Statistical Properties of *LAP-3*

.

Statistical Properties of the Project Sample

The Project Sample (N=363) included children with typical and atypical development from 30-78 months old (Mean = 54.3, SD = 10.6), distributed across the eight age categories as described in Chapter 3. The mean raw scores across domains for the Core Sample (children with typical development from 36-72 months old) ranged from 25.2 to 53.4. As stated earlier, the youngest (30-35 months) and oldest (73-78 months) age groups were included to demonstrate that the *LAP-3* is not appropriate for children younger or older than the 36-72 month old age range (unless they are functioning below their chronological age). Of the 10 children with typical development in the youngest age group (M=33.8, SD=1.40), the mean raw scores ranged from 11.0 to 27.8 for each domain. While most of the children in the youngest group were able to establish a basal, the small number of items in the lower ranges of each domain may provide a less accurate assessment of their functioning. We recommend using a measure designed for younger children, such as the Early Learning Accomplishment Profile (Early LAP), to provide a more appropriate assessment for children functioning below 36 months.

Of the 11 children in the oldest age group with typical development (M=75.8, SD=2.04), the mean raw scores for each domain ranged from 36.3 to 87.2, with 100% of the children completing the assessment before reaching a ceiling in five domains and an average of 81.8% unable to establish a ceiling in the remaining two domains, confirming that the LAP-3 is not an appropriate instrument for children with typical development above 72 months of age.

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 *LAP-3*, including correlations with age, internal consistency, standard error of measurement, test-retest reliability, and interrater reliability were conducted.

Correlations Between Chronological Age and LAP-3 Raw Scores

The correlations between the *LAP-3* raw scores and chronological ages were computed for the Core Sample (children with typical development in the 36-72 month age range) using Pearson product-moment correlation coefficients (*r)*. Table 9 presents the means, standard deviations, and correlation coefficients by domain for the Core Sample. These results indicate strong correlations (.77 to .84) between chronological age and the raw scores for six of the seven domains and moderate correlations (.61) for the Personal/Social domain. These results suggest that the raw scores on the *LAP-3* are reliably associated with chronological age, so that older children are likely to obtain higher scores than younger children. It should be noted that the number of items in each domain varies, and therefore the means and ranges will vary accordingly (see page 20).

Table 9. Means, Standard Deviations, and Domain/Chronological Age Correlations of *LAP-3* Raw Scores for Core Sample (N=251)

DOMAINS	Means	SD	*r*
Gross Motor	38.90	12.47	.84
Fine Motor	31.29	9.17	.82
Pre-Writing	25.25	9.20	.82
Cognitive	53.41	24.08	.82
Language	44.96	15.17	.77
Self-Help	44.50	10.94	.78
Personal/Social	36.64	8.38	.61

Note: For all correlations, $p < .01$
N: GM=245, FM=241, PW=243, C=246, L=242, SH=243, PS=229

Internal Consistency

The internal consistency of the *LAP-3* was examined to determine how well the items within each domain relate to one another. The internal consistency coefficient indicates how effectively the individual domain scores on the *LAP-3* are measuring defined constructs (e.g., gross motor, fine motor, cognitive skills). The higher the value, the greater was 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=251) by age groups. All items before the basal were counted as correct and all items above the ceiling were counted as incorrect for calculating the internal consistency coefficients.

Table 10 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 (.78 to .98). These results indicate that the *LAP-3* items show strong internal consistency within each domain across the various age groups covered by this measure.

Table 10. Internal Consistency of *LAP-3* Raw Scores by Age Group (N=251)

DOMAINS	36-41[a] months	42-47[b] months	48-53[c] months	54-59[d] months	60-65[e] months	66-72[f] months	Total[g]
Gross Motor	.92	.90	.94	.89	.94	.89	.97
Fine Motor	.89	.90	.92	.91	.91	.92	.96
Pre-Writing	.90	.92	.91	.92	.91	.89	.96
Cognitive	.94	.97	.98	.97	.97	.95	.99
Language	.91	.95	.95	.91	.95	.97	.97
Self-Help	.92	.92	.90	.93	.91	.91	.96
Personal/Social	.95	.93	.93	.78	.95	.85	.96

Note: For all correlations, $p < .01$
N: a (GM=15, FM=21, PW=22, C=21, L=21, SH=16, PS=17)
b (GM=21, FM=35, PW=36, C=35, L=32, SH=31, PS=29)
c (GM=18, FM=32, PW=37, C=36, L=37, SH=29, PS=25
d (GM=23, FM=23, PW=30, C=31, L=25, SH=19, PS=20)
e (GM=25, FM=24, PW=28, C=28 L=24, SH=21, PS=19)
f (GM=28, FM=31, PW=35, C=35, L=28, SH=14, PS=30)
g (GM=130, FM=166, PW= 188, C=186, L=167, SH=130, PS=140)

Standard Errors of Measurement

The Standard Error of Measurement (SEM) provides an estimate of the amount of error between an individual's observed score and the true score. The SEM has an inverse relationship with reliability so that as reliability increases the SEM decreases, indicating greater confidence in the accuracy of the observed scores. SEM's were calculated for each domain of the Core Sample (N=251) using the following formula, $SEM = s\sqrt{1 - r}$, where SEM 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 SEM's. Table 11 presents the SEM's for each domain of the LAP-3 by age group. Because any observed score includes some measurement error, these

Table 11. Standard Errors of Measurement of LAP-3 Raw Scores by Domain and Age Group (N=251)

DOMAINS	36-41[a] months	42-47[b] months	48-53[c] months	54-59[d] months	60-65[e] months	66-72[f] months	Total[d]
Gross Motor	1.76	2.17	2.08	2.27	1.72	1.52	2.16
Fine Motor	1.89	1.70	1.88	1.54	1.25	.91	1.83
Pre-Writing	1.63	1.83	1.73	1.61	1.42	1.19	1.84
Cognitive	2.40	2.61	2.31	3.05	2.30	1.45	2.41
Language	2.12	2.30	2.29	1.57	1.82	2.03	2.63
Self Help	2.68	2.20	1.94	1.55	1.23	3.13	2.19
Personal/Social	1.90	2.12	1.69	1.76	1.41	1.41	1.68

Note: For all correlations, $p < .01$
N: a (GM=15, FM=21, PW=22, C=21, L=21, SH=16, PS=17)
b (GM=21, FM=35, PW=36, C=35, L=32, SH=31, PS=29)
c (GM=18, FM=32, PW=37, C=36, L=37, SH=29, PS=25
d (GM=23, FM=23, PW=30, C=31, L=25, SH=19, PS=20)
e (GM=25, FM=24, PW=28, C=28 L=24, SH=21, PS=19)
f (GM=28, FM=31, PW=35, C=35, L=28, SH=14, PS=30)
g (GM=130, FM=166, PW= 188, C=186, L=167, SH=130, PS=140)

SEM's can be used to determine confidence intervals indicating the range within which a child's true score is likely to fall, based on the child's observed score and the SEM. For example, we can be 95% confident that the child's true score will be within the range of scores

indicated by the 95% confidence interval. Confidence intervals can be determined at different levels, based on standard formulas, with larger ranges for wider confidence intervals. The formula for calculating the 95% confidence interval is *observed score* $\pm$ *1.96 x SEM*, while the formula for the 99% confidence interval is *observed score* $\pm$ *2.58 x SEM.*

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 *LAP-3* 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 *LAP-3* 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 40 children from 37 to 72 months old (Mean = 57.00, SD = 10.19), including both typically and atypically developing children. The sample consisted of 55% females and 45% males, and was 5% African American, 5% Asian and Pacific Islander, 5% Hispanic origin, 65% White, and 15% "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 12 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) demonstrate very good test-retest reliability, indicating a high degree of stability in individual test scores over short intervals of time.

Table 12. Means, Standard Deviations, and Correlations of *LAP-3* Raw Scores by Domain for Test-Retest Reliability Sample (N=40)

DOMAINS	First Testing		Second Testing		
	Mean	SD	Mean	SD	*r*
Gross Motor	41.15	13.40	42.15	12.24	.96
Fine Motor	32.97	9.17	33.22	9.59	.98
Pre-Writing	27.18	9.35	27.65	9.12	.99
Cognitive	55.72	24.35	57.71	23.87	.98
Language	47.58	17.07	50.10	16.83	.96
Self-Help	45.90	10.60	46.28	10.19	.99
Personal/Social	38.11	7.52	37.97	8.33	.97

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

N: GM=39, FM=37, PW=40, C=39, L=40, SH =39, PS =38

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 *LAP-3* 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 33 children from 33 to 73 months old (Mean = 50.33, SD = 11.74), including 51.5% females and 48.5% males, and was 18.2% African American; 9.1% Asian and Pacific Islander, 6.1% Hispanic origin, 60.6% White; and 6.1% "Other" racial/ethnic origins.

Interrater reliability was determined by computing the correlations between the domain scores from the two test administrations by different examiners using Pearson's *r*. Table 13 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 (.84 to .98) when the *LAP-3* is administered by two different examiners.

Table 13. Means, Standard Deviations, and Correlations of *LAP-3* Raw Scores by Domain for Interrater Reliability Sample (N=33)

DOMAINS	First Testing		Second Testing		*R*
	Mean	SD	Mean	SD	
Gross Motor	35.94	14.81	33.91	14.64	.89
Fine Motor	27.55	11.00	27.55	11.45	.95
Pre-Writing	23.24	10.70	23.06	10.40	.97
Cognitive	46.61	28.57	47.34	28.89	.94
Language	40.41	16.22	42.31	17.65	.93
Self-Help	41.09	12.39	40.92	12.93	.84
Personal/Social	34.89	10.49	35.55	11.78	.98

Note: For all correlations, $p < .01$
N: GM=31, FM=33, PW=33, C=33, L=32, SH=32, SE=28

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. Two of these types of validity analyses are presented below: construct validity and criterion validity. Information about the content validity study can be found in Chapter 2.

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 raw scores for each domain for the Core Sample (N=251), as shown below the diagonal in Table 13. While these high positive correlations (.61 to .89) 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 domain raw scores, as depicted above the diagonal in Table 14. The magnitudes of the partial correlation coefficients are substantially smaller than the zero-order correlations (.26 to .57), in the modest to moderate range. These results suggest that while the different domains of the *LAP-3* are somewhat related, they are also measuring somewhat independent aspects of development.

Table 14. Zero-order Correlations[1] and Partial Correlations[2] Controlling for Age Among *LAP-3* Domains (N=251)

DOMAINS	Gross Motor	Fine Motor	Pre-Writing	Cognitive	Language	Self-Help	Personal/ Social
Gross Motor		.31	.30	.36	.33	.33	.26
Fine Motor	.79		.62	.56	.41	.52	.40
Pre-Writing	.80	.89		.56	.44	.44	.40
Cognitive	.80	.76	.86		.57	.36	.41
Language	.76	.78	.80	.85		.39	.48
Self-Help	.77	.83	.80	.77	.76		.47
Personal/Social	.61	.68	.68	.68	.71	.71	

Note: For all correlations, p < . 01
N: GM=245, FM=241, PW=243, C=246, L=242, SH=243, PS=229

1=Zero-order Correlations below diagonal.
2=Partial Correlations above diagonal.

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 *LAP-3* and the Battelle Developmental Inventory (BDI) was examined to investigate the criterion validity of the *LAP-3*. Of the Core Sample, 230 children (91.6%) were administered both the *LAP-3* and the BDI, either 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 *LAP-3* domain raw scores and the BDI total component raw scores for conceptually related areas. Table 15 presents the correlations between the raw scores for the *LAP-3* domains and the BDI components by age group. The results indicate fairly strong correlations between the *LAP-3* and BDI scores. Seventy-six percent of the domains had correlations between .70 to .92. The remaining 24% had correlations in the .54 to .69 range, and were primarily related to the Communication Domain on the BDI and the Personal/Social Domain on the LAP-3.

Table 15. Correlations Between *LAP-3* Raw Scores and BDI Total Component Raw Scores by Domain (N=230)

LAP-3 Domains	Gross Motor	Fine Motor	Pre-Writing	Cognitive	Language	Self Help	Personal/ Social
BDI Component Totals							
Personal/Social[a]	.68	.70	.70	.72	.76	.78	.88
Adaptive[b]	.73	.71	.71	.72	.76	.79	.70
Gross Motor[c]	.81	.77	.77	.75	.68	.72	.54
Fine Motor[d]	.80	.87	.92	.85	.80	.79	.66
Communication[e]	.64	.66	.69	.78	.82	.69	.68
Cognitive[f]	.76	.82	.84	.91	.86	.77	.68

Note: For all correlations, p < . 01

N: a (GM=191, FM=186, PW=189, C=192, L=190, SH=188, PS=180) b (GM=180, FM=178, PW=180, C=181, L=178, SH=178, PS=169)
c (GM=199, FM=193, PW=196, C=200, L=196, SH=196, PS=191) d (GM=192, FM=189, PW=190, C=193, L=191, SH=190, PS=181)
e (GM=175, FM=173, PW=174, C=176, L=174, SH=173, PS=167) f (GM=178, FM=175, PW=177, C=178, L=176, SH=175, PS=167)

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. As discussed earlier, a content or face validity study was conducted on the LAP-3 and adjustments made in accordance with the results of the review.

Children With Disabilities

Because the *LAP-3* is sometimes used in conjunction with standardized instruments to examine the skill development of children with developmental delays or diagnosed disabilities, a subsample of 28 children with disabilities (9.3%) 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 special education services. These children ranged in age from 33 to 73 months old (Mean = 55.21, SD = 11.26), were 39.3% females and 60.7% males, and were 10.7% African American, 14.3% Asian and Pacific Islander, 10.7% Hispanic origin, 53.6% White, and 10.7% "Other" racial/ethnic origins. The distribution of children across geographic areas was 7.1% from the Northeast, 25.0% from the South, 25.0% from the Midwest, and 42.9% from the Northwest. Of the 28 children in the sample, eight children had developmental delays, two children had motor disabilities, and seven children had speech and language disabilities, three children had Autism, one child had ADHD, and seven had multiple disabilities. 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 adaptive equipment for child with limited mobility). It is important to note that the information gathered through the *LAP-3* may be beneficial for older children functioning in the 36-72 month age range.

Table 16 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 raw scores for each domain are substantially lower than the mean of the children's chronological ages, and the correlations between raw scores and chronological age are substantially lower than the correlations for children with typical development (See Table 9). These results suggest that the *LAP-3* 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 *LAP-3* Raw Scores for Atypical Development Sample (N=28)

DOMAINS	Means	SD	*R*
Gross Motor	27.11	12.31	.33
Fine Motor	22.18	10.70	.42*
Pre-Writing	16.18	8.69	.63**
Cognitive	30.30	21.11	.61**
Language	32.54	17.76	.37
Self-Help	33.89	13.34	.54**
Personal/Social	25.52	11.25	.52**

Note: *Correlations significant at, p < . 05
**Correlations significant at, p < . 01
N: GM=27, FM=28, PW=27, C=26, L=26, SH=27, SE=27

Concluding Remarks

Overall, this research found the *LAP-3* to be reliable and valid in assessing the development of young children. The *LAP-3* was found to have relatively high correlations between raw domain scores and chronological age for children in the 36-72 month age range, while older children aged out on most domains. The *LAP-3* 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 *LAP-3*. Evidence of adequate construct validity was also shown. The *LAP-3* was found to have very good criterion validity, based on comparisons with the Battelle Developmental Inventory. In sum, the *LAP-3* evidences good reliability and validity characteristics, and is an appropriate tool for use in assessing young children's developmental functioning in the 36-72 month old age range.

References

Allen, K.E., & Schwartz, I.S. (2001). *The Exceptional Child: Inclusion in Early Childhood Education*. New York: Delmar Publishers.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *1999 Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing.* Seventh Edition. Upper Saddle River, NJ: Prentice Hall.

Bagnato, S. J., Neisworth, J. T., & Munson, S. M. (1997). *LINKing assessment and early intervention: An authentic curriculum-based approach*. Baltimore, MD: Paul H. Brookes Publishing Co.

Baroody, A. J. (2000). "Does mathematics instruction for three- to five-year-olds really make sense?" Young Children 55(4): 61-67.

Bredekamp, S. & Copple, C. (Eds.) (1997). *Developmentally Appropriate Practice in Early Childhood Programs*. Washington, DC: National Association for the Education of Young Children.

Gall, M. D., Borg, W. R.,Gall, J. P. (1996). *Educational Research*. White Plains, NY: Longman Publishers USA.

Hardin, B.J., Lohr, L. G., Wesley, P. (1996). *Planning* for Success: *A Teacher's Guide to a New Planning Guide to the Preschool Curriculum.* Lewisville, NC: Kaplan Press.

Head Start Bureau. (2000). *ACYF-HS-IM-00-18*. Washington, DC: US Department of Health and Human Services, Administration for Children and Families.

Kowalski, K., K. Pretti-Frontczak, et al. (2001). "Preschool teachers' beliefs concerning the importance of various developmental skills and abilities." Journal of Research in Childhood Education 16(1): 5-14.

Martens, P. A. (1999). "Mommy, how do you write 'Sarah'?": The role of name writing in one child's literacy." Journal of Research in Childhood Education 14(1): 5-15.

Neuman, S.B., Copple, C., & Bredekamp, S. (2000). *Learning to Read and Write: Developmentally Appropriate Practices for Young Children*. Washington, DC: National Association for the Education of Young Children.

Neuman, S.B., & Dickinson, D. K. (2001). Handbook of Early Literacy Research. NY: The Guilford Press.

Newborg, J., Stock, J.R., Wnek, L., Guidubaldi, J., and Svinicki, J. (1984). *Battelle Developmental Inventory*. Chicago, IL: Riverside Publishing Company.

Ross, M. E. (2000). "Science Their Way." Young Children 55(2): 6-13.

Sanford, A.R., Zelman, J.G. (1995). *Learning Accomplishment Profile-Revised Edition (LAP-R)*. Lewisville, NC: Kaplan Press.

Sanford, A.R., Zelman, J.G., Hardin, B.J., & Peisner-Feinberg, E.S. (2003). *Learning Accomplishment Profile-Third Edition (LAP-3)*. Lewisville, NC: Kaplan Press.

U.S. Census Bureau. (1995). *Population Profile of the United States: 1995*. Washington, DC: U.S. Government Printing Office.