Technical Qualities of the Early LAP

Executive Summary

Overall, research found the Early LAP to be reliable and valid in assessing the development of young children. The Early LAP was found to have relative 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 and interrater reliability were found for all domains of the Early LAP. 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.

Reliability

Analysis of the reliability for each domain of the Early LAP, including correlations with age, internal consistency, standard error of measurement, test-retest reliability, and inter-rater reliability, were conducted as part of the field study.

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*). The results indicate strong correlations (.90 to .95) between chronological age and development age in each domain.

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 individual domain scores on the Early LAP are measuring defined constructs (e.g., gross motor, fine motor, cognitive skills). 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 to .98). These results indicate that the Early LAP items show strong internal consistency within each domain.

Standard Errors of Measurement

The Standard Error of Measurement (SE) provides an estimate of the amount of error between an individual's observed score and the true score. The Standard Error of Measurement has an inverse relationship with reliability so that as reliability increase the SE decreases, indicating greater confidence in the accuracy of the observed scores. SE's were calculated for each domain of the Core Sample, the results of each of these calculations produced fairly small SE'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.

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 difference between administrations were more likely to reflect reliability rather than individual development. Test-retest reliability was determined by calculating the correlations between domain scores from the first and the second test administrations using Pearson's *r*. The resulting correlations (.96 to .99) indicate a high degree of stability in the individual test scores over short intervals of time.

Interrater Reliability

Interrater reliability measures the extent to which different examiners achieve the same results when independently assessing the same child. For the Early LAP, interrater reliability was determined by computing the correlations between domain scores from the test administrations by two examiners using Pearson's *r*. The resulting correlations indicate a high degree of reliability (.96 to .99) when the Early LAP is administered by two different examiners.

Construct Validity

Evidence of construct validity can be inferred by examining the intercorrelations among different areas of an assessment instrument. 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 development age domains scores for the Core Sample. 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 difference in skill performance as a result of age.

To separate these two elements, partial correlations controlling for age were calculated between development age domain scores. The magnitudes of the partial correlation coefficients are substantially smaller than the zero-order correlation (.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.

Criterion Validity

Criterion 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 difference doe to development or other variations in test conditions. For the ELAP, the correspondence between the Early LAP and the Mental and Motor Scores of the BSID-II was examined

to investigate the criterion validity of the Early LAP. Criterion validity was determined by examining the correlations using Pearson's *r* between the Early LAP development age domain scores and the BSID-II Mental and Motor Scale developmental age scores for conceptually related areas.

Correlations between the developmental age scores for the Early LAP domains and the BSID-II Mental and Motor Scales by age group indicate a strong correlation (.90 to .97) 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 25-3 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.

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 development 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. These 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

An Atypical Development Sample was composed of children who had been professionally diagnosed as having disabilities prior to this study. Children in the sample ranged from 6 to 43 months old. Using Pearson's *r* to evaluate the means, standard deviations, and correlations with chronological age for each domain, the mean developmental age scores for each domain are substantially lower than the correlations for children with typical development. 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 development skills of children with disabilities.