	6.50%
	Meets Expectations	42.10%	81.00%	77.70%
	Exceeds Expectations	0.20%	2.50%	15.90%
Social Emotional	Below Expectations	69.30%	34.90%	15.60%
	Meets Expectations	30.70%	64.70%	78.20%
	Exceeds Expectations	0.00%	0.40%	6.20%
Physical	Below Expectations	60.90%	31.20%	13.10%
	Meets Expectations	38.80%	67.70%	78.00%
	Exceeds Expectations	0.30%	1.10%	8.90%
Language	Below Expectations	78.80%	46.40%	22.10%
	Meets Expectations	20.80%	52.60%	71.50%
	Exceeds Expectations	0.40%	1.00%	6.50%
Mathematics	Below Expectations	87.10%	45.30%	19.10%
	Meets Expectations	12.70%	54.20%	75.20%
	Exceeds Expectations	0.20%	0.50%	5.70%