CEME
Technical Report

The Center for Educational Measurement and Evaluation

The Development Continuum for Infants, Toddlers & Twos Assessment System: The Assessment Component for the Creative Curriculum for Infants, Toddlers and Twos

Richard G. Lambert

CEME
The Center for Educational Measurement and Evaluation

College of Education
UNC CHARLOTTE

The Developmental Continuum for Infants, Toddlers & Twos Assessment System
The Assessment Component of The Creative Curriculum
for Infants, Toddlers & Twos

Technical Report

The Creative Curriculum for Infants, Toddlers, and Twos is an extension of *The Creative Curriculum®* for preschool children. As such, it is a comprehensive and integrative curricular model for early childhood programs. It is similarly rooted in a broad range of theoretical and research-based foundations concerning child development and learning. The curriculum is designed to help teachers and parents understand their infant, toddler, or two year old across all domains of development. The model outlines methods teachers can use to create a supportive, enriching, and responsive environment that engages the developing interests and capacities of the children. A variety of strategies are offered to teachers that include methods for structuring the physical environment and daily routine, establishing relationships with children, individualizing interactions that recognize differences between children, and enhancing the developmental progress of all children.

Purpose

The purpose of this report is to begin the process of accumulating evidence about the reliability and validity of the information that the assessment component of the curriculum can provide. It is important to note that all of the reports and suggestions to teachers provided by the teacher's guide and website are based on information from single items. The accompanying website does not provide, or utilize, scale scores in any of its processing and suggestions, and is intended for formative assessment, evaluation, and instructional planning purposes. This report is an attempt to further facilitate the proper use of the information that can be provided by the measure for other purposes such as program planning and quality improvement, monitoring the implementation of the curriculum, and teacher development. In addition, researchers may choose to use the information the measure provides for more summative purposes such as research and program evaluation. This report presents evidence that can be useful to program administrators, researchers, and evaluators who desire to use the information provided by the measure by offering guidance about the formation of scale scores and their measurement properties.

Measure

The data that were used for this study were collected by teachers working in child care and preschool settings that serve children younger than three years of age. Each teacher was trained to use *The Creative Curriculum for Infants, Toddlers, and Twos* and the accompanying assessment measure, the *Developmental Continuum for Infants, Toddlers, and Twos*. The measure is designed to help teachers record and organize their observations in terms of the progress each child is making on the four goals of the curriculum. These goals include enhancing the development of each child in the following areas: social/emotional development, physical development, cognitive development, and language development. Each goal is broken

into specific objectives and each objective has a corresponding item. Teachers using the measure are trained to make ratings of the developmental progress of each child using 21 items that represent the objectives. Each item includes a five-point rating scale, labeled Step 1 through Step 5. Specific anchors are given for each step along with example behaviors.

The measure is grounded in an expectation that children do not simply master a particular developmental task as an all or nothing proposition. Rather, there is an expected progression of successive attainments, or smaller steps toward the attainment of developmental milestones. In a sense, the anchor points or steps for each item represent these smaller steps, and the items represent larger developmental milestones. The measure is designed to offer teachers and parents information about the developmental level of a child in order to facilitate efforts to support and enhance development. Teachers and parents can use this developmental information and feedback to enhance their understanding of the child, leading to more sensitive and responsive interactions with the child and more supportive experiences in the learning environment of the classroom. The measure is also designed to help guide and focus a teacher's observations of children. It is designed primarily as a formative tool to help with instructional planning and communication with parents.

Norm Sample

The National Survey of America's Families (Urban Institute, 2002) was conducted in three rounds (1997, 1999, and 2002), and is designed to provide information about child and family well-being throughout the United States. The survey attempted to help social science researchers understand low-income families in particular. Over 40,000 families were interviewed across the waves of data collection. The survey results were used to investigate specific characteristics of children younger than three years of age who attend out-of-home care in center-based settings. The 2002 survey results indicate that 53.16% of these children are between 24 and 35 months old, 29.37% are 12 to 23 months old, and 17.49% are birth to 11 months old. The survey results were used to further segment these children into groups served by Head Start programs (5.99%) and those served in other group care settings (94.01%), and into six age and care setting specific cells: Head Start and 0-11 months (1.48%), Head Start and 12-23 months (1.83%), Head Start and 24-35 months (2.68%), other group care settings and 0-11 months (16.01%), other group care settings and 0-11 months (16.01%), and other group care settings and 24-35 months (50.47%). The study sample did not exactly conform to these cell percentages. Therefore, these estimates were used to create sampling weights so that the proportion of study children in each of these six cells would match the population.

The sample consisted of 2,256 children from programs in 26 states and the District of Columbia. These children were nested within 352 raters (teachers) who work within 167 different programs. Table 1 contains the demographic characteristics of the sample including geographic distribution. Ratings were included from all regions of the country and from urban (38.3%), suburban (43.7%), and rural (18.0%) locations. The northeastern part of the United States may have been under represented in the sample (8.4%) and the southeastern region may have been over represented (48.1%). Approximately one in ten study children (9.9%) have an Individualized Family Service Plan (IFSP). Hispanic ethnicity was reported for 15.9% of the children and the

teachers reported that 6.5% live in homes where Spanish is the primary language. Males may have been slightly over represented as they comprise 55% of the sample.

Descriptive Statistics, Reliability, and Normative Information

The data is comprised of teacher ratings of the developmental progress of the study children using the 21 items that outline developmental progress on each of the goals and objectives of the curriculum. The first step in the analysis was to examine the distributional properties for each of the items and goal-specific sections of the instrument using descriptive statistics. This analysis revealed some evidence of floor effects. Considering the range of ages that the instrument is designed to describe, and the amount of development, growth, and change that is typical for children under three years of age, this was not unexpected. Approximately one in three children less than 12 months of age (33.7%) received a rating of 1, the lowest step, for all 21 items. This finding is not surprising as the instrument was not designed to be sensitive to small differences between very young children. It is important to note that the infants who received a rating of 1 for all 21 items were 4.83 months of age on average and ranged in age from one month old to nine months of age. This floor effect was almost non-existent in the groups of older children. For children 12 to 23 months of age, 1.7% received ratings of 1 on all items. Similarly, 1.7% of children 23 to 35 months of age received ratings of 1 on all items. There were no substantial ceiling effects. No children under 24 months of age received ratings of 5 for all items, and only 2.0% of children 24 to 35 months of age received ratings of 5 on all items. These children were approximately 31 months of age on average and ranged in age from 26 to 35 months of age.

Next, the distributions of the ratings for each item were examined. The median rating for almost all of the items for the total sample was 3. The only exception was item 8, focused on gross motor skills, which had a median rating of 4. For children birth to 11 months of age, the median rating for all items except item 8 was 1. The median rating for item 8 was 2. For children 12 to 23 months of age, the median rating was 3 for almost all items. For five of the items, 15, 18, 19, 20, and 21, the median was 2. For children 24 to 35 months of age, the median ratings were 3 for ten of the items and 4 for 11 of the items. These patterns would be expected and suggest that teacher ratings based on the items and anchor points contained in this measure are sensitive to age and development. This information also suggests that item 8 may have a lower item difficulty and items 15, 18, 19, 20, and 21 may have higher difficulty levels.

Table 2 contains the percentage of children receiving ratings at each step on the ratings scale for the entire sample. Although the item level distributions of ratings contain only five discrete values, a reasonably unimodal and symmetrical distribution was observed for each item and is reflected in these percentages. Tables 3 through 5 contain the same information for each of the age levels and can function as a type of age-specific norm tables. As expected, the distributions of ratings for children birth to 11 months of age reflect a positively skewed shape with large numbers of children receiving ratings on the lower end of the scale. For children 12 to 23 months of age, the item level distributions are reasonably unimodal and symmetrical. For children 24 to 35 months of age, the item level distributions of ratings are somewhat negatively skewed as expected. Taken together, these findings present some evidence that the anchor points

on the rating scale represent a sequence of developmental steps on each of the curricular objectives.

For each goal, or section of the measure, a total score was created using the average of the ratings. The distributional properties of these scores for both the total sample and each age group are presented in Table 6. A total score across all 21 items was created in the same manner and is labeled Developmental Progress. The distributions of scores for each of these sections of the measure presented a reasonably unimodal and symmetrical shape with the additional feature of having a spike at 1, the lowest step. This feature represents the floor effect mentioned above and largely represents the scores of youngest children in the sample. Careful examination of the values in Table 6 reveals the expected progression of both mean and median scores in association with age. The correlations with the child's age in months at the time of the rating are presented and are all moderately high. It is important to note that these values are high enough to indicate the expected relationship with age and development, and yet low enough to indicate a substantial amount of within-age group variability. As children of the same chronological age can not be expected to present the exact same developmental stage, this finding indicates some evidence that the measure is successfully separating the children by developmental level.

Cronbach's alpha values, a measure of internal consistency reliability, are also reported in Table 6. Almost all of these values, both at the total sample and age-specific sub-sample levels, are in the acceptable range (.80 or greater). The only exception falls in the Physical Development goal, where the values for the children in the 12 to 23 month and 24 to 35 month groups were .774 and .797 respectively. However, Cronbach's alpha is sensitive to the number of items comprising a scale score and this goal has only two objectives. These values could be considered acceptable for such a small scale. The mean inter-item correlations are reported for each goal. Table 7 contains the correlations between the scale scores. All of these values are quite high, suggesting the possibility that the instrument yields information that is measuring a general construct of developmental progress, rather than goal-specific domains.

Construct Validity

In order to examine the underlying factor structure of the information contained in the teacher ratings, factor analytic techniques were used. Initially, Principal Components analysis was conducted using a randomly selected calibration sub-sample with the intention of conducting a confirmatory factor analysis using the remaining holdout sample. This analysis yielded a single factor solution that accounted for over 70% of the variance in ratings. This finding supports the single total score labeled here as Developmental Progress. In order to determine if the single factor solution was stable, additional analyses were conducted using both Principal Components analysis and Principal Axis Factoring. Analyses were conducted using the sub-sample and the entire sample of children, and with and without the sampling weights. Analyses were also conducted using the total sample and the separate age cohorts. They were also conducted by both eliminating and retaining those children who were assigned ratings with no variance (either a rating of 1 or 5 for all of the items). In every case the solution was the same, a single underlying dimension that accounted for a substantial majority of the variance in the ratings. When only children who were 30 months or older were used, a two factor solution

emerged in some of the analyses. However, none of these solutions conformed to the theoretical developmental domains represented by the goals of the curriculum. Taken together, these findings seem to suggest that that for children as young as those in this sample, the measure is capturing overall developmental progress. As might be expected from developmental theory, development across the theoretical domains can tend to happen simultaneously at these young ages. These findings may also suggest that even though children within age groups can differ in their rates of developmental progress, within individual children development may tend to occur at consistent rates across domains. As children reach the end of the age range for which this measure is intended, domain specific rates of development may begin to emerge.

The next step in the analysis involved the creation of an overall scale score using methodology based on Item Response Theory (IRT). The Winsteps software was used to apply the one parameter, or Rasch model, to these data. The randomly selected calibration sample was used to create a single scale score. Item difficulty parameters were estimated for each item. In addition, ability estimates were made for each child. Table 8 contains the mean and median scale scores for each of the age sub-groups. As expected, scores were associated with the child's age. The birth to 11 months of age group yielded an average scale score of 22.92. The 12 to 23 month group yielded an average score of 43.00 and for the two year old group, an average of 60.40 was observed. Reasonable within age cohort variability was also observed as indicated by the standard deviations.

IRT scaling allows the researcher to place items and persons on the same continuum or scale. Item locations are interpreted as item difficulties and person locations are interpreted as ability levels. A specific child can be expected to have a high probability of scoring at a high level on items with locations that are lower on the scale than the child's ability estimate. Similarly, a specific child can be expected to have a high probability of scoring at a lower level on items with locations that are higher than the child's ability estimate. Table 9 contains the item locations for each item, arranged according to descending difficulty. Note that item 20, a language development item, has the highest difficulty level. Item 8, gross motor development, has the lowest difficulty level. The general pattern to the item difficulty levels conforms to developmental theory. These item locations can be displayed graphically to form what can be viewed as a developmental pathway, or expected sequence of development. Figure 1 displays the developmental pathway for these items. The location of the median score for each age level is also displayed.

IRT methods also produce item and person fit statistics. These values allow the researcher to identify potentially problematic items. Such items may yield information that is outside the expected pattern of person ability and item difficulty levels. These items may also present problems with construct validity and may fail the test of unidimensionality, perhaps measuring at least in part a separate construct. Information yielded by such items may fall outside the expected pathway from easiest to most difficult developmental tasks. Standardized fit statistics greater than 2 indicate potentially problematic items. As indicated in Table 9, items 1, 5, 7, 9, 11, 12, 13, and 16 fall in this category. It may be important to examine the content and descriptions of the sample behaviors in the anchor points for these items to verify the theoretical sequence of development that they imply. It may also be important to examine the training of raters for these items.

IRT methods provide estimates of person level reliability and item level reliability. The item level reliability value indexes the capacity of the IRT model to estimate the true difficulty level for the items. It is an index of the precision of estimation of item separation and the reproducibility of the item locations. Similarly, the person level reliability is an index of the precision of person separation and the reproducibility of person ability estimates. Both of these values were .90 for this measure, using the data from this sample, and would be considered acceptable.

Conclusions

This report presents evidence that the measure, when properly used in the context of a program that been taken advantage of the proper training, technical assistance, and curriculum implementation strategies, can provide information that has adequate measurement properties. The distributional findings suggest that the information that the measure provides, at least with the large and diverse sample used in this study, is addressing developmental progress in the children and separates children both by age cohorts and within age cohorts. The reliability evidence is strong and suggestive of the reproducibility of the results about children over repeated observations.

Broadly speaking, the starting point in the measure evaluation process is an understanding of the intended purposes of a particular instrument. This understanding is fundamental to the proper interpretation and use of the information a measure provides. Judging the validity of the information provided by an instrument involves an evaluation of whether proper interpretation of its test scores obtained under the intended conditions is useful. Therefore, it is important to evaluate any instrument in light of its relevance to a particular purpose, research question, or proposed evaluative use.

The construct validity evidence suggests that the measure is addressing a single construct of global development. Researchers, evaluators, and program administrators may want to use a total score for summative evaluation purposes. However, information from the item and goal level scores may be useful to parents, teachers, and those who mentor and supervise teachers for formative purposes such as instructional planning. Program administrators, researchers, and evaluators who are interested in using the information the measure provides, are urged to do so in accordance with widely accepted standards of practice regarding the assessment of young children. In general, it is important to consider several broad principles:

- The reliability and validity of the information provided by assessments for young children tends to increase with the age of the children being assessed

- No source of information should be used as the sole source for decision-making purposes.

- Teacher ratings are only one source of information about children, and reflect the unique perspective of the teacher and the teacher's experience of the child within the classroom context.

- Multiple sources of information (informants, methods, and measures) provide a more complete picture of the child's developmental progress.

Directions for Future Research

This study has begun the process of establishing reliability and validity evidence for the measure. A diverse set of indicators discussed above, when taken collectively, represent a favorable set of technical properties. Validation and enhancement of measurement tools, particularly those for use with young children, can proceed for the life of the use of a particular measure. It is generally recognized that the consistency of children's responses to test stimuli and their behaviors when being observed, and therefore the reliability of test score information, tends to increase with age. Reliability evidence and standard errors of measurement should be reported separately for each age, grade level, or subgroup for which a test is intended. Reliability estimates based on scores from combined age groups or developmental levels can be spuriously high. The younger the age of the children for whom a measure is intended, the narrower the age range of the subgroups needs to be to collect reliability evidence. Therefore, in future research, it may be useful to pursue age standardized scores using IRT analyses at the level of more narrow age cohorts than were used in this study, including the selection of the representative age-specific samples needed to do so. Reliability and validity evidence could then be examined using these scores.

In addition, future research regarding the measurement properties of information yielded by this measure could focus on the following issues:

- Reanalysis of item fit statistics after examination of item content, anchor point example behaviors, and rater training for the potentially problematic items.

- Content validation of the items that presented possible problems with fit.

- Concurrent validity studies where the information from this measure is related to that collected by outside observers using standardized measures.
- Inter-rater reliability studies, including an examination of between and within teacher or rater variance.

- Examination of patterns of growth in individual children across multiple measures over time.

- Interviews with parents and teachers to study the usefulness of the measure for instructional planning and assistance in understanding individual children and their needs.

- Examination of the measurement properties of the information the measure provides when used with sub-groups such as children with special needs and those from ethnic and linguistic minorities.

Table 1
Demographic characteristics of the sample.

Characteristic	Category	Norm Sample
Age in Months	*0-11*	17.5%
	12-23	29.4%
	24-35	53.1%
Primary Language in the Home	English	88.7%
	Spanish	6.5%
	Other	4.8%
Disability Status	No IFSP	90.1%
	Has IFSP	9.9%
Gender	Male	55.0%
	Female	45.0%
Location	Urban	38.3%
	Suburban	43.7%
	Rural	18.0%
Agency Type	Head Start	6.0%
	Public School	5.2%
	University	4.5%
	Center-Based	84.3%
Region	West	22.6%
	Midwest	17.2%
	Northeast	8.4%
	Southeast	48.1%
	Southwest	3.8%

Note - n=2,256.

Table 2

Percentage of Children in the Norm Sample at each Developmental Step on the Fall Assessment.

Domain and Item	Step 1	Step 2	Step 3	Step 4	Step 5
Social / Emotional Development					
1 Trusts known, caring adutls.	15.2%	15.1%	28.0%	29.5%	12.2%
2 Regulates own behavior.	17.2%	20.9%	22.7%	28.5%	10.8%
3 Manages own feelings.	15.3%	25.6%	24.5%	25.4%	9.2%
4 Responds to others' feelings with growing empathy.	17.5%	26.5%	31.7%	14.4%	9.8%
5 Plays with other children.	14.5%	16.1%	27.2%	27.4%	14.8%
6 Learns to be a member of a group.	13.2%	18.7%	38.3%	15.6%	14.2%
7 Uses personal care skills.	14.7%	19.4%	30.9%	19.5%	15.5%
Physical Development					
8 Demonstrates basic gross motor skills.	10.4%	14.5%	21.4%	36.0%	17.6%
9 Demonstrates basic fine motor skills.	12.6%	17.9%	29.5%	29.1%	10.8%
Cognitive Development					
10 Sustains attention.	17.2%	20.0%	30.5%	21.4%	10.9%
11 Understands how objects can be used.	15.4%	21.5%	41.5%	14.2%	7.5%
12 Shows a beginning understanding of cause and effect.	14.0%	21.8%	31.0%	24.0%	9.2%
13 Shows a beginning understanding that things can be grouped.	14.2%	23.5%	35.5%	17.5%	9.3%
14 Uses problem-solving strategies.	21.1%	17.0%	34.3%	18.7%	8.9%
15 Engages in pretend play.	17.0%	20.8%	33.3%	17.7%	11.3%
Language Development					
16 Develops receptive language.	14.7%	20.4%	28.6%	22.8%	13.5%
17 Develops expressive language.	14.2%	22.4%	25.4%	24.7%	13.2%
18 Participates in conversations.	15.3%	23.9%	29.4%	18.4%	12.9%
19 Enjoys books and being read to.	16.2%	26.6%	32.8%	13.4%	11.0%
20 Shows an awareness of pictures and print.	18.0%	29.8%	26.0%	19.5%	6.8%
21 Experiments with drawing and writing.	15.3%	20.9%	33.3%	22.4%	8.1%

Note - n=2,256.

Table 3

Percentage of Children 0-11 Months of Age at each Developmental Step on the Fall Assessment.

Domain and Item	Step 1	Step 2	Step 3	Step 4	Step 5
Social / Emotional Development					
1 Trusts known, caring adutls.	54.0%	30.9%	8.5%	5.5%	1.1%
2 Regulates own behavior.	63.8%	28.3%	4.0%	2.0%	1.9%
3 Manages own feelings.	53.9%	37.3%	7.7%	1.1%	0.0%
4 Responds to others' feelings with growing empathy.	60.3%	29.4%	9.6%	0.7%	0.1%
5 Plays with other children.	55.4%	30.2%	10.3%	3.9%	0.2%
6 Learns to be a member of a group.	56.6%	29.2%	13.0%	1.2%	0.1%
7 Uses personal care skills.	60.7%	31.1%	7.0%	1.0%	0.1%
Physical Development					
8 Demonstrates basic gross motor skills.	47.5%	44.7%	4.8%	1.7%	1.5%
9 Demonstrates basic fine motor skills.	56.0%	34.8%	5.8%	2.9%	0.5%
Cognitive Development					
10 Sustains attention.	68.3%	24.8%	4.0%	2.8%	0.1%
11 Understands how objects can be used.	68.2%	27.3%	3.2%	1.2%	0.1%
12 Shows a beginning understanding of cause and effect.	56.4%	34.4%	8.5%	0.6%	0.1%
13 Shows a beginning understanding that things can be grouped.	56.0%	37.7%	5.0%	1.3%	0.0%
14 Uses problem-solving strategies.	80.6%	14.9%	3.8%	0.7%	0.0%
15 Engages in pretend play.	73.4%	21.2%	4.2%	1.2%	0.1%
Language Development					
16 Develops receptive language.	63.1%	30.5%	5.2%	1.0%	0.1%
17 Develops expressive language.	61.8%	31.1%	5.9%	1.2%	0.0%
18 Participates in conversations.	62.2%	33.3%	3.2%	1.2%	0.1%
19 Enjoys books and being read to.	69.3%	25.5%	2.8%	1.8%	0.6%
20 Shows an awareness of pictures and print.	71.0%	24.0%	2.4%	2.6%	0.0%
21 Experiments with drawing and writing.	65.6%	27.8%	4.2%	1.6%	0.8%

Note - n=394.

Table 4
Percentage of Children 12-23 Months of Age at each Developmental Step on the Fall Assessment.

Domain and Item	Step 1	Step 2	Step 3	Step 4	Step 5
Social / Emotional Development					
1 Trusts known, caring adutls.	6.4%	20.5%	39.3%	29.6%	4.2%
2 Regulates own behavior.	9.3%	34.4%	31.4%	20.9%	4.0%
3 Manages own feelings.	8.6%	37.9%	33.6%	16.6%	3.3%
4 Responds to others' feelings with growing empathy.	11.6%	38.0%	39.2%	6.2%	5.0%
5 Plays with other children.	10.2%	24.6%	38.2%	20.5%	6.5%
6 Learns to be a member of a group.	6.8%	25.9%	50.5%	12.4%	4.4%
7 Uses personal care skills.	8.8%	30.1%	46.9%	11.0%	3.2%
Physical Development					
8 Demonstrates basic gross motor skills.	4.1%	18.1%	41.3%	29.4%	7.0%
9 Demonstrates basic fine motor skills.	6.7%	29.1%	45.3%	17.0%	1.8%
Cognitive Development					
10 Sustains attention.	11.8%	34.0%	37.9%	13.6%	2.8%
11 Understands how objects can be used.	10.5%	38.2%	42.8%	6.5%	2.1%
12 Shows a beginning understanding of cause and effect.	9.9%	37.9%	36.0%	13.5%	2.6%
13 Shows a beginning understanding that things can be grouped.	9.8%	34.8%	45.8%	7.3%	2.3%
14 Uses problem-solving strategies.	16.5%	27.5%	43.4%	9.5%	3.0%
15 Engages in pretend play.	9.8%	41.2%	36.9%	8.4%	3.7%
Language Development					
16 Develops receptive language.	7.3%	34.9%	39.6%	14.4%	3.8%
17 Develops expressive language.	6.5%	43.0%	34.8%	14.3%	1.4%
18 Participates in conversations.	9.4%	42.9%	37.9%	7.7%	2.2%
19 Enjoys books and being read to.	11.9%	47.9%	33.0%	4.1%	3.1%
20 Shows an awareness of pictures and print.	14.2%	51.2%	27.5%	6.3%	0.7%
21 Experiments with drawing and writing.	13.8%	37.5%	35.4%	12.1%	1.2%

Note - n=662.

Table 5

Percentage of Children 24-35 Months of Age at each Developmental Step on the Fall Assessment.

Domain and Item	Step 1	Step 2	Step 3	Step 4	Step 5
Social / Emotional Development					
1 Trusts known, caring adutls.	8.1%	7.2%	27.7%	36.9%	20.1%
2 Regulates own behavior.	7.5%	11.3%	23.6%	40.4%	17.1%
3 Manages own feelings.	7.3%	15.3%	24.6%	37.5%	15.3%
4 Responds to others' feelings with growing empathy.	8.7%	19.5%	33.9%	22.7%	15.2%
5 Plays with other children.	6.1%	7.7%	25.7%	37.3%	23.2%
6 Learns to be a member of a group.	4.9%	12.0%	38.4%	21.3%	23.4%
7 Uses personal care skills.	4.7%	10.1%	29.0%	29.5%	26.7%
Physical Development					
8 Demonstrates basic gross motor skills.	2.6%	3.3%	15.5%	50.2%	28.4%
9 Demonstrates basic fine motor skills.	3.1%	6.9%	27.8%	43.4%	18.7%
Cognitive Development					
10 Sustains attention.	4.9%	10.9%	34.4%	31.2%	18.6%
11 Understands how objects can be used.	3.9%	10.8%	50.9%	21.9%	12.4%
12 Shows a beginning understanding of cause and effect.	4.5%	9.4%	34.3%	36.3%	15.4%
13 Shows a beginning understanding that things can be grouped.	5.3%	13.6%	38.2%	27.2%	15.6%
14 Uses problem-solving strategies.	7.5%	11.7%	37.6%	28.8%	14.5%
15 Engages in pretend play.	5.7%	9.6%	39.2%	27.2%	18.3%
Language Development					
16 Develops receptive language.	4.4%	9.5%	29.4%	33.8%	22.9%
17 Develops expressive language.	5.2%	8.8%	25.7%	37.0%	23.4%
18 Participates in conversations.	6.1%	11.3%	31.9%	28.7%	22.1%
19 Enjoys books and being read to.	5.1%	15.3%	40.2%	21.4%	18.0%
20 Shows an awareness of pictures and print.	7.2%	19.7%	30.8%	30.6%	11.7%
21 Experiments with drawing and writing.	5.9%	10.5%	38.0%	32.2%	13.3%

Note - n=1,200.

Table 6
Properties of the Distributions of the Domain and Total Scores.

Domain	Age	Mean	SD	Min	Percentiles 25th	Percentiles 50th	Percentiles 75th	Max	Number of Items	Alpha	Mean Iter-item Correlation	Correlation with Age
Social / Emotional Development	*0-11*	1.540	0.658	1.000	1.000	1.286	2.000	4.860	7	0.929	0.660	
	12-23	2.753	0.793	1.000	2.250	2.805	3.286	5.000		0.910	0.592	
	24-35	3.451	0.959	1.000	2.857	3.571	4.143	5.000		0.928	0.648	
	0-35	2.919	1.111	1.000	2.143	3.000	3.714	5.000		0.949	0.728	0.653
Physical Development	*0-11*	1.586	0.714	1.000	1.000	1.500	2.000	5.000	2	0.899	0.817	
	12-23	2.973	0.832	1.000	2.500	3.000	3.500	5.000		0.774	0.635	
	24-35	3.821	0.857	1.000	3.500	4.000	4.500	5.000		0.797	0.667	
	0-35	3.199	1.158	1.000	2.500	3.500	4.000	5.000		0.887	0.798	0.725
Cognitive Development	*0-11*	1.372	0.543	1.000	1.000	1.167	1.667	4.500	6	0.934	0.708	
	12-23	2.558	0.783	1.000	2.000	2.500	3.000	4.830		0.911	0.633	
	24-35	3.346	0.906	1.000	3.000	3.333	4.000	5.000		0.919	0.656	
	0-35	2.786	1.090	1.000	2.000	3.000	3.500	5.000		0.950	0.762	0.690
Language Development	*0-11*	1.312	0.551	1.000	1.000	1.000	1.724	4.330	6	0.924	0.673	
	12-23	2.487	0.754	1.000	2.000	2.500	3.000	5.000		0.920	0.659	
	24-35	3.420	0.940	1.000	2.833	3.500	4.000	5.000		0.926	0.677	
	0-35	2.808	1.124	1.000	2.000	2.833	3.667	5.000		0.956	0.784	0.718
Developmental Progress	*0-11*	1.438	0.556	1.000	1.000	1.200	1.714	4.480	21	0.973	0.645	
	12-23	2.643	0.722	1.000	2.143	2.619	3.095	4.810		0.966	0.579	
	24-35	3.447	0.869	1.000	3.000	3.476	4.048	5.000		0.970	0.607	
	0-35	2.863	1.076	1.000	2.048	2.952	3.663	5.000		0.982	0.723	0.723

Note. 0-11 n=394, 12-23 n=662, 24-35 n=1,200, 0-35 n=2,256.

Table 7

Correlations between domain scores.

Domain	Social Emotional Development	Physical Development	Cognitive Development	Language Development
Physical Development	0.843			
Cognitive Development	0.906	0.869		
Language Development	0.891	0.847	0.915	
Developmental Progress	0.966	0.907	0.968	0.963

Note. $p < .001$ for all coefficients.

Table 8

Total Scale Score by Age Range.

Age Range	Mean	Median	SD
0-11	22.92	20.00	7.56
12-23	43.00	42.70	15.12
24-35	60.40	62.35	16.75
0-35	48.76	50.51	20.60

Note. n=1,239.

Table 9

Item Locations and Fit Statistics.

Item	Item Location	Z Infit
20	58.65	-1.2
4	56.45	1.2
19	55.59	0.4
11	55.36	-6.3
14	55.31	1.1
13	54.16	-2.5
21	53.87	-0.1
15	52.53	-0.3
18	52.01	-1.5
10	51.07	-0.1
3	51.03	0.2
12	50.85	-4.2
6	49.29	0.9
2	49.24	1.9
17	48.58	-1.4
16	48.05	-4.4
7	47.47	-2.3
9	45.63	-3.0
5	45.53	3.9
1	44.52	9.1
8	37.51	1.0

Note. n=1,239.

Figure 1. Developmental Pathway with Item and Person Locations.

Persons scale:			Items scale:
most developed	**80**		most difficult

	70		
	24-35 60		
	Q20		
Q11		Q19 Q14 Q4	
Q13	Q21		
Q12	Q18 Q15 Q10 **50**	Q3	
Q16 Q7	Q17	Q6 Q2	
Q9			Q5
			Q1
	12-23 40		
		Q8	
	30		
	0-11 20		

least developed			least difficult
z=-2	Infit	z=+2	