Technical Qualities

Sample Description

To investigate the reliability and validity of the Early LAP, a sample representative of the United States was selected based on population projections for the Year 2000 by the U.S. Census Bureau (1995). The sample included 285 children ages 2 to 44 months old, including children with typical and atypical development. A stratified sampling procedure was used based on geographic region, age, race, gender, and type of setting.

Geographic Distribution of Project Sites

Four sites were selected to represent the geographic regions of the United States with similar sample sizes in each region: Northeast (New York City and Westchester County, New York), South (Orange County, North Carolina), North Central (Greater Metropolitan Area of Kansas City and Lawrence, Kansas), and West (San Jose, California). Table 3 depicts the distribution of the sample across the four sites.

Table 3. Geographic Distribution of Sample

Geographic Area	Number of Children	Percentage of Sample
Northeast	59	20.7%
South	68	23.9%
North Central	81	28.4%
West	77	27.0%
Total	285	100%

Age

Children from 2 to 44 months were selected for the study in the following seven categories: birth to 6 months, 7-12 months, 13-18 months, 19-24 months, 25-30 months, 31-36 months, and 37-44 months. Children in the last category, 37-44 months, were included to demonstrate the upper age limit for the Early LAP. Table 4 shows the distribution of the Project Sample by age category and geographic region.

Table 4. Number of Participants by Age and Geographic Region

Age	Northeast	South	North Central	West	Total
Birth - 6 months	0	7	16	12	35 (12.3%)
7-12 months	6	14	13	11	44 (15.4%)
13-18 months	11	10	12	12	45 (15.8%)
19-24 months	16	9	14	11	50 (17.5%)

25-30 months	10	11	8	12	41 (14.4%)
31-36 months	5	8	14	13	40 (14.0%)
37-44 months	11	9	4	6	30 (10.5%)
Total	59	68	81	77	285 (100%)

Gender

An approximately equal number of males and females were selected for the sample. The distribution of the Project Sample by gender and geographic region is indicated in Table 5.

Table 5. Number of Participants by Gender and Geographic Region

Gender	Northeast	South	North Central	West	Total
Female	18	32	40	35	125 (43.9%)
Male	41	36	41	42	160 (56.1%)
Total	59	68	81	77	285 (100%)

Race/Ethnicity

To represent the variety of cultural and ethnic groups in the United States, children were proportionally selected for the sample to reflect the major racial/ethnic groups indicated in the U.S. Census Bureau population projections for the Year 2000 (1995). These groups included the following categories: African American; American Indian, Eskimo, and Aleut; Asian and Pacific Islander; Hispanic origin; and White. In addition, an Other category included mostly bi-racial children who were categorized in this group by their parents. Table 6 depicts the racial/ethnic distribution by geographic region.

Table 6. Number of Participants by Race/Ethnicity and Geographic Region

Race or Ethnic Group	Northeast	South	North Central	West	Total**
African American	23	8	20	6	57 (20.0%)
American Indian, Eskimo, and Aleut	2	0	2	1	5 (1.8%)
Asian and Pacific Islander	1	1	1	5	8 (2.8%)
Hispanic origin	10	3	1	13	27 (9.5%)
White	15	52	50	52	169 (59.3%)
Other*	8	4	7	0	19 (6.6%)
Total	59	68	81	77	285 (100%)

* The majority of the children classified as Other were bi-racial children.

** The 1995 U.S. Census Bureau population projections for Year 2000 were: African American - 13%; American Indian, Eskimo, and Aleut - 1%; Asian and Pacific Islander - 4%; Hispanic origin - 11%; and White - 71%.

Program Types

Because the Early LAP is used for in-home services as well as in group settings for infants and toddlers, children were recruited from three different settings: child care center programs (N=179, 62.8%); Early Head Start programs (N=37, 13.0%); and individual homes (N=69, 24.2%). Thus the majority of the children participating in the study were assessed in a child care program (N=216, 75.8%). Of the children who were included from individual homes, 65.2% were located in the West, 29% were located in the North Central site, 5.8% were located in the South, and none in the Northeast. A total of 29 child care programs participated in the study, with some variation in the types of centers across the four geographic regions. The Northeast site (New York City and Westchester County, New York) included four Early Head Start centers, one developmental day center (centers serving only children with disabilities), one YMCA child care facility, and four community day care centers. In the South (Orange County, North Carolina), six community child care centers, one Early Head Start, and one developmental day center participated in the study. The participants in the North Central site (Greater Metropolitan Area of Kansas City and Lawrence, Kansas) included five community day care centers and one center located on an American Indian college campus. The site in the West (San Jose, California) included one developmental day care, a perinatal unit in a substance abuse program, and three community child care centers

Children With Disabilities

Because the Early LAP is frequently used in conjunction with standardized instruments to examine the skill development of infants and toddlers with developmental delays, a subsample of children with disabilities (6%) was selected that reflected the U.S. rates for children under age 18 with disabilities (U.S. Census Bureau, 1995). These children had been professionally diagnosed and were receiving early intervention services. Where possible, appropriate adaptations in the use of materials and procedures were used to allow children to respond to test items independent of their particular impairment (e.g., use sign language for hearing impaired child, use adaptive equipment for child with limited mobility). It is important to note that older children functioning in the birth to age three range may benefit from the information gathered through the Early LAP.

Characteristics of the Project Sample and Core Sample

The Project Sample (N=285) included children with typical and atypical development from 2-44 months old (Mean = 21.28, SD = 11.29), distributed across the seven age categories as described in Chapter 3. As stated earlier, the oldest age group (37-44 months) was included to demonstrate that the Early LAP is not appropriate for older children (unless they are functioning below their chronological age). Of the 26 children in the oldest age group with typical development, the mean developmental age scores for each domain ranged from 35.0 to 48.6, with an average of 81% of the children completing the assessment before reaching a ceiling in the domain, confirming that the Early LAP is not an appropriate instrument for children with typical development above 36 months of age.

To establish a Core Sample of children with typical development in the birth to 36 month age range, the scores of children in the oldest age group (37- 44 months old) and the children with professionally diagnosed disabilities were excluded. Thus, the resulting Core Sample (N=242) was comprised of children with presumedly typical development from 2-36 months old (Mean = 18.97, SD = 9.77). Table 7 presents the means and standard deviations for the Early LAP developmental age domain scores and for each of the age categories in the Project Sample and Core Sample.

Table 7. Means and Standard Deviations of Early LAP Developmental Age Domain Scores by Chronological Age Clusters for Project Sample and Core Sample

Domain/ Chronological Age Clusters	Project Sample*		Core Sample**	
	Mean	Standard Deviation	Mean	Standard Deviation
Gross Motor				
Birth -6 months	4.41	1.65	4.45	1.66
7-12 months	10.44	3.42	10.95	2.90
13-18 months	17.61	5.64	17.88	5.41
19-24 months	24.53	1.65	25.05	5.24
25-30 months	29.68	6.03	29.51	6.02
31-36 months	32.00	7.75	33.56	4.48
37-44 months	34.11	4.79	--	--
Fine Motor				
Birth -6 months	4.76	2.05	4.78	2.08
7-12 months	10.34	2.40	10.73	1.94
13-18 months	16.05	3.08	16.16	2.96
19-24 months	22.56	4.64	22.91	4.46
25-30 months	28.49	5.08	28.60	5.09
31-36 months	31.45	7.49	33.35	3.60
37-44 months	34.10	3.82	--	–
Cognitive				
Birth -6 months	5.03	1.69	5.03	1.69
7-12 months	10.19	2.38	10.48	2.14
13-18 months	15.60	3.66	15.77	3.51
19-24 months	22.62	4.50	22.69	4.59
25-30 months	28.53	5.75	28.81	5.55
31-36 months	32.51	7.09	34.37	2.94
37-44 months	34.30	3.91	--	–
Language				
Birth -6 months				1.27
7-12 months				3.13
13-18 months				4.29
19-24 months	5.00	1.26	5.03	5.60
25-30 months	9.98	3.28	10.28	6.07
31-36 months	16.36	4.53	16.60	3.93
37-44 months	22.96	5.85	23.41	–
	28.66	6.82	29.19	
	32.31	7.86	34.20	
	33.70	4.64	–	
Domain/	Project Sample*		Core Sample**	

Chronological Age Clusters	Mean	Standard Deviation	Mean	Standard Deviation
Self Help				
Birth -6 months	6.78	.67	6.78	.67
7-12 months	12.58	2.69	12.76	2.53
13-18 months	16.69	3.05	16.80	2.99
19-24 months	23.19	5.16	23.61	4.96
25-30 months	28.37	5.46	28.65	5.25
31-36 months	32.53	6.43	33.77	5.06
37-44 months	45.79	12.13	--	--
Social Emotional				
Birth -6 months	4.61	1.61	4.61	1.61
7-12 months	10.32	3.08	10.63	2.92
13-18 months	17.45	7.24	17.62	7.26
19-24 months	28.03	7.47	28.67	6.91
25-30 months	31.97	5.80	32.50	5.02
31-36 months	33.03	7.63	34.96	2.88
37-44 months	36.00	.00	–	--

*Project Sample N= 285 (2 - 44 months old with typical and atypical development)

**Core Sample N=242 (2-36 months old with typical development)

Procedures

The data were collected by a team of nine professionals trained and supervised by the project co-directors. Each examiner had a master's or doctoral degree in education, early intervention services, or another related field. The examiners participated in a two-and-a-half day training session on the Early LAP, BSID-II, and procedures for data collection in the fall of 1999.

A total of 285 children participated in the study with approximately equal numbers of children recruited from each geographic region. Within each site, children were recruited from child care centers and individual families. An effort was made to include settings representing children from a range of socioeconomic groups. Each center director, or parent in the case of home settings, was contacted by phone and invited to participate in the study. Copies of the Early LAP manual and letters describing the study and requesting consent to participate were shared and discussed during a subsequent meeting. In the case of child care and Early Head Start programs, each center director agreed to distribute and collect permission forms for parents interested in participating in the study.

The data were collected from November, 1999 to July, 2000. After the children were recruited, each examiner was responsible for scheduling assessment visits with center directors or individual families within home settings, completing the assessments, and submitting completed protocols to the project co-directors.

After the data collection was completed, the individual item scores were entered into a database. Once all data had been entered, each item was verified against the original protocol independently by two different individuals, and all errors were reconciled and corrected in the database. An analysis data set based on the final database was then programmed in SAS 8.1 and then converted into SPSS 8.0. Statistical analyses were generated in SPSS 8.0 for each component of the study.

Reliability

The reliability of an assessment instrument refers to its accuracy and consistency over time. For example, an assessment instrument should produce roughly the same results when the same individuals are tested under similar conditions within a short period of time. Analyses of the reliability for each domain of the Early LAP, including correlations with age, internal consistency, standard error of measurement, test-retest reliability, and interrater reliability were conducted. Although children were recruited for the sample in 6-month increments of age, for purposes of analysis the data were collapsed into 12-month increments to ensure sufficient sample sizes in each cell. Every effort was made to gather complete data for each child, however, in some cases there were missing items that prevented calculation of a domain score. In most cases, the missing data were caused by the inability to observe particular behaviors due to the inaccessibility of large materials (e.g., furniture, stairway) or a restricted number of test items in a developmental range (e.g., self-help begins at six months and social emotional has a limited number of items for some age ranges).

Correlations Between Chronological Age and Developmental Age Scores

The correlations between the Early LAP developmental age scores and chronological age were computed for the Core Sample (children with typical development in the birth to 36 month age range) using Pearson product-moment correlation coefficients (*r)*. Table 8 presents the correlation coefficients by domain and age group. These results indicate strong correlations (.90 to .95) between chronological age and developmental age in each domain for the overall sample. Within age groups, correlations for the 2-12 month old age range (.80 to .91) and the 13-24 month old age range (.63 to .73) also indicate strong relations between chronological age and developmental age for each domain. The correlations for the 25-36 month old age range (.31 to .61) are moderate. These findings suggest that the developmental age scores on the Early LAP are reliably associated with chronological age for younger children, but that the association decreases somewhat as children become older and begin to age out of certain items and/or domains on the test.

Table 8. Correlations Between Chronological Age and Developmental Age Scores by Domain and Age Group

DOMAINS	2 months to 12 months[a]	13 months to 24 months[b]	25 months to 36 months[c]	Total[d]
Gross Motor	.88	.63	.45	.91
Fine Motor	.91	.72	.51	.94
Cognitive	.92	.73	.61	.95
Language	.82	.70	.43	.92
Self-Help	.80	.73	.60	.91
Social Emotional	.85	.68	.31	.90

Note: For all correlations, $p < .01$ N: a (GM=73, FM=73, C=73, L=68, SH=47, SE=69) b (GM=85, FM=86, C=88, L=87, SH=85, SE=75) c (GM=71, FM=74, C=72, L=72, SH=68, SE=57) d (GM=229, FM=233, C=233, L=227, SH=200, SE=201)

Internal Consistency

The internal consistency of the Early LAP was examined to determine how well the items in each domain relate to one another. The internal consistency coefficient indicates how effectively the individual domain scores on the Early LAP are measuring defined constructs (e.g., gross motor, fine motor, cognitive skills). The higher the value, the greater the consistency of items within the domain. Cronbach's coefficient alpha was used to calculate the internal consistency of each domain for the total Core Sample (N=242) by age groups. All items before the basal were counted as correct and all items above the ceiling were counted as incorrect.

Table 9 presents the results of the internal consistency analyses. The alpha coefficients for the total Core Sample (.96 to .99) indicate strong internal consistency for each domain. The alpha coefficients for the individual age groups are also quite high (.84-.98). These results indicate that the Early LAP items show strong internal consistency within each domain.

Table 9. Internal Consistency of Early LAP Developmental Age Domain Scores by Age Group

DOMAINS	2 months to 12 months[a]	13 months to 24 months[b]	25 months to 36 months[c]	Total[d]
Gross Motor	.98	.97	.84	.99
Fine Motor	.96	.94	.90	.98
Cognitive	.97	.96	.96	.99
Language	.91	.96	.95	.98
Self-Help	.97	.96	.93	.98
Social Emotional	.91	.91	.87	.96

Note: For all correlations, $p < .01$ N: a (GM=75, FM=75, C=75, L=75, SH=72, SE=74) b (GM=80, FM=88, C=88, L=88, SH=87, SE=84) c (GM=62, FM=75, C=72, L=73, SH=72, SE=72) d (GM=217, FM=238, C=235, L=236, SH=231, SE=230)

Standard Errors of Measurement

The Standard Error of Measurement (SE_M) provides an estimate of the amount of error between an individual's observed score and the true score. The SE_M has an inverse relationship with reliability so that as reliability increases, the SE_M decreases, indicating greater confidence in the accuracy of the observed scores. SE_M's were calculated for each domain of the Core Sample (N=242) by the following formula, $SE_M = s \sqrt{1 - r}$, where SE_M is the standard error of measurement, s is the standard deviation of the observed scores, and r is the reliability of the assessment instrument. The internal consistency reliability coefficients reported in the previous section were used to calculate the SE_M. Table 10 presents the SE_M's for each domain of the Early LAP by age group. The results of each of these calculations produced fairly small SE_M's, indicating a high degree of confidence that the observed scores on the Early LAP will provide an accurate representation of an individual's skills.

Table 10. Standard Errors of Measurement of Early LAP Developmental Age Domain Scores by Age Group

DOMAINS	2 months to 12 months[a]	13 months to 24 months[b]	25 months to 36 months[c]	Total[d]
Gross Motor	0.57	1.11	2.27	1.08
Fine Motor	0.71	1.24	1.59	1.43
Cognitive	0.58	1.07	1.05	1.04
Language	1.08	1.21	1.27	1.51
Self-Help	0.57	1.07	1.52	1.26
Social Emotional	1.16	2.69	1.57	2.41

Note: For all correlations, p <. 01 N: a (GM=73, FM=73, C=73, L=68, SH=47, SE=69) b (GM=85, FM=86, C=88, L=87, SH=85, SE=75) c (GM=71, FM=74, C=72, L=72, SH=68, SE=57) d (GM=229, FM=233, C=233, L=227, SH=200, SE=201)

Test-Retest Reliability

Test-retest reliability indicates the extent to which scores on an assessment instrument are consistent from one time period to the next. Because the Early LAP measures a continuum of developmental skills, the test-retest reliability was measured over a short period of time so that any differences between administrations were more likely to reflect reliability rather than individual development. Therefore, the Early LAP was administered by the same examiner on two separate occasions one to three weeks apart for a subset of children from the overall Project Sample (Test-Retest Sample). The Test-Retest Sample was composed of 92 children from 2 to 44 months old (Mean = 18.55, SD = 9.74), including both typically and atypically developing children. The sample consisted of 48.9% females and 51.1% males, and was 15.2% African American, 3.3% Asian and Pacific Islander, 15.2% Hispanic origin, 62% White, and 4.3% Other racial/ethnic origins. Test-retest reliability was determined by calculating the correlations between domain scores from the first and the second test administrations using Pearson's *r*. Table 11 presents the means and standard deviations for the first and second test scores and the test-retest correlation coefficients for each domain. The resulting correlations (.96 to .99) indicate a high degree of stability in individual test scores over short intervals of time.

Table 11. Means, Standard Deviations, and Correlations of Early LAP Developmental Age Domain Scores for Test-Retest Reliability Sample

	First Testing		Second Testing		
DOMAINS	Mean	SD	Mean	SD	*r*
Gross Motor	19.16	11.33	19.15	11.65	.98
Fine Motor	19.03	10.13	19.81	10.72	.97
Cognitive	18.59	10.43	19.26	10.52	.99
Language	18.93	10.44	18.57	10.52	.99
Self-Help	20.91	9.29	20.94	9.41	.96
Social Emotional	19.85	11.66	20.20	11.90	.99

Note: For all correlations, $p < .01$

N: GM=79, FM=85, C=84, L=81, SH =72, SE =69

Interrater Reliability

Interrater reliability measures the extent to which different examiners achieve the same results when independently assessing the same child. The results of the assessment should reflect the developmental skills of the child independent of the particular person administering the test, assuming proper procedures have been followed. In order to determine the level of interrater reliability, the Early LAP was administered to a subset of children from the overall Project Sample by two different examiners on two separate occasions one to three weeks apart (Interrater Reliability Sample). The Interrater Reliability Sample was comprised of 49 children from 2 to 43 months old (Mean = 20.80, SD = 10.27), including 40.8% females and 59.2% males, and was 22.4% African American; 4.1% American Indian, Eskimo, and Aleut; 4.1% Asian and Pacific Islander, 4.1% Hispanic origin, 63.3% White; and 2.0% Other racial/ethnic origins.

Interrater reliability was determined by computing the correlations between the domain scores from the test administrations by two different examiners using Pearson's *r*. Table 12 presents the means and standard deviations for both test administrations and the interrater reliability correlation coefficients for each domain. The resulting correlations indicate a high degree of reliability (.96 to .99) when the Early LAP is administered by two different examiners.

Table 12. Means, Standard Deviations, and Correlations of Early LAP Developmental Domain Age Scores for Interrater Reliability Sample

	First Testing		Second Testing		
DOMAINS	Mean	SD	Mean	SD	*r*
Gross Motor	22.63	10.41	22.02	10.02	.97
Fine Motor	21.46	10.06	21.74	9.89	.99
Cognitive	21.42	10.37	21.98	10.32	.99
Language	22.63	10.37	22.89	10.28	.96
Self-Help	24.17	10.78	23.27	9.03	.96
Social Emotional	23.45	12.06	24.07	11.60	.98

Note: For all correlations, $p < .01$

N: GM=44, FM=47, C=47, L=46, SH=39, SE=43

Validity

The foremost authoritative reference on validity and other test matters, the *1999 Standards for Educational and Psychological Testing*, defines validity as, "The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test." (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, p.184). This definition emphasizes that inferences derived from test scores give meaning to them beyond simply reporting numbers. Four types of analyses are recognized by the *1999 Standards for Educational and Psychological Testing* as demonstrating the validity of test score inferences: (1) construct-related evidence; (2) content-related evidence; (3) predictive evidence; and (4) concurrent evidence. Three of these types of validity analyses are presented below: construct validity, and criterion validity which were conducted as part of this study, and content validity which was conducted in a previous study.

Construct Validity

Evidence of construct validity can be inferred by examining the intercorrelations among different areas of an assessment instrument. Thus, to examine the extent to which the different domains measure different skills, the intercorrelations among domains were calculated. High correlations among areas would suggest that they are measuring similar underlying constructs, while low correlations would suggest that they are measuring different underlying constructs. Domains that are more strongly related conceptually and that have more items in common would be expected to have relatively stronger intercorrelations. Zero-order correlations using Pearson's *r* were calculated between developmental age domain scores for the Core Sample (N=242) as shown below the diagonal in Table 13. While these high positive correlations (.87 to .97) potentially indicate a single underlying construct, because these zero order correlations were calculated across age groups, they also indicate differences in skill performance as a result of age.

To separate these two elements, partial correlations controlling for age were calculated between developmental age domain scores as depicted above the diagonal in Table 13. The magnitudes of

the partial correlation coefficients are substantially smaller than the zero-order correlations (.27 to .76) in the modest to moderate range. The relatively higher correlations among the fine motor, cognitive, and language domains are likely a result of a number of shared items, while the less conceptually related domains evidence lower correlations. These results suggest that while the different domains of the Early LAP are somewhat related, they are also measuring somewhat independent aspects of development.

Table 13. Zero-order Correlations and Partial Correlations Controlling for Age Among Early LAP Domains

DOMAINS	Gross Motor	Fine Motor	Cognitive	Language	Self-Help	Social Emotional
Gross Motor		.50	.53	.51	.31	.58
Fine Motor	.93		.76	.57	.33	.39
Cognitive	.93	.97		.71	.48	.34
Language	.92	.94	.96		.27	.44
Self-Help	.88	.91	.93	.88		.27
Social Emotional	.93	.97	.90	.91	.87	

Note: For all correlations, $p < .01$

N: GM=229, FM=233, C=233, L=227, SH=200, SE=201

Criterion Validity

Criterion validity (also known as concurrent validity) is the extent to which individual scores on one test correspond to scores on an established test of similar constructs. These two tests must be administered consecutively, so as to minimize differences due to development or other variations in test conditions. The established test is the criterion used to validate the new test (Gall, Borg, & Gall,1996). In this study, the correspondence between the Early LAP and the Mental and Motor Scales of the BSID-II was examined to investigate the criterion validity of the Early LAP. The Core Sample (N=242) was administered both the Early LAP and the BSID-II Mental and Motor Scales during the same testing session or in two sessions in close proximity. Criterion validity was determined by examining the correlations using Pearson's *r* between the Early LAP developmental age domain scores and the BSID-II Mental and Motor Scale developmental age scores for conceptually related areas.

Table 14 presents the correlations between the developmental age scores for the Early LAP domains and the BSID-II Mental and Motor Scales by age group. The results indicate a strong correlation (.90 to .97) between the Early LAP and BSID-II scores in each domain for the overall sample. Fairly high correlations were found within the 2-12 month old age range (.83 to .95) and the 13-24 month old age range (.72 to .88). The correlations for the 25-36 month old age range (.47 to .83) were somewhat lower, particularly in the self-help and social emotional domains. These somewhat lower correlations for the oldest age group may reflect some aging out of the Early LAP assessment for these children, similar to the analyses discussed earlier.

Table 14. Correlations Between the Early LAP Domains and the BSID-II Mental and Motor Scales for Developmental Age Scores by Age Group

	2 months to 12 months[a]		13 months to 24 months[b]		25 months to 36 months[c]		Total[d]	
Early LAP DOMAINS	BSID-II Mental	BSID-II Motor	BSID-II Mental	BSID-II Motor	BSID-II Mental	BSID-II Motor	BSID-II Mental	BSID-II Motor
Gross Motor		.95		.72		.61		.92
Fine Motor		.90		.72		.66		.94
Cognitive	.93		.88		.83		.97	
Language	.87		.88		.75		.96	
Self-Help	.83		.83		.58		.91	
Social Emotional	.85		.72		.47		.90	

Note: For all correlations, $p < .01$ N: a (GM=73, FM=73, C=73, L=68, SH=47, SE=69) b (GM=85, FM=86, C=88, L=87, SH=85, SE=75) c (GM=71, FM=74, C=72, L=72, SH=68, SE=57) d (GM=229, FM=233, C=233, L=227, SH=200, SE=201)

Content Validity

Content validity examines the extent to which the scores on an assessment actually represent the content they purport to measure. Content validity is determined through a systematic examination of an assessment instrument by content experts. Fleming (2000) conducted a study at Johns Hopkins University that included a content validity examination of each item on the Early LAP. Four different experts evaluated the content of the Early LAP both in terms of the developmental ages assigned to items and the representativeness of the items for the intended content areas. Experts examined the Early LAP items in comparison with four other standardized measures and with four widely used textbooks on infant and toddler development. The results of this study indicated that all of the items were representative of the skills tested and that 387 of the 414 items (93%) on the Early LAP were appropriately categorized by developmental age. The distribution of the remaining 27 items in which the experts disagreed with the developmental age categories on the Early LAP included nine items in the gross motor domain, eleven items in the fine motor domain, one item in the cognitive domain, two items in the language domain, four items in the self-help domain, and none in the social emotional domain. The discrepancies between the Early LAP classification and the experts' opinions for gross motor items differed by only one month and by no more than four months for fine motor. Discrepancies for the other domains differed by six to eight months. Although there were no differences in the social emotional domain, Fleming noted that some developmental age ranges had very few items. In sum, the Early LAP was found to have good content validity.

Children With Disabilities

The Atypical Development Sample was composed of a subset of 17 children who had been professionally diagnosed as having disabilities prior to this study. These children ranged from
6 to 43 months old (Mean = 25.47, SD = 12.04), were 29.4% females and 70.6% males, and were 11.8% African American, 23.5% Hispanic origin, 58.8% White, and 5.9% Other racial/ethnic origins. The distribution of children across geographic areas was 17.6% from the Northeast, 35.3% from the South, 11.8% from the North Central, and 35.3% from the West. Of the 17 children in the sample, 10 children had developmental delays, five children had motor disabilities, and two children had speech and language disabilities. Table 15 depicts the means, standard deviations, and correlations with chronological age (using Pearson's *r*) for each domain for the Atypical Development Sample. As expected, the mean developmental age scores for each domain are substantially lower than the children's chronological ages, and the correlations between developmental age scores and chronological age are substantially lower than the correlations for children with typical development (See Table 8). These results suggest that the Early LAP discriminates children's skill levels independently of their age, and that it can be used effectively to assess the developmental skills of children with disabilities.

Table 15. Means, Standard Deviations, and Domain Correlations of Early LAP Developmental Age Domain Scores for Atypical Development Sample

DOMAINS	Means	SD	*r*
Gross Motor	17.29	13.31	.70
Fine Motor	16.29	10.59	.74
Cognitive	17.13	9.80	.72
Language	14.94	10.80	.69
Self-Help	19.42	8.67	.73
Social Emotional	18.00	13.03	.72

Note: For all correlations, $p < .01$
N: GM=17, FM=17, C=15, L=17, SH=14, SE=13

Concluding Remarks

Overall, this research found the Early LAP to be reliable and valid in assessing the development of young children. The Early LAP was found to have relatively high correlations between developmental age domain scores and chronological age, especially for children in the birth to two-year-old range, while older children aged out on some items and/or domains. The Early LAP also evidenced good internal consistency and fairly low standard errors of measurement for each domain. Very good test-retest reliability and

interrater reliability were found for all domains of the Early LAP. Evidence of adequate construct validity was also shown. The Early LAP was found to have very good criterion validity, based on comparisons with the Bayley Scales of Infant Development–Second Edition. A separate study indicated that the Early LAP also demonstrated good content validity (Fleming, 2000). In sum, the Early LAP evidences good reliability and validity characteristics, and is an appropriate tool for use in assessing young children's developmental functioning.

Administration

- Administration time takes about 1-1½ hours. Domains may be administered in more than one session.
- Administered by trained professionals or paraprofessionals
- Administered at specific intervals or ongoing to monitor progress

Scoring and Interpretation
If the child meets the criteria of an item, a plus (+) should be recorded to indicate the presence of the criterion-referenced behavior. A minus (-) is recorded if the skill is not demonstrated by the child.

Basal and Ceiling Criteria for Early LAP

Basal	**8 consecutive items successfully completed**
Ceiling	**3 errors out of 5 consecutive items**

Early LAP results indicate both mastered and emerging skills in each developmental domain. Although an approximate developmental age can be calculated for each domain, it should not be used as the sole criterion to diagnose a child for early intervention services. Rather, it should be used as supporting evidence in conjunction with other instruments that are norm-referenced.