DOCUMENT RESUME

ED 353 068 PS 021 038

AUTHOR Hutchens, Teresa A.; And Others
TITLE The Stability of I. Q. in Preschool Years: A Review.
PUB DATE Apr 91
NOTE 23p.; Paper presented at the Meeting of the American Association for Counseling and Development (Reno, NV, April 21-24, 1991).
PUB TYPE Reports - Evaluative/Feasibi' ty (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Intellectual Development; *Intelligence Quotient; *Intelligence Tests; *Measurement Techniques; Measures (Individuals); *Preschool Children· Preschool Education; Testing Problems; Test Reliability; Test Validity
IDENTIFIERS *Age Appropriateness; *Temporal Stability; Test Retest Reliability

ABSTRACT

The question of reliability in the intellectual assessment of young children is cause for concern among developmental psychologists and diagnosticians. The issue of reliability is confounded by normal variability in skills during early childhood, by the problem of consistency across time of age-appropriate assessment measures, and by the selection of subsequent or concurrent measures. Stability across time of measures of children's intellectual ability during infancy and the preschool years is of particular interest to those practitioners faced with diagnosis, placement, and treatment decisions. Research to address temporal stability in mental measurement in the preschool period has yielded inconsistent findings. Differential results appear to be due, at least in part, to children's age at the time of evaluation, the choice of intellectual assessment instrument, the length of the test-retest interval, and unique characteristics of the sample studied. Because the implications of the reliability of intellectual measures for controversial placement issues and long-range educational programming are significant, practitioners should identify trends in assessment practices that relate to the stability of assessment across time and across measures. (MM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATI- J
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED353068

The Stability of I. Q. in Preschool Years: A Review

Teresa A. Hutchens, S. Elaine Hamilton, Patricia A. Town,
Lena R. Gaddis, Rodney J. Presley

The University of Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Teresa A. Hutchens
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PS 021038

Correspondence and requests for reprints should be addressed to Teresa A. Hutchens, Department of Educational Psychology, 325 Aderhold, University of Georgia, Athens, Georgia, 30602.

BEST COPY AVAILABLE

ERIC

The Stability of I. Q. in Preschool Years:

A Review

Abstract

The question of reliability in the intellectual assessment of young children is cause for concern among developmental psychologists and diagnosticians. The issue is confounded, not only by normal variability in skills during early childhood, but also by the temporal consistency of age-appropriate assessment instruments and the selection of subsequent or concurrent measures. Temporal stability during infancy and preschool years is of particular interest to those practitioners faced with diagnosis, placement, and treatment decisions.

Research to address stability in mental measurement within the preschool period has yielded inconsistent findings, particularly with regard to age and specific target population. Differential results appear to be due, at least in part, to age at the time of evaluatio , choice of intellectual assessment instrument, length of the test-retest interval, and unique characteristics of the sample studied. It, therefore, seems important to identify trends in the concurrent and temporal stability of intellectual assessment in normal development, as well as the exceptional. This review presents a consolidation of the information available in an attempt toward clarification of these issues.

ERIC

The Stability of I. Q. in Preschool Years:

A Review

Reliability in the assessment of intelligence has long been an area of concern to the psychologist and diagnostician. Categorical placement decisions are necessary for the receipt of some educational services; treatment decisions and considerations for intervention are often dependent on assessment results demonstrating a reasonably stable picture of an individual's level of intellectual functioning.

As early as 1899, longitudinal study has been recognized as important to address questions of intelligence and developmental changes (Mills, 1899). Longitudinal investigation was difficult for many reasons, one of which was a lack of institutional support over the length of time required for such study. In the 1920's, several child research institutes were established across the United States, equipped to engage in such investigation (Cairns, 1983). Research centers, such as those located at Berkeley, Fels Institute, Minnesota, and Harvard, have hosted many of the longitudinal studies that yield information about the nature and consistency of intellectual assessment during the early developmental period.

Intelligence quotients obtained during middle childhood have been reported to be consistent and highly positively correlated in major longitudinal studies (e.g., McCall, Appelbaum, & Hogarty, 1972). IQ scores obtained between 2 and 6 years of age, during the preschool period, have shown at least moderate validity in predicting later intelligence test performance (Anastasi, 1978). In a longitudinal study of 140 children at Fels Research Institute, Sontag, Baker, and Nelson (1958) found that Stanford-Binet scores obtained at 3 and at 4 years of age yielded a high, positive correlation (r = .83). Scores obtained in subsequent years were also postively correlated with IQ's obtained at 3; magnitude of the correlation decreased as the test-retest interval increased. Moderately high correlations remaine . with retest at age 12 (r = .46). The strength of correlations between preschool and later years increased markedly with increasing age at initial testing (e.g., between 3 and 6 years). IQ's obtained in childhood, after age 6, were found to correlate with those obtained at age 18 at a level of .80 and above (Bayley, 1949). Bradway, Thompson, and Cravens, (1958) used a subgroup of the 1937 Stanford-Binet standardization sample who were originally tested at 2 to 5 1/2 years of age. They were retested at 10- and 25-year retest intervals; correlations of

Temporal consistency has been found to be poor for infant testing, especially in the first year of life; however, infant assessment is shown to have some predictive validity for performance on preschool instruments (Anastasi, 1978). Wilson (1978) assessed a group of infants at 3, 6, 9, 12, 18, and 24 months using the Bayley Scales of Infant Development; they were tested at 3 years on the Stanford-Binet. Correlations increased in magnitude as did the age of the infants, with the strongest relationship between scores obtained at 24 and 36 months of age (r = .73). A notable increase in predictive power was shown at 18 months (Wilson, 1978).

The Collaborative Perinatal Project (Broman and Nichols, 1975) included extensive investigation of the relationships between infant, preschool, and school-age mental development and social class indices. A racially-mixed group (14,665 white and 16,293 black) was tested across a 7-year period, receiving the Bayley Mental Scale at 8 months, the Stanford-Binet Intelligence Test at 4 years, and the Wechsler Intelligence Scale for Children (WISC) at 7 years. While the Bayley Mental Development Scales were good predictors of severe mental retardation at 7 years, they were not strongly related to IQ's obtained by normals at 4 and 7 years.

One of the most carefully executed longitudinal studies in the literature is the Berkley Growth Study (Bayley, 1949). Five different intelligence tests were used across an 18-year age span: California First Year Tests, until 15 months; California Preschool Tests, until 5 years; The Stanford-Binet (forms vary), ages 6 through 12; the Terman-McNemar, ages 13 and 15; and the Wechsler-Bellevue, ages 16 and 18. Scores were more consistent with advancing age at testing. Scores obtained before age two were not closely related to those of school-age evaluation. After two years, correlations with later testing were more positive, rarely less than .50. Authors note that scores at 1 year had a zero correlation with those at age 17; however, IQ's obtained at age 4 were positively correlated with those at age 17 (r = .71). Bayley (1949) concluded that the magnitude of correlations was a combined function of the age of the children and the length of time between testing.

In addition to pairwise correlational data, aggregate data were also addressed by Bayley (1949). Group scores of California First Year Tests at ages 10, 11, and 12 months were positively correlated with intelligence measures given at 17 and 18 years (r = .41). In general, combining of scores from several administrations had a marked effect on the increasing magnitude of correlations over time. Bloom (1964) suggested that aggregation across administrations corrects somewhat for unreliability of the tests, producing higher correlations. Task demands and the qualities measured by existing intelligence tests change from infancy to maturity (McCall, Hogarty, & Hurlburt, 1972; Sattler, 1982). Tests used in the first 18 months are highly saturated with demands in motor and physical development, whereas the focus of tests used at 17 or 18 years of age is on cognitive measures and verbal ability. A gradual shift in item focus from perceptual-motor to verbal skills emerges with increasing age even within a single instrument, the Stanford-Binet (Chase and Sattler, 1980).

Another important issue in the stability of IQ is that of intra-individual differences, sometimes called "instability" (Anastasi, 1978). Results from the California Guidance Study (Honzik, Macfarlane, & Allen, 1948) revealed that individual IQ scores fluctuated by as much as 50 points. It was noted that over the period 6 to 18 years, when test-retest correlations have been generally reported to be high, 59% of children tested had scores which differed by 15 or more IQ points; 37%, by 20 or more; and 9%, by 30 or more. It was also of note that the changes in obtained scores were not usually random or erratic in nature but, rather, were exhibited as consistent upward or downward trends over several consecutive years.

A recent study by Hutchens, Town, Hamilton, Gaddis, and Presley (1988) addressed "instability" of individuals' scores across the preschool period. Over a 5-year span, individually administered IQ tests [Stanford-Binet, McCarthy Scales of Children's Abilities (MSCA), Bayley Mental Development Index (MDI), and Weschler Intelligence Scale for Children - Revised (WISC-R)] were given to a group of normals, ranging in age from 1 to 7 years. Of the total sample (N = 224), 119 received a second evaluation using a different instrument; 59, a third; and 16, a fourth. Grouped data yielded significance ($p < .005$), with positive correlations ranging from .62 to .89 across the four evaluations. The authors noted that, despite the level of positive correlations in group data, there were great differences in individual scores. For 113 of those receiving a second

ERIC

evaluation, more than 60% had differences of at least 1 standard deviation in their obtained scores; 11% had differences of 30 points or more (Hutchens et al., 1988).

Reliability Across Instruments

Another factor in the reliability of preschool/infant intellectual assessment is the choice of instrument. A child's chronological age and psychometric properties of specific tests may dictate a number of instruments potentially appropriate for the evaluation; however, the examiner's role in instrument selection is an important one. Although each of the tests under consideration may yield an IQ-related standard score, they are not identical measures.

Concurrent studies with a number of tests reflect moderate to high correlations; however, it should be noted that much information in test development reflects limited comparison testing. The majority of test manuals report correlations only with the Stanford-Binet, e.g., Bayley's (1969) Mental Development Index (MDI). Of the 350 California children tested, results were reported for 120 of this sample. The correlations ranged from .47 to .64 across groups (ages 24 - 30 months).

Similar investigations have been conducted with the McCarthy, using the Stanford-Binet, WISC, WISC-R, and WPPSI (Wechsler Preschool and Primary Scales of Intelligence) for comparison. A median correlation of .75 was obtained (Sattler, 1982). McCarthy (1972) reports the MSCA manual's only evidence of concurrent validity using the Stanford-Binet and the WPPSI on a restricted sample of 35 children, 6-0 to 6-7 years. The MSCA General Cognitive Index (GCI) was positively correlated (r = .7) with the WPPSI Full Scale IQ and with the Stanford-Binet, 1960 norms (r = .81). The mean GCI in this study was 10 points lower than the mean Stanford-Binet IQ (McCarthy, 1972; Silverstein, 1978). Sattler (1982) suggests that the MSCA standard scores may be about 6 points lower, while Stanford-Binet and WISC-R scores are more similar.

The relationship between the McCarthy's GCI and the WISC-R IQ's was studied using a sample of 51 children, from 7-0 to 8-7 years (Davis & Walker, 1977). Test-retest intervals were 1 to 18 days between counterbalanced administrations. The obtained correlations were .65, .62, and .75 for the Verbal, Performance, and Full Scale IQs, respectively.

Wechsler (1978) reported a study of the WISC-R and Stanford-Binet using a sample of 118 normals at 6, 9 1/2, 12 1/2, and 16 1/2 years of age. The administration of the WISC-R preceeded the Stanford-Binet and intervals varied between testing, from 1 day to 5 1/2 months (median = 1 month) at age 6 and from 2 weeks to 9 1/2 months (median = 3 1/2 months) for the older groups. Average correlations were .71, .60, and .73 for the Verbal, Performance, and Full Scale IQs, respectively. Wechsler (1978) interprets these trends to suggest that the WISC-R and the re-normed Stanford-Binet yield similar scores for normals 6 to 16 years.

A normal sample (mean = 113.7) was also the target of longitudinal study by Hutchens et al. (1988). Results were analyzed in pairwise comparisons by both age and the instrument used for testing across test-retest intervals of one year. The Stanford-Binet and the McCarthy were most often significantly correlated by age. The Stanford-Binet when administered at 4 years of age was also significantly correlated with the WISC-R at age 7 (Hutchens et al, 1988).

An extensive literature is devoted to comparative and concurrent investigations with the Stanford-Binet (e.g., Brooks, 1977). Many of these studies include samples of exceptional populations, previously defined by performance outside the average range. Sewell and Manni (1977) were the first to examine the relationship between the WISC-R and the Stanford-Binet in a normal sample since publication of the 1974 WISC-R manual. Counterbalanced administration with a racially mixed sample of 106 middle class children, (6 to 16-6 years) was conducted over intervals of 3 to 6 weeks. Both tests yielded a higher mean IQ at younger ages (115.70 for the 6-0 to 8-0 group vs. 105.62 for the group 8-2 to 16-10). Average correlation coefficients were .86, .71, and .86 with the Verbal, Performance, and Full Scale IQs, respectively.

Use of different instruments within a variety of exceptionalities may yield differences in consistency. Some studies report stability in the significant, positive relationships between various instruments (e.g., Brooks, 1977; Kaufman & van Hagen, 1977). Others present inconsistencies, even with the use of these tests in categorical educational placement (e.g., Kaufman & Kaufman, 1977). One such investigation was conducted with children referred for learning problems (Bloom, Raskin, & Reese, 1976). Results from the WISC-R and the Stanford-Binet were positively correlated; however, discrepancies between test results and corresponding intellectual classification systems (test publishers and the AAMD) yielded discrepancies in a full 54% of the sample. Investigations support concern regarding the consistency of these instruments with gifted children, mentally retarded, and the learning disabled (summarized in Sattler, 1982). Bloom et al. (1976) emphasized a need for evaluators to be aware of differences in classification systems, individual performance, and differential requirements of the instruments themselves.

Test-Retest Reliability

Test-retest reliability compares performance across administrations of the same instrument. Two major trends have been noted by Sattler (1978) and Anastasi (1978). First, reliability tends to be greater with a short time interval between the first and second administration. Secondly, the magnitude of the correlation appears to increase with increasing age at the time of initial testing. The latter may suggest that skills measured by IQ tests become more stable as with maturity.

Test-retest correlations, commonly reported in test manuals, are often obtained over short time intervals. For example, using a one week test-retest interval, Bayley (1969) reported a 76.4% agreement between administrations, with initial testing by the Bayley Scales of Infant Maturity (MDI) at eight months of age (n=28). Wilson (1978) used the Bayley MDI with test-retest intervals of 3 to 18 months, reporting more variable results, with correlations ranging from .22 to .61.

McCarthy (1972) reported correlations for the IQ-related General Cognitive Index (GCI) over a test-retest interval of three to five weeks. Using three age groups (3 - 3 1/2, 5 - 5 1/2, 7 1/2 - 8 1/2), correlations clustered at .90. McCarthy (1972) and Hunt (1978) report a range of .75 to .90 on all McCarthy Scales using the standardization sample of 125 children with a one month test-retest interval; of all scales, the highest correlation was with the GCI (r = .90). These results were consistent with those reported in a separate study with an interval of three to six weeks. A test-retest correlation of .88 was obtained for the GCI with a sample of 38 middle class suburban children, initially tested between the ages of 5 and 6 (Roffe and Bryant, 1979).

Use of more lengthy test-retest intervals have also produced consistent findings. Davis and Slettedahl (1976) used an interval of one year and found a correlation of .85 for the McCarthy GCI with a culturally mixed sample (n = 43) of rural

ERIC

kindergarten students. The McCarthy GCI was also used by Ernhardt and Callahan (1980); they reported a correlation coefficient of .61 over a five year test-retest interval. Their sample consisted of 68 urban black children tested initially within the preschool period.

The standardization sample of the Wechsler Intelligence Scale for Children - Revised (WISC-R) was evaluated with a test-retest interval of three to five weeks. For the Full Scale IQ, a test-retest correlation of .96 was reported (Wechsler, 1978). For the youngest children in the sample, 6 1/2 to 7 1/2 years of age, a correlation coefficient of .95 was found.

The Stanford-Binet was used by Payne, Hallahan, Ball, and Obenauf (1972) in the evaluation of 158 Head Start students. With a test-retest interval of one year, differential correlations were found for male and females. Correlations for two groups of boys were .77 and .65; for the two groups of girls, correlations were .24 and .50. It was suggested by the authors that environmental influences contributed to differential development of cognitive abilities.

A study by Schwartz and Blonen (1975) indicated that changes in an individual's IQ scores over successive evaluations were common. Fifty-eight subjects were tested at varying intervals between one and six years of age; they were tested again at age 16. The Stanford-Binet was administered after age two, with an alternative measure used before that time. The authors reported

ERIC

significant differences in individuals' score across administrations; 50% of the subjects had differences of 13 points or more in obtained scores on at least three different evaluations (Schwartz & Blonen, 1975). Group IQs were suggested to remain stable while individual scores may fluctuate.

Across four years of longitudinal study by Hutchens et al. (1988), annual administrations of either the Bayley MDI, McCarthy, Stanford-Binet, and the WISC-R, grouped IQ data revealed positive correlations from .62 to .82; all were significant ($p < .005$). Without regard to the instrument used, correlations across consecutive years between ages 2 and 7, inclusive, were positive and significant, with a single exception. IQ scores obtained at 4 years were significantly correlated with those of age 5, but not at age 6. No significant correlations were found when the test-retest interval exceeded 2 years. These results add further evidence that the greater the time interval between testing during the preschool years, the greater the likelihood of obtaining variable scores.

This was also reflected in the analysis of scores obtained by 57 individual subjects across 3 or more years. Without regard to the choice of instrument, 50% of this sample had a difference of 16 points or more (~ 1 sd); 23% had a range of 25 points or more, and 9% had a difference of 33 or more points (Hutchens et al, 1988). Findings lend support to the interpretation of Schwartz and Blonen (1985), suggesting that the stability of

ERIC

intelligence scores suggested by group data may mask the wide variability obtained by individuals in the preschool years.

Discussion

The results of group testing using intellectual assessment during the preschool period appear to be fairly stable with regard to both temporal and inter-test reliability. However, the practitioner should be aware of exception to this indication of stability, particularly when considering grouped data and individual variability.

Intelligence scores obtained during the first year of life do not adequately predict IQ's during later testing, especially when the test-retest interval is broad. Predictions appear to be more reliable when the target population is mentally deficient. Research suggests that this may be due, at least in part, to the differences in task demands and quality of skills assessment at the youngest ages.

The length of time between evaluations may also effect stability, particularly in the early years. Reliability is greatest when the subjects were older, at the upper limits of the preschool period, and the test-retest interval is short. Instrumentation plays a interactive role with consideration of these variables, such that, greater consistency is likely when using the same measure in subsequent testing. Tests which tap similar developmental skills seem to yield more consistent results over time. Practitioners should, therefore, consider the age of the child at initial testing, the retest period, and the nature of the assessment instruments used.

Perhaps the most significant finding in this body of research is the high incidence of significant variability (one standard deviation or more) in individual scores. Obtained scores may vary in a somewhat predictable fashion when the initial testing contributes to differential analysis of strengths and weaknesses and are compared to task demands of subsequent measures. However, the most important fact of individual variability, particularly in the developmental period, is its poor predictability over time. Research suggests that within a span of a single year, obtained scores may vary by as much as 1 standard deviation in 50% of the normal preschool population and as much as 2 standard deviations in 10%.

The implications for controversial placement issues and long-range educational programming are significant. It appears that the most carefully formulated placement decisions for preschool/early intervention may be generated from an assessment strategy that, not only includes a second estimate of intellectual abilities, but also one that takes into account the valuable component of temporal variability. This would suggest the inclusion of a "reasonable" test-retest interval between administrations of individual IQ measures. This strategy would require that evaluations be extended; yet, it would more carefully insure the reliable and appropriate use of data when

ERIC

addressing service issues, particularly where results impact on placement decisions and long-range planning. Those who are faced with placement decisions should weigh the possibility of extended the evaluation period when there are questionable levels of global functioning.

It must be remembered that factors other than age, testing intervals, or instrumentation also influence the stability of the obtained IQ. Environmental characteristics, educational experiences, ability to benefit from experiences, etc. are obvious (Clarizio, 1979; Madden, 1980). Operationalizing for curriculum-based assessment is relevant and vital for a comprehensive approach to individual evaluation; however, preschool evaluation and intervention planning continues to rely on traditional assessment techniques. It behooves the astute practitioner to utilize such information in the most dependable, reliable way possible.

References

Anastasi, A. (1978). Psychological testing (5th edition). New York: MacMillan Publishing Co., Inc.

Bayley, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. Journal of Genetic Psychology, 75, 165-196.

Bayley, N. (1969) Manual for the Bayley Scales of Infant Development. New York: The Psychological Corporation.

Bloom, B. S. (1964). Stability and change in human characteristics. New York: John Wiley & Sons, Inc.

Bloom, A., Radkin, L., & Reese, A. (1976). A comparison of WISC-R and Stanford-Binet Intelligence Scale classifications of developmentally Diosabled Children. Psychology in the Schools, 13, 288-290.

Bradway, K. P., Thompson, C. W., & Cravens, R. B. (1958). Preschool IQ's after twenty-five years. Journal of Educational Psychology, 49, 278-281.

Broman, S. H., & Nichols, P. L. (1975). Early mental development, social class, and school-age IQ. Paper presented at the meeting of the American Psychological Association, Chicago.

Brooks, C. (1977). WISC, WISC-R, S-B, L&M, WRAT relationships and trends among children ages 6 to 10 referred for psychological evaluation. Psychology in the Schools, 14, 30-33.

Cairns, R. B. (1983). The emergence of developmental psychology. In P. H. Mussen (Ed.), Handbook of Child Psychology: Volume I (4th ed.). New York: John Wiley & Sons.

Chase, C. H., & Sattler, J. M. (1980). Determining areas of strengths and weaknesses on the Stanford-Binet. The School Psychology Review, 9, 174-177.

Clarizio, H. F. (1979). In defense of the IQ test. The School Psychology Digest, 8, 79-88.

Davis, E. E., & Slettedahl, R. W. (1976). Stability of the McCarthy Scales over a one-year period. Journal of Clinical Psychology, 32, 798-800.

Davis, E. E., & Walker, C. (1977). McCarthy Scales and WISC-R. Perceptual and Motor Skills, 44, 966.

Ernhardt, C. B., & Callahan, R. (1980). The McCarthy Scales: Predictive validity and stability of scores for urban black children. Educational & Psychological Measurement, 40, 1183-1188.

Honzik, M. P., Macfarlane, J. W., & Allen, L. (1948). The stability of mental test performance between two and eighteen years. Journal of Experimental Ed cation, 17, 309-324.

Hunt, J. V. (1978). The McCarthy Scales of Childrens' Abilities. In O. K. Buros (Ed.) Eighth Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press.

Hutchens, T. A., Town, P. A., Hamiliton, S. E., Gaddis, L. R., & Presley, R. J. (1988). A longitudinal study of reliability in intellectual assessment: The preschool period. Manuscript submitted for publication.

Kaufman, A. S., & Kaufman, N. L. (1977). Clinical evaluation of young children with the McCarthy Scales. New York: Grune & Stratton.

Kaufman, A. S., & van Hagen, J. (1977). Investigations of the WISC-R for use with retarded children: Correlation with the 1972 Stanford-Binet and comparison of WISC and WISC-R profiles. Psychology in the Schools, 14, 10-14.

McCall, R. B., Appelbaum, M. I., & Hogarty, P. S. (1973). Developmental changes in mental performance. Monographs of the Society for Research in Child Development, 38 (3, Serial #150).

McCall, R. B., Hogarty, P. S., & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. American Psychologist, 27, 728-748.

McCarthy, D. (1972). Manual for the McCarthy Scales of Children's Abilities. Cleveland, Ohio: The Psychological Corporation.

Madden, P. B. (1980) Intelligence tests on trial. The School Psychology Review, 9, 149-153.

Mills, W. (1899). The nature of animal intelligence and the methods of investigating it. Psychological Review, 6, 262-274.

Payne, J. S., Hallahan, D. P., Ball, D. W., & Obenauf, P. A. (1972). Sex differences in reliability and congruent validity of Stanford-Binet Short Form and Peabody Picture Vocabulary Tests. Psychological Reports, 31, 934.

Roffe, M. W., & Bryant, C. K. (1979). How reliable are MSCA profile interpretations? Psychology in the Schools, 16, 1-18.

Sattler, J. M. (1982). Assessment of children's intelligence and special abilities (2nd edition). Boston: Allyn and Bacon.

Sattler, J. M. (1978). The McCarthy Scales of Children's Abilities. In O. K. Buros (Ed.) Eighth Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press.

Schwartz, E. M., & Blonen, A. S. (1975). A 16-year longitudinal study of variation in intellectual performance. Journal of Clinical Psychology, 31, 687-694.

Sewell, T., & Manni, J. (1977). Comparison of scores of normal children on the WISC-R and Stanford-Binet, Form L-M. Perceptual and Motor Skills, 45, 1057-1058.

Silverstein, A. B. (1978). The McCarthy Scales of Childrens' Abilities. In O. K. Buros (Ed.) Eighth Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press.

Sontag, L. W., Baker, C. T., & Nelson, V. L. (1958). Mantal growth and personality development: A longitudinal study. Monographs of the Society for Research in Child Development, 23 (2, Serial #68).

Wechsler, D. (1978). Manual for the Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation.

Wilson, R., S. (1978). Sensorimotor and cognitive development. In F. D. Minife & L. L. Lloyd (Eds.), Communicative and Cognitive Abilities - Early Behavioral Assessment. Baltimore, Md: University Park Press.