CEME Technical Report

CEMETR-2010-06
DECEMBER 2010

The Center for Educational Measurement and Evaluation

Technical Manual for the *Teaching Strategies GOLD*™ Assessment System

Richard G. Lambert
Do-Hong Kim
Heather Taylor
Jennifer R. McGee

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

Richard G. Lambert
Do-Hong Kim
Heather Taylor
Jennifer R. McGee
Center for Educational Measurement and Evaluation
UNC Charlotte
December 2010

Norm Sample

A total population of 111,059 children was rated using the *Teaching Strategies GOLD™* assessment system for the fall, 2010 checkpoint. These children received educational services in 735 different programs at 3,792 different centers that were located in all regions of the United States. These programs and centers included Head Start, private childcare, and school-based sites. All fifty states and the District of Columbia were represented. Most of the participating centers use The *Creative Curriculum*® and had been using *The Creative Curriculum Developmental Continuum for Ages 3-5* assessment system prior to this study. A total of 8,042 different raters (teachers) provided the ratings. Each teacher received training in the use of the *Teaching Strategies GOLD™* assessment system and rated an average of 13.8 children.

The total population of children rated using the *Teaching Strategies GOLD™* assessment system was 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 was October 29, 2010. This date represents the final date for teachers to enter ratings into the online assessment system. 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 population span the entire age range for which the assessment system is intended (birth through kindergarten).

The teachers collected information about the race and ethnicity of each child and entered this information into the online 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. A norm sample was created by sampling from the total population of children rated using the measure. The goal was to represent each of the 24 three month age strata with 500 randomly selected children for a total norm sample of 12,000 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.

For 17 of the 24 age strata, there were sufficient numbers of children in the population to sample 500 children in their proper proportions so as to match the U.S. Census Bureau 2009 estimates with respect to ethnic subgroup. These strata covered ages 15 months to 65 months and represent the ages for which the assessment is most widely used. At the extreme upper and lower ends of the age distribution, 0 to 14 months and 66 to 71 months, there were not enough children in the total population to allow for sampling. Each of these strata contained less than 500 children and all children in the population that

fell into these age strata were retained in the norm sample. Therefore the total norm sample was less than the targeted 12,000 (n=10,963) and the total distribution of the seven ethnic subgroups across the norm sample did not exactly match the national population estimates.

As shown in Table 1, the total population of children rated using the *Teaching Strategies GOLD™* assessment system for the fall, 2010 checkpoint does not exactly match the U.S. Census Bureau estimates for the U.S. population of children under the age of 6. White children are somewhat under-represented and African-American children are somewhat over-represented. In the norm sample the ethnic subgroups were more closely represented in their proper proportions. Hispanic children were very closely matched to national estimates (25.5% Census estimates vs. 26.0% norm sample). White children were slightly under-represented (52.1% Census estimates vs. 48.5% norm sample). African-American children were slightly over-represented (13.6% Census estimates vs. 16.3% norm sample). Children of all other ethnic subgroups were closely represented in their overall proper proportion (8.9% Census estimates vs. 9.2% 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.

A total of 10,963 children were retained in the norm sample. These children received educational services in 618 different programs at 2,525 different centers that were located in all regions of the United States. These programs and centers included Head Start, private childcare, and school-based sites. Forty eight states and the District of Columbia were represented. A total of 4,580 different raters (teachers) provided the ratings.

Exploratory Factor Analysis

The first step in demonstrating evidence for the reliability and validity of the information yielded by the use of the *Teaching Strategies GOLD™* assessment system was to conduct an exploratory factor analysis. The full norm sample of 10,963 was randomly divided into two samples of approximately 5,500 children. Half of the norm sample was used for an exploratory factor analysis and the remaining half was used for a confirmatory factor analysis. The exploratory factor analysis was conducted using principal axis factoring and oblimin rotation with Kaiser Normalization. These methods were chosen because the theoretical domains of child development upon which the measure is based, are expected to correlate with each other as operationalized into a series of teacher ratings.

A five factor solution emerged that accounted for 71.99% of the variance in the ratings. The five factor solution very closely matched the theoretical domains of child developmental that the test developers intended. Simple structure was clearly achieved as no item loaded on more than one factor. All of the items from the Social-Emotional domain (1a, 1b, 1c, 2a, 2b, 2c, 2d, 3a, and 3b) loaded on the same factor with factor loadings that ranged from .429 to .846. The highest loading was for item 3a. This item focuses on children cooperating in group situations by balancing their needs with the needs of others and helps represents the overall construct of social development for young children.

The second factor replicated the Physical domain of development and contained all of the items from this section of the measure (4, 5, 6, 7a, and 7b) with loadings that ranged from .507 to .762. The highest loading was for item 5 which focuses on balancing skills. The third factor contained all of the items from the Language section of the measure (8a, 8b, 9a, 9b, 9c, 9d, 10a, and 10b) with loadings that ranged from .398 to .894. The highest

loading was for item 9c which focuses on using spoken language to express thoughts and needs with a particular focus on understandable grammar. The fourth factor contained all of the items from the Cognitive domain of development (11a, 11b, 11d, 11e, 12a, 12b, 13, 14a, and 14b) with loadings that ranged from .301 to .456. The highest loading was for 11e which focuses on demonstrating positive approaches to learning by showing flexibility and inventiveness. The final factor contained all of the items from both the Literacy and Mathematics domains (15a, 15b,15c, 16a, 16b, 17a, 17b, 18a, 18b, 18c, 19a, 19b, 20a, 20b, 20c, 21a, 21b, 22, and 23) with loadings that ranged from .527 to .952. The highest loading was for item 16b which focuses on alphabet knowledge and letter-sound knowledge.

It is not clear why all of these items loaded together rather than on separate factors. However, it is important to note that while the ratings for all of the items on the other factors (1a to 14b) yielded approximately normal distributions of ratings on the 0 to 9 scale, the items on these two sections of the measure (15a to 23) consistently yielded positively skewed distributions. Since these developmental milestones generally emerge in children toward the upper end of the intended age range for the measure, teachers naturally were giving ratings at the lowest end of the rating scale for most of the very youngest children in the norm sample, thus creating very skewed distributions for these items. It is therefore possible that natural maturation and the fact that these items are more appropriate for children toward the upper end of the intended age range for the measure may be impacting the inter-item correlations.

Confirmatory Factor Analysis

The factorial structure of the *Teaching Strategies GOLD™* was examined using confirmatory factor analysis (CFA) with full-information maximum likelihood (FIML) estimation in Mplus (Muthén & Muthén, 2008). A six-factor model at the item level that corresponds to the designed structure of the instrument was examined. The chi-square test can be used to evaluate model fit. However, given the sensitivity of this test to sample sizes, alternative goodness-of-fit indices were used to evaluate model fit including Standardized Root Mean Square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI). RMSEA values ≤ .05 are considered a good fit, values between .05 and .08 indicates reasonable fit and values ≥ .10 are unacceptable (Browne & Cudeck, 1993). More recently, RMSEA values < 0.06 and SRMR values < 0.08 were recommended as a good fit (Hu & Bentler, 1999). Generally, a CFI value of at least .90 is required to accept a model (Hu & Bentler, 1998) and more recently, a more stringent criterion of CFI values ≥ .95 has been recommended (Hu & Bentler, 1999).

Results of the CFA showed a significant chi-square statistics ($\chi^2_{1209} = 29683.91, p < .01$), which is not surprising given this study's large sample size. The six-factor model fit the data reasonably well, as evidenced by SRMR=0.033, CFI= 0.931, and RMSEA= .066. The standardized factor loadings for this model are provided in Tables 5 and 6. All factor loadings were generally large and statistically significant at p< .001. The correlations between the six scales were also large (values ranged from 0.786 to 0.960) and statistically significant at p< .001.

The overall results supported the first-order six-factor model. However, the high correlations between the scales may indicate either the presence of a higher-order factor structure or a single first-order factor structure. Both the higher-order factor and the single

factor model were tested, but did not improve the fit of the model and produced even worse fit (For the higher-order factor model, SRMR=0.041, CFI= 0.923, and RMSEA= .070; for the single factor model, SRMR=0.046, CFI= 0.828, and RMSEA= .104) as compared with the first-order factor model with the six factors. Based on these results, the data best fits the current first-order six-factor model.

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, 2009). A separate Rasch analysis was conducted for each of the six domains of development identified in the factor analysis.

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, 2009).

Social Emotional Scale (9 items)

The principal component analysis of the residuals showed that for the Social Emotional scale, the Rasch dimension explained 82.9% of the variance in the data, with its eigenvalue of 43.6. The first contrast (the largest secondary dimension) had an eigenvalue of 2.2 and accounted only for 4% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Physical Scale (5 items)

The principal component analysis of the residual showed that for the Physical scale, the Rasch dimension explained 86.0% of the variance in the data, with its eigenvalue of 30.7. The first contrast (the largest secondary dimension) had an eigenvalue of 1.8 and accounted only for 5% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Language Scale (8 items)

The principal component analysis of the residual showed that for the Language scale, the Rasch dimension explained 88.4% of the variance in the data, with its eigenvalue of 61.2. The first contrast (the largest secondary dimension) had an eigenvalue of 1.6 and accounted only for 2% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Cognitive Scale (10 items)

The principal component analysis of the residual showed that for the Cognitive scale, the Rasch dimension explained 85.5% of the variance in the data, with its eigenvalue of 58.9. The first contrast (the largest secondary dimension) had an eigenvalue of 2.1 and accounted only for 3% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Literacy Scale (12 items)

The Rasch dimension explained 79.9% of the variance in the data, with its eigenvalue of 47.7. The first contrast (the largest secondary dimension) had an eigenvalue of 2.2 and accounted for 3.6% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Mathematics Scale (7 items)

The Rasch dimension explained 82.2% of the variance in the data, with its eigenvalue of 32.4. The first contrast (the largest secondary dimension) had an eigenvalue of 1.7 and accounted for 4.2% of the unexplained variance. These results indicated that the data satisfied the unidimensionality assumption of the Rasch model.

Model Fit

The fit statistics for all of the Social Emotional items were well within acceptable limits: the infit MNSQ ranged from 0.81 to 1.27; the outfit MNSQ ranged from 0.78 to 1.28. The fit statistics for all of the Physical items were well within acceptable limits: the infit MNSQ ranged from 0.89 to 1.21; the outfit MNSQ ranged from 0.89 to 1.30. The fit statistics for all of the Language items were well within acceptable limits: the infit MNSQ ranged from 0.81 to 1.27; the outfit MNSQ ranged from 0.78 to 1.28. The fit statistics for all of the Cognitive items were well within acceptable limits: the infit MNSQ ranged from 0.84 to 1.27; the outfit MNSQ ranged from 0.83 to 1.25. All Literacy items except one (item 16a) exhibited good fit to the unidimensional Rasch model: the infit MNSQ ranged from 0.69 to 1.33; the outfit MNSQ ranged from 0.62 to 1.31. Item 16a slightly beyond the 1.4 cutoff (infit MNSQ = 1.79; outfit MNSQ = 1.40). All Mathematics items except one (item 20c) exhibited good fit to the unidimensional Rasch model: the infit MNSQ ranged from 0.67 to

1.14; the outfit MNSQ ranged from 0.67 to 1.10. Item 20c slightly beyond the 1.4 cutoff (infit MNSQ = 1.57; outfit MNSQ = 1.36). With very few exceptions, these model fit statistics suggest that the data does in fact fit the Rasch rating scale model very well. Acceptable model fit indexes are also an indication that the assumption of unidimensionality within each construct is met.

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 the Social Emotional Scale, the average measure increased with the category level and the thresholds advanced with the categories. For the Physical Scale, the average measure increased with the category level and the thresholds advanced with the categories. For the Language Scale, the average measure generally increased with the category level and the thresholds advanced with the categories. However, there was substantial overlap

between category 0 and category 1. It is possible that a rating of 1 may not have been quite distinct and seemed somewhat redundant with the 0 category for the raters. For the Cognitive Scale, the average measure increased with the category level. The thresholds advanced with the categories. For the Literacy Scale, the average measure increased with the category level. The thresholds advanced with the categories, except between category 7 and category 8. An examination of the Rasch category probability curves indicated that category 7 may not have been quite distinct and seemed somewhat redundant with adjacent categories. For the Mathematics Scale, the average measure increased with the category level except between category 8 and category 9. The thresholds advanced with the categories.

Item Difficulty Measures

Social Emotional Scale

The item location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, the item pertaining to a child's ability to solve social problems (3b) was found to be the most difficult item, whereas the item pertaining to a child's ability to respond to emotional cues (2b) was estimated as the easiest item.

Physical Scale

The item location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, 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 (4a) 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 location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, the item pertaining to a child's ability to tell about another time or place (9d) was found to be the most difficult item, whereas the item pertaining to a child's ability to ability to comprehend language (8a) was estimated as the easiest item.

Cognitive Scale

The item location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, 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 location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, 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 location hierarchy appeared to be consistent with the expected developmental trajectory for typically developing children. For example, 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 4.88, person reliabilities of .96, item separation indexes of 56.39, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .969, 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 4.32, person reliabilities of .95, item separation indexes of 38.59, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .957, 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 5.76, person reliabilities of .97, item separation indexes of 49.70, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .977, 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 6.65, person reliabilities of .98, item separation indexes of 50.79, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .980, 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 4.75, person reliabilities of .96, item separation indexes of 57.36, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .977, 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 4.64, person reliabilities of .96, item separation indexes of 43.41, and item reliabilities of .99. The Cronbach's alpha reliability coefficient for this scale was .972, indicating high internal consistency reliability.

Inter-Rater Reliability

An inter-rater reliability study was conducted by examining the correlations between the rating of a master trainer and the ratings of teachers who are current users of the system. This study was conducted by first having a master trainer rate 18 children on all items contained in the measure. These children ranged in age from 5 to 71 months of age. Next, a sample of 557 teachers examined video tapes of these same children and provided their ratings across all items in the assessment system. These teachers provided 2,558 separate child assessments and rated an average of 4.59 children (minimum 1,

maximum 9). Each teacher rated only those children who matched the age group with they work and thus no teacher rated all 18 children.

Raw scores were created for both the master trainer and the participating teachers by taking the sum of the ratings across the items for each of the 6 developmental domains. The correlations between the teacher ratings and the ratings by the master trainer were all high. All were above .80 and all but one were above .90 (see table 8). The highest level of agreement was found for the Literacy section of the measure (r=.939). Given that reliability is an extremely important concern when considering the value of teacher ratings, these values are very encouraging.

Future analyses will focus on variance decomposition whereby the percentage of variance in the ratings that can be attributed to child age, within teacher variability, and between teacher variability will be examined. These results will provide another important indicator of the magnitude of any potential teacher or rater effects. In addition, future analyses will examine the percentage of agreement between teachers and the master trainer at the item level. It will be important to examine extent to which teachers can agree with a master trainer at not just the level of the scale score, but also at the level of the individual item by item ratings. These item level data will be used by teachers for instructional planning, communicating with parents, and generally developing a greater understanding of the developmental progress of individual children. Therefore high levels of agreement for each item will also be important to demonstrate.

Scale Scores and Norm Tables

The totality of the evidence from the factor analysis and Rasch modeling suggested 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.

The children were separated into three month age bands based on their age in months at the time of the first assessment of the academic year. The data contained in this report was collected in October of 2010 and was used as the fall assessment by the participating teachers, schools, and programs. For each scale score and three month age band the tables include the mean, standard deviation, and quartile boundaries for both the raw scores and the scale scores. The raw scores are simply the sum of the item scores across the 0 to 9 rating scales for each developmental domain. Tables are included for all three month age bands from 6 months to 71 months of age. The youngest two age bands, 0-2 months and 3-5 months, are not included in the norm tables due to insufficient numbers of children. An attempt will be made to gather more data on children in these age bands for the winter and spring assessments in order to include them in subsequent reports.

Table 9 includes the distributional characteristics for the overall norm sample across age bands for each scale score. The means for each scale score are generally occurring most commonly for children around 36 months of age which is the middle of the age range for which the measure is intended. Correlations with age in months are also included. The scale scores correlate moderately strongly with age (r = .673 - .731). 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.

As can be seen in both the raw score and scale score norm tables (Tables 10-21) 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.

Given the limited amount of data for the very youngest children (0-5 months), teachers may choose to use the scale scores with caution for these children. Item level data may be more important for instructional planning for these children. Future analyses will continue to examine this issue as more data becomes available. Future analyses will also focus on including winter and spring norms.

Summary

Overall, the *Teaching Strategies GOLD™* assessment system appears to be yield highly reliable scores as indicated by both the classical and Rasch reliability statistics. Results of the factor analyses showed that the ratings loaded onto the constructs generally as intended by the test 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 teacher's are able to make valid

ratings of the developmental progress of children across the intended age range, from birth through 72 months.

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.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), Testing structural Equation Models (pp. 136-162). Newbury Park, CA : Sage.

Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4): 424-453.

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

Linacre, J. M. (2009). Winsteps (Version 3.68.0) [Computer Software]. Chicago, IL: Winsteps.com.

Muthén, L. K., & Muthén, B. O. (2008). Mplus version 5.1 [Computer software], Los Angeles, CA: Muthén & Muthén.

Table 1
Norm Sample by Ethnic Subgroup.

Racial and Ethnic Subgroup	GOLD Population	2009 Census Bureau Estimates	Norm Sample
White, not Hispanic	43.2%	52.1%	48.5%
African American, not Hispanic	21.7%	13.6%	16.3%
Native American, not Hispanic	1.2%	0.9%	1.5%
Asian, not Hispanic	1.9%	4.5%	3.2%
Hawaiian or Pacific Islander, not Hispanic	0.4%	0.2%	0.2%
Multirace, not Hispanic	4.0%	3.2%	4.2%
Hispanic	27.7%	25.5%	26.0%

Note. GOLD population n=111,059; US population n=25,485,229; norm sample n = 10,963.

Table 2
Frequency by Three Month Age Band.

Months	Frequency	Percent
0-2	225	2.1%
3-5	371	3.4%
6-8	417	3.8%
9-11	391	3.6%
12-14	484	4.4%
15-17	500	4.6%
18-20	500	4.6%
21-23	500	4.6%
24-26	500	4.6%
27-29	500	4.6%
30-32	500	4.6%
33-35	500	4.6%
36-38	500	4.6%
39-41	500	4.6%
42-44	500	4.6%
45-47	500	4.6%
48-50	500	4.6%
51-53	500	4.6%
54-56	500	4.6%
57-59	500	4.6%
60-62	500	4.6%
63-65	500	4.6%
66-68	330	3.0%
69-71	245	2.2%
Total	10963	100.0%

Table 3
Norm Sample by Census Region and State

Region	State	n	Percent	Region	State	n	Percent
Northeast and DC	CT	78	18.8%	South	AL	101	3.9%
	MA	91	22.0%		AR	14	0.5%
	ME	1	2.0%		DE	6	0.2%
	NH	4	1.0%		FL	527	20.2%
	NJ	53	12.8%		KY	70	2.7%
	NY	102	24.6%		LA	36	1.4%
	PA	49	11.8%		MD	3	0.1%
	RI	27	6.5%		MS	14	0.5%
	VT	9	2.2%		NC	208	8.0%
	DC	503	4.6%		OK	491	18.8%
	Total	917	8.4%		SC	27	1.0%
					TN	160	6.1%
					TX	187	7.2%
					VA	45	1.7%
					WV	716	27.5%
					Total	2605	23.8%
Midwest	IA	329	6.6%	West	AK	61	2.5%
	IL	2378	47.9%		AZ	137	5.5%
	IN	309	6.2%		CA	124	5.0%
	KS	272	5.5%		CO	1140	46.0%
	MI	402	8.1%		HI	113	4.6%
	MN	5	0.1%		ID	8	0.3%
	MO	189	3.8%		NM	20	0.8%
	ND	74	1.5%		NV	52	2.1%
	NE	361	7.3%		OR	384	15.5%
	OH	446	9.0%		SD	30	1.2%
	WI	196	4.0%		UT	10	0.4%
	Total	4961	45.3%		WA	360	14.5%
					WY	41	1.7%
					Total	2480	22.6%

Table 4
Individual Characteristics of Participants

Characteristic		Frequency	Percent
Sex	Female	5231	47.7%
	Male	5706	52.0%
Primary Language	English	8606	48.5%
	Spanish	1750	16.0%
	Other	604	5.5%
Disability Status	IFSP	323	2.9%
	IEP	902	8.2%
	Typically Developing	9738	88.8%

Note. IFSP = Individual Family Service Plan,
IEP = Individual Education Program.

Table 5
Model Estimates for Relationships Between Latent Consturcts and Observed Variables

Domain	Item	Estimate
Social-Emotional	1. Regulates own emotions and behaviors	
	a. Manages Feelings	0.835
	b. Follows limits and expectations	0.884
	c. Takes care of own needs appropriately	0.881
	2. Establishes and sustains positive relationships	
	a. Forms relationships with adults	0.838
	b. Responds to emotional cues	0.895
	c. Interacts with peers	0.901
	d. Makes friends	0.896
	3. Participates cooperatively and constructively in group situations	
	a. Balances needs and rights of self and others	0.920
	b. Solves social problems	0.901
Physical	4. Demonstrates traveling skills	0.916
	5. Demonstrates balancing skills	0.918
	6. Demonstrates gross motor manipulative skills	0.907
	7. Demonstrates fine-motor coordination	
	a. Uses fingers and hands	0.913
	b. Uses writing and drawing tools	0.903
Language	8. Listens to and understands increasingly complex language	
	a. Comprehends language	0.914
	b. Follows directions	0.895
	9. Uses language to express thoughts and needs	
	a. Uses an expanding expressive vocabulary	0.928
	b. Speaks clearly	0.916
	c. Uses conventional grammar	0.941
	d. Tells about another time or place	0.919
	10. Uses appropriate conversational and other communication skills	
	a. Engages in conversations	0.925
	b. Uses social rules of language	0.929

Table 6
Model Estimates for Relationships Between Latent Consturcts and Observed Variables

Domain	Item	Estimate
Cognitive	11. Demonstrates positive approaches to learning	
	a. Attends and engages	0.895
	b. Persists	0.887
	c. Solves problems	0.912
	d. Shows curiosity and motivation	0.898
	e. Shows flexibility and inventiveness in thinking	0.931
	12. Remembers and connects experiences	
	a. Recognizes and recalls	0.918
	b. Makes connections	0.933
	13. Uses classification skills	0.912
	14. Uses symbols and images to represent something not present	
	a. Thinks symbolically	0.928
	b. Engages in sociodramatic play	0.914
Literacy	15. Demonstrates phonological awareness	
	a. Notices and discriminates rhyme	0.867
	b. Notices and discriminates alliteration	0.898
	c. Notices and discriminates smaller and smaller units of sound	0.856
	16. Demonstrates knowledge of the alphabet	
	a. Identifies and names letters	0.852
	b. Uses letter-sound knowledge	0.831
	17. Demonstrates knowledge of print and its uses	
	a. Uses and appreciates books	0.850
	b. Uses print concepts	0.934
	18. Comprehends and responds to books and other texts	
	a. Interacts during read-alouds and book conversations	0.929
	b. Uses emergent reading skills	0.932
	c. Retells stories	0.927
	19. Demonstrates emergent writing skills	
	a. Writes name	0.863
	b. Writes to convey meaning	0.868
Mathematics	20. Uses number concepts and operations	
	a. Counts	0.930
	b. Quantifies	0.935

Table 7
Model Estimates of the Relationships Between Latent Constructs

Domain 1	Domain 2	*Coefficient*
Physical	Social-Emotional	0.904
Language	Social-Emotional	0.927
	Physical	0.904
Cognitive	Social-Emotional	0.951
	Physical	0.916
	Language	0.960
Literacy	Social-Emotional	0.866
	Physical	0.786
	Language	0.867
	Cognitive	0.897
Mathematics	Social-Emotional	0.880
	Physical	0.828
	Language	0.900
	Cognitive	0.919
	Literacy	0.959

Table 8
Inter-rater Reliability Between Master Teachers and Trainees

Domain	*r*
Social-Emotional	0.902
Physical	0.859
Language	0.931
Cognitive	0.922
Literacy	0.939
Mathematics	0.920

Note. All values are significant at the .001 level.

Table 9
Overall Characteristics of Each Distribution of Scale Score Across Age Bands

Domain	*Mean*	*SD*	*Percentiles* *25th*	*50th*	*75th*	Correlation with Age
Social-Emotional	499.520	98.43	432	505	570	0.673
Physical	500.530	99.42	443	512	578	0.686
Language	501.460	99.98	429	505	573	0.685
Cognitive	498.804	99.57	424	494	571	0.695
Literacy	499.330	100.47	428	502	572	0.731
Mathematics	497.120	99.88	432	506	570	0.720

Table 10

Scale Score Norm Tables for Each Developmental Domain by Age Band (6-23 months)

		Age in Months					
Domain	Statistic	6-8	9-11	12-14	15-17	18-20	21-23
Social-Emotional	*n*	392	363	449	465	468	452
	M	396.24	389.56	404.64	424.68	433.48	450.43
	SD	109.20	82.92	69.56	55.74	49.93	50.20
	25th	325	348	359	396	404	426
	50th	350	369	396	426	432	452
	75th	452	404	426	452	458	476
	SEM	18	18	17	16	16	16
Physical	*n*	399	372	470	485	477	480
	M	380.95	379.06	404.18	431.77	439.13	454.19
	SD	112.38	81.79	69.93	60.09	51.36	54.17
	25th	309	339	366	399	410	432
	50th	339	354	399	421	432	455
	75th	443	389	421	455	466	478
	SEM	18	18	17	18	18	18
Language	*n*	388	358	443	460	455	454
	M	397.80	389.59	395.10	411.72	424.80	440.39
	SD	107.54	84.13	65.40	53.26	47.48	47.30
	25th	334	342	358	381	403	417
	50th	354	366	389	410	423	441
	75th	468	389	410	429	447	469
	SEM	14	14	14	13	13	12

Table 11

Scale Score Norm Tables for Each Developmental Domain by Age Band (6-23 months)

		Age in Months					
Domain	Statistic	6-8	9-11	12-14	15-17	18-20	21-23
Cognitive	*n*	390	354	439	451	452	440
	M	406.26	395.19	400.54	414.75	421.94	433.60
	SD	99.45	83.31	62.99	50.22	43.81	42.37
	25th	343	356	372	387	397	408
	50th	362	372	387	408	419	436
	75th	467	387	408	424	442	453
	SEM	12	12	12	12	13	13
Literacy	*n*	390	358	439	458	454	454
	M	395.46	380.91	392.46	409.58	422.34	432.35
	SD	104.02	90.15	68.17	57.31	46.07	41.85
	25th	296	346	346	375	394	408
	50th	346	346	375	408	419	428
	75th	476	394	408	428	437	451
	SEM	14	30	25	22	20	19
Mathematics	*n*	389	357	437	454	451	435
	M	391.31	368.74	374.71	404.08	420.87	434.40
	SD	101.37	87.81	66.48	63.92	54.70	47.61
	25th	331	331	331	331	394	409
	50th	331	331	331	409	432	441
	75th	459	331	402	441	450	459
	SEM	24	32	32	20	18	17

Table 12
Scale Score Norm Tables for Each Developmental Domain by Age Band (24-47 months)

		Age in Months							
Domain	Statistic	24-26	27-29	30-32	33-35	36-38	39-41	42-44	45-47
Social-Emotional	*n*	456	463	467	472	468	477	474	472
	M	461.24	477.04	485.23	503.25	498.02	502.91	512.61	533.05
	SD	53.18	53.22	54.51	61.50	62.20	66.26	57.52	58.64
	25th	432	452	458	470	464	470	482	499
	50th	464	476	482	505	502	511	516	534
	75th	488	505	511	534	534	546	546	570
	SEM	15	15	15	15	15	15	15	15
Physical	*n*	480	482	480	482	474	477	479	485
	M	468.51	483.99	496.11	510.78	500.29	508.45	519.81	531.60
	SD	56.89	58.99	57.71	58.09	61.36	63.13	54.70	62.40
	25th	443	455	466	478	466	478	490	502
	50th	466	478	490	512	502	512	523	543
	75th	502	512	523	543	533	554	554	565
	SEM	18	18	18	17	18	18	17	17
Language	*n*	465	462	462	468	472	473	479	470
	M	457.99	477.81	489.93	512.61	497.41	505.88	520.88	534.82
	SD	49.54	51.19	54.25	60.37	67.21	63.80	58.29	64.66
	25th	429	452	464	481	464	472	486	499
	50th	464	481	492	511	499	511	524	536
	75th	486	505	517	548	536	548	560	573
	SEM	12	12	13	13	13	13	13	13

Table 13
Scale Score Norm Tables for Each Developmental Domain by Age Band (24-47 months)

		Age in Months							
Domain	Statistic	24-26	27-29	30-32	33-35	36-38	39-41	42-44	45-47
Cognitive	*n*	445	448	441	468	454	456	467	457
	M	447.37	464.96	475.46	496.56	494.26	501.07	516.90	531.71
	SD	46.23	49.56	48.73	58.00	61.52	54.81	57.29	60.05
	25th	422	436	448	459	453	465	482	500
	50th	448	459	471	494	494	506	519	533
	75th	471	488	500	526	526	540	548	571
	SEM	13	13	13	13	13	13	14	14
Literacy	*n*	440	446	445	453	447	451	457	456
	M	446.87	465.48	473.67	490.91	494.77	505.12	520.15	532.44
	SD	47.17	44.24	43.60	54.04	52.67	52.78	51.20	51.08
	25th	419	437	451	464	464	476	486	502
	50th	444	470	476	492	497	502	521	533
	75th	476	492	502	521	530	539	552	565
	SEM	17	15	15	14	14	14	14	14
Mathematics	*n*	433	437	437	451	449	455	461	455
	M	453.29	470.68	480.81	498.37	496.53	508.09	521.23	535.54
	SD	50.10	45.10	46.58	52.93	55.46	55.45	53.24	52.95
	25th	432	450	459	467	467	475	491	506
	50th	459	475	483	499	499	506	527	539
	75th	483	491	506	533	533	545	551	570
	SEM	16	16	16	15	15	15	14	14

Table 14
Scale Score Norm Tables for Each Developmental Domain by Age Band (48-71 months)

		Age in Months							
Domain	Statistic	48-50	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Social-Emotional	*n*	467	474	479	482	480	448	264	182
	M	550.31	567.72	578.64	583.28	593.46	593.39	599.92	601.44
	SD	66.92	65.53	69.44	66.08	62.56	81.78	86.75	88.68
	25th	511	528	534	540	552	546	546	552
	50th	552	570	589	589	595	601	601	601
	75th	595	608	621	627	634	647	660	668
	SEM	15	15	15	15	16	16	16	16
Physical	*n*	479	484	481	490	487	488	324	243
	M	548.61	562.62	572.85	577.82	586.72	593.97	602.46	608.65
	SD	64.22	55.14	62.00	56.37	55.31	71.10	72.38	74.15
	25th	523	533	543	543	543	554	554	565
	50th	554	565	578	578	592	605	617	617
	75th	592	592	617	617	627	646	646	656
	SEM	17	18	19	19	19	19	19	19
Language	*n*	472	475	478	480	479	486	319	239
	M	551.33	566.70	581.17	584.74	598.04	588.81	599.39	609.22
	SD	66.27	65.64	59.08	64.86	62.25	81.42	85.53	87.34
	25th	511	524	542	542	554	548	548	554
	50th	554	567	588	580	597	588	597	606
	75th	597	606	628	628	639	639	658	667
	SEM	13	13	14	15	16	15	16	16

Table 15
Scale Score Norm Tables for Each Developmental Domain by Age Band (48-71 months)

		Age in Months							
Domain	Statistic	48-50	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Cognitive	*n*	457	468	471	475	473	443	255	182
	M	550.77	573.12	582.06	590.35	601.72	600.22	610.15	609.85
	SD	70.53	66.51	72.15	68.57	71.18	81.03	92.40	94.37
	25th	506	526	540	548	548	556	548	548
	50th	548	571	591	591	597	610	616	603
	75th	597	616	622	634	647	647	670	670
	SEM	15	14	14	13	13	13	13	13
Literacy	*n*	454	462	463	469	469	482	312	235
	M	555.95	570.32	579.05	587.20	601.82	604.03	619.71	627.25
	SD	57.37	57.54	57.79	53.93	57.40	75.84	88.10	87.04
	25th	521	535	544	556	569	561	569	572
	50th	556	572	580	588	599	607	622	622
	75th	588	607	618	622	633	647	678.25	694
	SEM	13	12	12	12	12	12	12	12
Mathematics	*n*	449	457	469	468	468	440	246	175
	M	555.39	569.99	579.16	586.70	599.00	599.91	605.61	614.94
	SD	59.92	56.76	57.20	54.76	55.49	75.38	84.32	84.76
	25th	520	539	551	551	564	557	557	570
	50th	557	570	582	589	602	608	608	602
	75th	589	602	615	621	634	647	661	668
	SEM	14	14	14	14	14	14	14	14

Table 16

Raw Score Norm Tables for Each Developmental Domain by Age Band (6-23 months)

		Age in Months					
Domain	Statistic	6-8	9-11	12-14	15-17	18-20	21-23
Social-Emotional	*n*	392	363	449	465	468	452
	M	17.52	15.72	17.17	19.58	20.74	23.31
	SD	15.26	11.87	9.92	8.02	7.33	7.52
	25th	8	9	11	15	16	19
	50th	10	12	15	19	20	23
	75th	23	16	19	23	24	27
Physical	*n*	399	372	470	485	477	480
	M	12.08	11.50	13.57	15.95	16.59	17.93
	SD	9.38	7.00	6.00	5.25	4.52	4.70
	25th	6	8	10	13	14	16
	50th	8	9	13	15	16	18
	75th	17	12	15	18	19	20
Language	*n*	388	358	443	460	455	454
	M	14.98	12.90	13.50	15.75	17.71	20.31
	SD	15.50	12.19	9.42	7.95	7.24	7.24
	25th	5	6	8	11	14	16
	50th	8	9	12	15	17	20
	75th	25	12	15	18	21	25

Table 17

Raw Score Norm Tables for Each Developmental Domain by Age Band (6-23 months)

		Age in Months					
Domain	Statistic	6-8	9-11	12-14	15-17	18-20	21-23
Cognitive	*n*	390	354	439	451	452	440
	M	16.00	13.92	15.24	17.91	19.31	21.44
	SD	16.30	13.45	10.33	8.44	7.47	7.18
	25th	5	7	10	13	15	17
	50th	8	10	13	17	19	22
	75th	27	13	17	20	23	25
Literacy	*n*	390	358	439	458	454	454
	M	8.86	6.23	5.32	6.01	6.65	7.59
	SD	14.63	14.10	9.74	7.84	6.65	5.98
	25th	0	1	1	2	3	4
	50th	1	1	2	4	5	6
	75th	13	3	4	6	7	9
Mathematics	*n*	389	357	437	454	451	435
	M	5.72	3.51	2.75	4.33	5.20	6.21
	SD	10.34	9.34	6.46	5.99	5.29	4.41
	25th	0	0	0	0	2	3
	50th	0	0	0	3	5	6
	75th	8	0	3	6	7	8

Table 18
Raw Score Norm Tables for Each Developmental Domain by 3 Month Age Band (24-47 months)

		Age in Months							
Domain	Statistic	24-26	27-29	30-32	33-35	36-38	39-41	42-44	45-47
Social-Emotional	*n*	456	463	467	472	468	477	474	472
	M	25.04	27.49	28.83	31.87	31.02	31.97	33.44	36.70
	SD	7.97	8.30	8.54	9.79	9.77	10.22	9.21	9.56
	25th	20	23	24	26	25	26	28	31
	50th	25	27	28	32	31	33	34	37
	75th	29	32	33	37	37	39	39	43
Physical	*n*	480	482	480	482	474	477	479	485
	M	19.30	20.57	21.62	22.96	22.03	22.76	23.79	24.81
	SD	4.95	5.15	5.02	5.15	5.29	5.54	4.81	5.49
	25th	17	18	19	20	19	20	21	22
	50th	19	20	21	23	22	23	24	26
	75th	22	23	24	26	25	27	27	28
Language	*n*	465	462	462	468	472	473	479	470
	M	23.16	26.43	28.39	31.95	29.49	30.91	33.29	35.34
	SD	7.83	8.07	8.40	9.25	10.23	9.85	9.02	9.66
	25th	18	22	24	27	24	26	28	30
	50th	24	27	29	32	30	32	34	36
	75th	28	31	33	38	36	38	40	42

Table 19
Raw Score Norm Tables for Each Developmental Domain by 3 Month Age Band (24-47 months)

		Age in Months							
Domain	Statistic	24-26	27-29	30-32	33-35	36-38	39-41	42-44	45-47
Cognitive	*n*	445	448	441	468	454	456	467	457
	M	23.82	26.73	28.55	31.93	31.49	32.63	35.12	37.35
	SD	7.71	8.13	7.93	9.14	9.79	8.78	8.91	9.33
	25th	19	22	24	26	25	27	30	33
	50th	24	26	28	32	32	34	36	38
	75th	28	31	33	37	37	39	40	43
Literacy	*n*	440	446	445	453	447	451	457	456
	M	9.7	12.35	13.67	17.18	17.88	19.88	22.83	25.46
	SD	7.09	7.39	7.25	9.86	9.52	9.95	10.65	10.98
	25th	5	7	9	11	11	13	15	18
	50th	8	12	13	16	17	18	22	25
	75th	13	16	18	22	24	26	29	32
Mathematics	*n*	433	437	437	451	449	455	461	455
	M	8.18	10.05	11.31	13.72	13.56	15.09	16.87	18.94
	SD	5.17	5.16	5.35	6.64	6.92	7.23	7.21	7.39
	25th	5	7	8	9	9	10	12	14
	50th	8	10	11	13	13	14	17	19
	75th	11	12	14	18	18	20	21	24

Table 20
Raw Score Norm Tables for Each Developmental Domain by 3 Month Age Band (48-71 months)

		Age in Months							
Domain	Statistic	48-50	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Social-Emotional	*n*	467	474	479	482	480	448	264	182
	M	39.58	42.43	44.21	44.87	46.35	46.46	47.42	47.75
	SD	10.76	10.38	10.79	10.32	9.93	12.69	13.59	13.55
	25th	33	36	37	38	40	39	39	40
	50th	40	43	46	46	47	48	48	48
	75th	47	49	51	52	53	55	57	58
Physical	*n*	479	484	481	490	487	488	324	243
	M	26.31	27.55	28.43	28.83	29.58	30.27	30.98	31.46
	SD	5.58	4.85	5.32	4.87	4.70	6.10	6.17	6.24
	25th	24	25	26	26	26	27	27	28
	50th	27	28	29	29	30	31	32	32
	75th	30	30	32	32	33	35	35	36
Language	*n*	472	475	478	480	479	486	319	239
	M	37.72	39.84	41.82	42.31	44.12	42.61	43.83	45.12
	SD	9.63	9.16	9.32	8.62	7.95	10.94	11.23	11.02
	25th	32	34	37	37	39	38	38	39
	50th	39	41	44	43	45	44	45	46
	75th	45	46	48	48	49	49	51	52

Table 21
Raw Score Norm Tables for Each Developmental Domain by 3 Month Age Band (48-71 months)

		Age in Months							
Domain	Statistic	48-50	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Cognitive	*n*	457	468	471	475	473	443	255	182
	M	40.19	43.10	44.85	46.10	47.72	47.36	48.73	48.79
	SD	10.62	9.97	10.61	10.08	10.42	11.99	13.62	13.93
	25th	34	37	39	40	40	41	40	40
	50th	40	43	46	46	47	49	50	48
	75th	47	50	51	53	55	55	58	58
Literacy	*n*	454	462	463	469	469	482	312	235
	M	30.98	34.45	36.61	38.51	42.25	43.40	47.82	49.46
	SD	13.50	13.55	13.89	13.50	14.66	18.36	20.92	21.56
	25th	22	25	27	30	33	31	33	34
	50th	30	34	36	38	41	43	47	47
	75th	38	43	46	47	50	54	63	67
Mathematics	*n*	449	457	469	468	468	440	246	175
	M	21.93	24.11	25.55	26.60	28.46	28.77	29.59	30.83
	SD	8.67	8.31	8.35	8.29	8.38	10.63	11.90	12.16
	25th	16	19	21	21	23	22	22	24
	50th	22	24	26	27	29	30	30	29
	75th	27	29	31	32	34	36	38	39