1982

An Investigation of a Slosson Intelligence Test Classification Schema as an Aid in Diagnostic-Educational Hypothesis Formulation

Andrea Rolsky
Loyola University Chicago

AN INVESTIGATION OF A SLOSSON INTELLIGENCE TEST CLASSIFICATION SCHEMA AS AN AID IN DIAGNOSTIC-EDUCATIONAL HYPOTHESIS FORMULATION

by

Andrea Rolsky

A Dissertation Submitted to the Faculty of the Graduate School of Education of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of Doctor of Education

January

1982

Andrea Rolsky

Loyola University of Chicago

AN INVESTIGATION OF A SLOSSON INTELLIGENCE TEST CLASSIFICATION SCHEMA AS AN AID IN DIAGNOSTIC-EDUCATIONAL HYPOTHESIS FORMULATION

In the present three-step investigation, a SIT classification system, patterned after Sattler's (1965) Stanford-Binet schema, was developed by having at least two out of four experienced judges, using content analysis and a sorting technique, agree on the assignment of each SIT item (from year two to year twenty-seven) to either the Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor or Social Intelligence-Reasoning Categories. A high percentage of agreement by three out of three judges on 73.7% of the items suggests that the resultant SITFILE appears to have some face validity. Analysis of the distribution of the items within both the Sattler S-B Binetgram and SITFILE categories at four different age levels suggests that both the S-B and SIT share similar function assessments but different developmental designs.

One hundred-fifty Chicago parochial school students, grades two through eight, participated in an exploration of the SITFILE's reliability. Ninety-five students attending a university diagnostic service center participated in the study

of the SITFILE's validity. Individual category scores were calculated by using chronological age as the reference point for standard deviation scatter analysis.

Only the Language, Memory and Numerical Reasoning Categories were found generally to possess sufficient reliability for middle class white students in grades two through eight. Adequate specificity, while somewhat lower for the Memory Category than for the Language and Numerical Reasoning Categories, was reported. Corrected Pearson stability coefficients between .73 and .98 for the Language, Memory and Numerical Reasoning Categories were also reported, as were small standard error of measurements.

A measure of each Language, Memory and Numerical Reasoning Categories' validity was obtained by correlating SITFILE category scores with age scores achieved on either the ITPA or the Detroit and the WRAT. Significant correlations (p .05) suggest that the SITFILE Language, Memory and Numerical Reasoning Categories measure functions related to those measured by these frequently employed diagnostic instruments. However, interactions suggested by large amounts of common variance and multiple correlations between the Language, Memory and Numerical Reasoning Categories and identified diagnostic tests argue against any independent interpretation of isolated SITFILE category scores.

ACKNOWLEDGEMENTS

Many individuals have been instrumental in providing the necessary information and support for the completion of this study. Certainly, this investigation would not have been possible without the help of my colleagues at the William Itkin Children's Service Center and at Saint Mary of the Woods School -- particularly that of Lillian Stevenson, Marge Thomas, Margaret Cushing, Marie Egan and Sister Lynette Lamphere. Also, the cooperation and tenacity of many Northeastern Illinois University Graduate Students, who are too numerous to mention, must be recognized. Without their conscientious efforts it would have been impossible to test individually the two hundred forty-five (245) children included in this study. The children and their parents must also be thanked for their willingness to participate in programs with primary research objectives. Finally, I must thank my dissertation committee members: Doctor Jack Kavanagh and Doctor Martha Wynne and especially my Director, Doctor Ronald Morgan. Only through their guidance, insight and encouragement was it possible to design and successfully conclude this investigation.

VITA

The author, Andrea Louise Rolsky, is the daughter of Morton and Ruth Rolsky. She was born in Kansas City, Missouri on July 23, 1949.

Her elementary and secondary education was obtained in the Kansas City Public School Systems. She graduated from Southwest High School in 1967. She attended Drake University from 1967 to 1969 and the University of Missouri at Kansas City from 1969 to 1970. She completed her undergraduate work at California State University at Northridge in May of 1970, receiving her Bachelor's of Art Degree. She began graduate work at California State University in September of 1970 and completed her Master's of Art Degree with specialization in teaching children with learning and/or behavior disabilities at Northeastern Illinois University in May of 1976.

In September of 1976 she became Program Director and Coordinator of Saint Mary of the Woods School's Learning Enrichment Program. At Saint Mary's she conducted educational evaluations as well as supervised the special instruction of children referred to the Learning Enrichment Program. In 1977 she joined the Loyola University of Chicago part-time faculty, teaching undergraduate foundations of education courses. In January of 1978 she became a visiting instructor at Northeastern Illinois University. At Northeastern Illinois University she supervised and taught both undergraduate and graduate students enrolled in the Learning Disabilities Special Education Sequences. Also, during the 1978-79 academic year she completed a school psychology internship in Skokie, Illinois School District #67. The following academic year she continued working at the District #67 junior high school conducting psychological assessments of children referred for individual evaluations due to educational problems. Presently, Ms. Rolsky continues to function as the Program Director at Saint Mary of the Woods School as well as a part-time lecturer at Loyola University of Chicago and a part-time instructor at Northeastern Illinois University.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table | Page

CONTENTS OF APPENDIX

INTRODUCTION

Since the passage of Public Law 94-142 and its enforced compliance beginning in 1978, the public schools have found themselves mandated to explore the learning problems of a larger segment of our school age children. While state and district interpretations of this law have resulted in varying programs, the law is clear in its specification of the need for both an initial psychological and educational evaluation, as well as periodic re-evaluations. Consequently, increased interest has been placed on the development and employment of screening and/or multi-purpose test batteries.

One instrument that has been utilzed to a large extent in educational evaluations is the Slosson Intelligence Test (SIT). Steward and Jones (1976) report that usage of the SIT has greatly increased during the past decade. Slosson originally published the SIT in 1961 with the primary intent that it be used as a screening instrument to evaluate the general intelligence of individuals between four years of age and adulthood. Since the test is composed of different kinds of items (language, memory, numerical reasoning, etc.) a number of educators and psychologists charged with evaluation and development of educational prescriptions, have suggested systems for extending the SIT's utility by incorporating scatter analytic procedures (Canfield, 1972; Boyd, 1974; Stone, 1975; Hedberg & Shapiro, 1976).

Two interpretative systems have been published (Stone, 1975; Boyd, 1974). However, neither of these schemes nor any of the available unpublished schemes (Canfield, 1972; Project Success, 1975; Hedberg & Shapiro, 1976) have reported any significant normative data to support the reliability of their porposed "subscales" or the validity of employing SIT scatter analysis. The employment of such an unproven approach appears highly questionable as educators and psychologists must carefully scrutinize their interpretive techniques as well as their instruments.

Overall, the present study investigates the use of the SIT as an aid for generating diagnostic hypotheses concerning children's learning aptitudes. A SIT classification schema was developed and correlated with the Sattler Stanford Binet (Form L-M) Binetgram to assist in the clarification of the construct validity of the SIT classification system. The reliability of SIT scatter analysis was investigated by evaluating three hundred (300) test-retest protocals of children in grades one through eight. Furthermore, in an attempt to explore the concurrent validity of a SIT classification schema and scatter analysis, the SIT responses of ninety-five (95) students, between five and fourteen years of age, were correlated with results from selected educational assessment instruments (The Illinois Test of Psycholinguistic Abilities, 1968; The Detroit Tests of Learning Ability, Revised, 1967; The Wide Range Achievement Test, 1965).

It is intended that the results of the present study will enable one to judge more accurately the validity of using SIT scatter analysis when making educational decisions. As long as the absence of such data persists, psychologists and educators continue to question seriously the use of the SIT and scatter analysis as diagnostic aids.

REVIEW OF THE LITERATURE

When systematically reviewing the literature pertinent to the development of a SIT qualitative diagnostic system, it becomes necessary to consider four areas of previous investigation: research dealing specifically with the SIT's characteristics; relevant data regarding the design of classification systems; research pertinent to the development of a SIT scatter analysis format; and finally, research regarding previous SIT interpretive systems.

Description of the SIT

When the SIT was published in 1961 it was presented as an abbreviated intelligence test which could be administered to children or adults. It is an age scale of graded test questions from year two to year twenty-seven, modeled after those of the Gesell Developmental Schedules and the 1966 Stanford Binet (S-B) Form L M. In designing his test, Slosson intended that its ease of administration would make it possible not only for psychologists to administer the test, but also for teachers or counselors to do so.

Slosson's standardization population for the SIT was geographically restricted to New York state. However, inclusion of all English speaking intellectural, racial and socio-economic groups make his sample broadly representative.

SIT graded test questions are presented to subjects auditorily and depend heavily upon language skills both for comprehension of the stimuli and item response. An average of twenty minutes is required for a SIT administration; however, with either a very slow individual or one who evidences a great deal of variability, it may take up to thirty minutes to reach a ceiling on the test. The basal age for a child is determined at that point where the individual achieves a series of ten successful passes. The ceiling for an individual is that point where ten items in a row are missed. Administration of the SIT results in a ratio IQ with a mean of 100. While Slosson (1963) presents an IQ classification chart for interpreting IQ scores, the relationship between it and his reported SIT standard deviations of 24.7 and 25.1 is not clear. The SIT standard deviation as calculated according to the data included in Stewart and Jones' (1976) rather comprehensive review is seventeen points.

SIT Reliability and Validity

Test-retest investigations have shown the SIT to be a reliable measure of student potential (Hammill, 1969 r=.97; Hammill, Crandell, and Colarusso, 1970 r=.96). SIT internal consistency coefficients derived by the splithalf procedure have been reported as ranging between .81 to .97 (Hammill, 1969; Hammill, Crandell, and Colarusso, 1970). Many studies have investigated the validity of the SIT as an index of general intelligence. (Slosson, 1963; DeLapa, 1967; Houston and Otto, 1968; Jongeward, 1969; Kaufman and Ivanhoff, 1969; Carlisle, 1970; Meissler, 1970; Swanson and Jacobson, 1970; Armstrong and Mooney, 1971; Johnson and Johnson, 1971; Stewart, Wood and Gallman, 1971; Lessler and Galinsky, 1971; Maxwell, 1971; Jerrolds, Callaway and Gwaltney, 1972; Armstrong and Jensen, 1972; Machen, 1972; Martin and Rudolp; 1972; Lamp and Traxler, 1973; Ritter, Duffey and Fischman, 1973; Steward and Myers, 1974). A review of the results from these investigations reveals that when the range of subjects is not restricted: SIT rankings and scores are comparable to S-B rankings and scores; SIT rankings and scores are comparable to Wechsler Full Scale rankings and scores; SIT rankings and scores are comparable to Wechsler Verbal Scales rankings and scores; but, SIT-Wechsler Performance Scales correlations are lower and more variable. The lower SIT-Wechsler Performance Scales' correlations are important as they suggest that the intellectual skills assessed by the SIT are only moderately related to those assessed by the Wechsler Performance Scales. The implications of these findings underline cautions against employing the SIT with the same expectations as one might have for the Wechsler

Intelligence Scales. High SIT-Stanford Binet correlations (.90's range) and SIT-Wechsler Verbal and Full Scales correlations (Low .80's range) do suggest, however, that limited interpretations concerning a child's intelligence can be made with confidence.

Studies have also been designed to investigate the relationship between SIT scores and measures of school achievement (DeLapa, 1976; Hammill, 1969; Shepherd, 1969; Stewart, Wood and Gallman, 1971; Hutton, 1972; Martin and Rudolph, 1972; Lamp, Traxler and Gustafson, 1973). Correlations between the SIT and the various achievement measures included in these studies ranged between .24 to .75.

In sum, empirical evidence has consistently shown that the SIT is a reasonably reliable and valid standardized instrument measuring many of the same attributes that the S-B and Wechsler Scales measure. Its ease of administration and scoring has resulted in its frequent use and acceptance by professionals as a useful tool for screening purposes.

Classification Systems

A test classification system is a systematic division of test items into groups. This division is done according to a definite plan and makes possible a particularized examination of an individual's performance. The diagnostic utility of a classification system is based upon the observation that while many individuals may achieve a similar number of correct responses leading to a similar total score, these correct responses are themselves not necessarily made to the same items.

Utilization of a classification system can make it possible to describe an individual's intra-test variability by looking for patterns of successes and failures. With a classification system, one can derive manageable qualitative information. Employment of classification systems has reportedly provided useful clues for more specific follow-up testing and has furthered diagnostic decision making (Sattler, 1975).

For the most part, classification systems have been systematically developed through content analysis as well as factor analysis. When content analysis is employed, variables that adhere to restrictive criteria are grouped into categories with face validity. This grouping of items can be accomplished by either a single judge or by a panel of expert judges working independently. Content analysis reliability may be improved through the use of a panel of judges and consensus criteria.

Factor analysis is a statistical procedure that summarizes the interrelationships among different variables in a parsimonious fashion. This empirical method identifies variables that are qualitatively different from one another as well as the degree of generalizability between each variable. Thus, factor analysis is an objective means of grouping homogeneous test items so as to aid conceptualization and interpretation. However, the educational usefulness and generality of resultant factor analytic conceptualizations has been questioned. While factor analysis does reduce a large number of variables to a smaller number, because it seeks to explain variance, it can also overlook important asymptotic functions that combine differentially within a factor. Sattler (1975) notes that this failure to recognize asymptotic functions may too narrowly constrict qualitative behavioral analysis.

Stanford Binet Classification Schemes

In his description of the SIT, Slosson (1963) states that he modeled a portion of the SIT items after S-B items. When proposing a classification system for the SIT, it is consequently appropriate to review first those schemes devised for extending the interpretation of the S-B.

The generation of S-B schemes began shortly after the introduction of the 1916 S-B with at least fifteen systems proposed for the 1916 edition. An early system by Brighman (1917) incorporated nine categories including: ideation, judgement, school training, association, memory imagination, kinesthetic discrimination, suggestibility and perception. Roe and Shakow (1942) classify S-B items into two broad categories: learned material dependent upon factual recall and thought material dependent upon integration and synthesis.

Publication of the 1937 S-B generated additional classification systems which were largely variations on the earlier schemes (Davis, 1941; McNemar, 1942; Bradway, 1945; Slutzky, Justman and Wrightstone, 1953; Bradway and Thompson, 1962). McNemar (1942) while presenting his system seriously questions the employment of S-B classification systems to measure special abilities. In proposing his vocabulary, nonverbal, and memory scales, he concludes that only the vocabulary scale should be utilized to aid in making specific diagnostic statements. A complex schema proposed by Fromm, Hartman and Marschak (1954) concentrated on providing insights into a child's psychodynamics rather than his learning difficulties.

The publication of the 1960 S-B Form L-M was followed by additional classification schemes. Meeker (1969) utilized Guilford's Structure of Intellect Model to classify not only the test items of the S-B but also the Wechsler Scales. By coding test items according to a three letter system, corresponding to the Guilford dimensions of operations, content and products, two hundred and forty-nine (249) classifications were specified for the one hundred and twenty-two (122) tests of the L-M Form excluding test alternates. A template, available for the whole scale, keying the test items to the Meeker system, was devised to aid an examiner in evaluating an individual's strengths and weaknesses.

Valett (1964) also devised a classification system and published an interpretive chart. Valett's schema categorizes test items as assessing: judgement and reasoning, vocabulary and verbal fluency, general comprehension, memory and concentration, visual motor, and arithmetic reasoning. Sattler (1965) proposed a similar classification system and developed the Binetgram for charting an individual's responses. He identified seven functions: Language, Reasoning, Numerical Reasoning, Memory, Visual Motor and Social Intelligence. Sattler notes that his classifications are somewhat arbitrary. However, Sattler's classification and the Binetgram serve as a model for the development of a classification system to assist in making test interpretations. Kaufman (1978), in his proposal of a simplification of the standard deviation method with the Binetgram, notes the continuing absence of reliability and validity investigations of the Sattler schema. Still, he recommends cautious application of the model when one maintains awareness that it does not assess some essential abilities.

Thus, classification systems have been employed to describe individual intra-test variability. They have been used with one of the SIT's parent instruments, the S-B, since the S-B's earliest introduction and evolution. Investigators such as Valett (1964), Sattler (1965) and Kaufman (1978) have endorsed the employment of classification systems, as a means by which to obtain manageable information regarding individual response patterns, even in the absence of supporting empirical data. It is suggested that through the employment of classification systems one can be provided with useful additional information, which will consequently help direct future testing and hypothesis formulation.

Scatter Analysis

Generally, on a test composed of increasingly difficult items, it is anticipated that normal subjects will systematically fail a greater proportion of test items as the individuals progress through the scale. However, total response consistency is not expected. Normal individuals reportedly display some irregular performances (Schafer, 1944; Rapaport, 1945; Jastak, 1948; Kaufman, 1979). This tendency to evidence irregular performance on a given test is referred to as test scatter. Scatter analysis is an attempt to systematically summarize and/or quantify this phenomenon. Scatter analysis provides a framework for additional qualitative analysis and a method for generating diagnostic and or educational hypotheses.

By inspecting indices of a subject's response variability, it is believed that some consistent traits of the individual may be revealed. Kaufman (1975) points out that the validity of these interpretations is related to the administered test's specificity. Confidence in scatter analysis implies the belief that factors such as testee motivation and location of test items with regard to their level of difficulty have been considered and minimized. Scatter analysis can involve assessment of intra-test variability and/or assessment of patterns of inter-test variability.

Purpose of Scatter Analysis

The interpretation of scatter analysis derived from a desire to employ well accepted, valid assessment instruments in a way that would yield data regarding specific variables. It was felt that variability studies could be as important as a final score and provide more information than an IQ quotient. Practically, scatter analysis interpretation is based upon the belief that behavioral acts are an expression of both intellectual and non-intellectual factors. The practice, when used to identify personality variables reflects a theory of intelligence postulating a dominant general factor and group or specific factors of such small loadings as would not account for variability from task to task. Scatter analysis when used to identify cognitive variables reflects a theory of assessment which postulates that tests or subtests can have sufficient specificity so as to clarify mental organization and consequently further diagnostic and or educational planning.

Over the years, scatter analysis has frequently been employed as a clinical tool when attempting to understand individual differences. Kaufman (1976) states that when an abnormal amount of scatter occurs, analysis often can further the evaluation process. Scatter analysis has been used when attempting to differentiate between normal and emotionally disturbed, cognitively limited, those who are cognitively inefficient or have specific learning problems and those who are mentally superior (Kendig and Richmond, 1940; Babcock, 1941; Rabin, 1941; Bijou, 1942; Rabin, 1942; Schafer and Rapaport, 1944; Strother, 1944; Sloan and Cutts, 1945; Justak, 1948; Olch, 1948; Garfield, 1945; Heyer, 1949; Levine, 1949; Clark and Moore, 1950; Furvitz, 1950; Harper, 1950; Warner, 1950; Seashore, 1951; Schneider and Smilles, 1959; Vane, Weitzman and Applebaum, 1966; Kaufman, 1976). Scatter analysis has also been employed when attempting to clarify the dimensions of a particular disorder (Piotrowski, 1937; Kendig and Richmond, 1940; Bijou, 1942; Magaret, 1942; Roe and Shakow, 1942; Gilleland, 1943; Magaret and Wright, 1943; Schafer, 1944; Silverstein, 1968; Rugel, 1974; Ackerman, Dykman and Peters, 1976; Anderson, Kaufman and Kaufman, 1976; Kaufman and Van Hagin, 1976; Vance, Gaynor and Coleman, 1976; Smith, Coleman, Dokecki and Davis, 1977; Zingale and Smith, 1978). The usefulness of test scatter is difficult to assess in view of the differing results obtained. The significance of scatter as a diagnostic sign or dimension can not be fully evaluated unless one knows that such scatter occurs infrequently in the normal population. The specific results of the directly relevant S-B scatter analysis studies mentioned above are discussed more fully after reviewing pertinent scatter analysis methodology.

Interpreting Scatter Analysis

As stated earlier, scatter analysis can involve either intra- or inter-test variability. When it involves inter-test variability, it is frequently referred to as profile analysis. Intra-test scatter analysis can focus on the range of scatter, the area of scatter for the entire test or clusters of test items, or on a combination of range and area scatter. Range of scatter refers to the age levels covered. Area of scatter refers to the number of items failed below and number of items passed above a designated point.

When calculating scatter within or between tests or between clusters of tests, one must choose a reference point from which to measure scatter. The decision as to choice of reference point should be based upon consideration of its statistical stability and psychological relevance. Reference points from which to measure scatter have included test basals and ceilings, a single test score believed to be a good measure of general intelligence, such as mental age, test and subtest means, and a stable individual score such as chronological age. When evaluating the benefits of one reference point over another, it should be noted that the constancy of a mean score is an asset. However, it should also be realized that if the trait being measured is included and consequently has a variable effect on the reference point (the mean), then the reference point may be contaminated.

When interpreting scatter within or between tests, or between clusters of tests one must determine a means for evaluating the significance of the observed scatter. A "cardinal rule" of profile analysis is that statistically significant differences must exist between scales or subtests (Sattler, 1974). Also, when scatter analysis is being employed for the purposes of classification, statistically significant differences must also exist between the subject's degree of scatter and that seen in the normal population (Kaufman, 1975).

Scatter Analysis and the Stanford Binet

As it was appropriate to review classification systems developed for employment with the S-B, as a parent instrument of the SIT, so it is appropriate to review studies of scatter analysis and the S-B. All three methods of scatter analysis (range, area and combined range and area techniques) have been used to derive qualitative S-B data.

Range scatter was employed by Doll (1919), Mateer (1921) and McFadden (1931). Harris and Shakow (1937) criticize scatter measures soley dependent upon range or span as being too coarse. Wells (1927) defends range scatter techniques on the basis of their simplicity of computation. Area scatter techniques were employed with the S-B by Doll (1919), Wells and Kelly (1920), and Wallin (1922, 1927 and 1929). This method totals earned credits and does not consider the range of levels over which successes are spread. However, it is logical that there may be a correlations between the range and the number of advance credits earned. Combination range and area S-B scatter techniques have been employed by Pressey and Cole (1919), Mathews (1921), Merril (1924), Woodworth (1928), Emch (1931), Shakow and Millard (1935), Weisenberg, Roe and McBride (1936), Riggs and Burchard (1952), Vane, Weitzman and Applebaum (1966), and Gittleman and Birch (1967). These combined scatter analytic techniques have the advantage of considering both the number of levels over which successes and failures are distributed, and the regularity or the degree of success at each level.

Several studies relying on combined scatter techniques have incorporated standard deviation methodologies to measure variability.(Thomson, 1926; Merrill, 1924; Woodworth, 1928; Sattler, 1965; Kaufman, 1975). While these standard deviation techniques seem to employ more objective criteria, the assumption that successes are normally distributed in a cumulative frequency form ignores the observation that the distribution of S-B successes and failures sometimes shows significant deviations from normal kurtosis (Harris and Shakow, 1937).

Results of Stanford Binet Scatter Analysis

The empirical results of the previously cited scatter studies of children have been equivocal. A number of studies suggest no significant differences between scatter of feebleminded, delinquent, neurotic and normal children (Pressey and Cole, 1919; Doll, 1919; Wallin, 1922, 1927; Emch, 1931; Schneider and Smillie, 1959) and only moderate differences between the scatter of bright and average children (Merrill, 1924; Wallin, 1927; Emch, 1931). While Pressey and Cole (1919) found that scatter was not systematically related to mental age, Vane, Weitzman and Applebaum (1966) found greater scatter among children identified as emotionally disturbed than among non-emotionally disturbed children. Berko (1955) found a correlation between the learning efficiency and scatter of brain-injured children. Wallin (1929), McNemar (1942), and Vane, Weitzman and Applebaum (1966) conclude that scatter may be related to the nature of the S-B, and consequently they emphasize uncertainties regarding the usefulness of scatter as a pathognomonic sign.

Studies investigating the scatter of adults on the S-B have also failed to provide conclusive interpretive evidence. Pressey and Cole (1919) and McFadden (1931) reported higher scatter in feebleminded adults than in normal children. Suggestive differences between groups of psychotic adults were identified by Pressey and Cole (1919) and Wells and Kelly (1920). Harris and Shakow (1938) studied the scatter of schizophrenic, normal, and delinquent adults, but found only mental age to be related to degree of scatter.

Schofield (1952) summarized the results of scatter S-B studies previous to 1952 by writing that numerous investigations had failed to confirm scatter analysis as a valid determinant of diagnostic signs. Subsequent studies have failed to provide any further conclusive evidence. The usefulness of scatter analysis may be limited by such testee behaviors as temporary shifts in effort, general distractability or momentary confusions. Specific problems with S-B scatter analysis may be attributable to problems with the test's construction and the lack of perfect correlations among the tests and limitations of specific test's discrimination power. However, the failure of scatter analysis to significantly improve diagnostic decisions may also be attributable to problems with external validity criteria (Jastak, 1949).

Summary of Pertinent Scatter Analysis Research

Pattern or scatter analysis has been attempted since the introduction of discriminate heterogeneous scales. Scatter has been an observered characteristic of normal and atypical examinee test behavior. Overall, two general rationales for the employment of pattern analysis are distinguishable. One rationale is grounded in the belief that psychometric tests measure intelligence and that mental disorders or inefficiencies will be detectable by their effects on cognitive processes as revealed by an analysis of test responses. A second rationale is grounded in the belief that through factor analysis independent functions can be identified and profile analysis employed to explain individual differences. However, the utility of factor analysis for educational planning is clouded by the problem of asymptotic functions.

Growing primarily out of intelligence theory, inter-personal and intra-personal comparisons have been made. Intra-individual norms have been established by studying inter-test discrepancies and intra-test response patterns.

When determining a reference point for scatter analysis, whether of an intra- or inter-personal nature, it is suggested that the decision of its choice be based upon its relevancy and its statistical stability. Results of numerous studies indicate that the method chosen for quantifying scatter should incorporate considerations of both the number of levels over which the successes and failures are distributed and the amount of success at each level. It is also noted that a measure of scatter should not be systematically related to any other irrelevant variable. The results of specific S-B scatter analysis studies, as previously discussed, have been somewhat discouraging. S-B scatter has not conclusively differentiated normal from abnormal children or adults.

Previous SIT Interpretive Systems

After a discussion of the development and purpose of classification schemes and a review of systems that have been employed with the S-B, it is appropriate to review SIT classification systems. Slosson (1963) notes the need to consider individual examinee responses as well as final quantitative results. While many systems have been distributed informally, two systems were published in national journals (Boyd, 1974; Stone, 1975).

Stone (1975) published a system he had developed in 1969. His schema utilized the Valett (1964) S-B classification system as a model. A jury of three psychologists utilizing the Valett format assigned test items to the following categories: Information and Comprehension, Vocabulary and Verbal Fluency, Arithmetic Reasoning, Memory, and Visual Motor. While Stone presents no empirical data, he suggests that there is a correlation between the functions assessed on the S-B and the SIT. With the caution that both the SIT and S-B favor the middle class child, Stone recommends employment of a classification system to derive a deeper understanding of the meaning of a SIT IQ score.

Canfield (1972) employed a multi-letter code, similar to Meeker's technique for classification of the S-B and Wechsler Scales, to interpret SIT performances. Canfield designated ten categories, assigning corresponding letters from "a" to "j". Each test item was assigned one or more letters based on the functions supposedly involved. Canfield's ten categories include: a. Sensory and perceptual discrimination; b. Motor coordination; c. Comprehension; d. Ideation judgement; e. Practical judgement; f. Imagery; g. Comparisons; h. Vocabulary; i. Arithmetic Reasoning; and j. Memory. For example, item one-eight is coded "djg" (ideation, judgement, imagery and comparisons). While Canfield's system focuses on the operations and contents of the SIT, no evidence of category reliability or validity is presented.

Boyd (1974), also, published a classification system of the SIT. He employed Wechsler subtest descriptions as his category definitions. Utilizing item analysis, he categorized each SIT item between year four-eight and fifteen-ten as measuring: Information, Arithmetic, Similarities, Vocabulary, or Digit Span. Boyd refers to Strang's employment of the Wechsler subtests of Information, Arithmetic, Digit Span and Vocabulary, for the diagnosis of reading problems and suggests that his SIT categories can be used in a-comparable fashion. However, Boyd presents no analysis of item distributions within the designated categories or statistical evidence of the comparability of assessment a-cross age intervals.

Directors of Project Success (1975) proposed an informal system, classifying SIT items into five major categories. Auditory Memory items were classified into non-meaningful (auditory memory for number) and meaningful (auditory memory for sentences). Conceptualization was divided into seven subcategories: prepositions, size comparisons, math counting, health, math fractions, math numbers sequence and vocabulary. While the Project Success Schema attempts to discriminate specific skill areas, no evidence of sub-scale validity is presented.

Hedberg and Shapiro (1976) proposed a classification system which incorporated the Sattler S-B classification model. Content analysis was utilized to classify SIT items into seven major categories and three subcategories adhering to the Sattler definitions. A comparison made of the SIT and S-B tests suggested that both instruments assess common functions with the exception of non-verbal reasoning and visual memory. However, it was also noted that while function assessment by the SIT and S-B within age intervals is similar, items pertinent to each classification are not distributed evenly throughout the tests. Consequently, information gained from the proposed interpretive profile is limited by the structure of the SIT itself.

While the Hedberg and Shapiro study failed to explore subscale reliability, it did attempt to look informally at subscale validity. Teacher consultant summaries were compared with the SIT profiles of sixty-three children between ages five-six and seventeen-nine. Scatter analysis was employed to determine agreement between teacher diagnoses and SIT profiles. While agreement was found, statistical significance was not reported. Diagnostic tests were also administered to seven additional children to correlate SIT performances with specific diagnostic instruments. Learning quotients were calculated with a score of ninety or lower as suggestive of a deficit by which to correlate interpretive profiles and corroborative tests. Again a high percentage of agreement between learning quotients and SIT performances was presented but no significance was reported.

While the Hedberg and Shapiro study attempted to introduce empirical data regarding the use of a SIT classification schema, the lack of control over previous evaluations leading to teacher consultant summaries and the small sample to whom specified tests were administered, as well as the lack of sophistication of the statistics employed, seriously limit the generalizability of the study.

Summary of SIT Classification Systems

This selected survey of previous SIT classification systems suggests that a number of SIT classification schemes have been proposed to further the diagnostic process which are similar to previously proposed S-B classification systems. However, the legitimacy of employing them has not been substantiated.

The SIT was introduced as a quick measure of general intelligence with sufficient reliability and validity to support its employment for purposes of educational planning. Techniques of classification and scatter analysis that were developed and applied to other instruments like the S-B, have also been suggested for employment with the SIT. The employment of these techniques has grown out of a preoccupation with the belief that valuable information could be derived to supplement quantitative indices of brightness. The validity and reliability of doing so has not been sufficiently investigated. The SIT can possibly be accepted as a reasonably useful evaluative instrument, but it can not be accepted as a differential diagnostic instrument without considerable additional supportive data.

Recapitulation of Related Literature

In reviewing the pertinent literature, four areas of previous investigation have been discussed. Accordingly, the SIT has been observed to be a reliable and valid "quick" individual measure of intelligence. Secondly, the endorsement of classification systems, developed through either content or factor analysis, as a method for organizing behavioral observations, has been summarized. The continued use of classification systems, even in the absence of supportive empicial data, has been recommended by a number of investigators. Research regarding classification systems developed and employed in conjunction with S-B administrations was also summarized, due to close S-B and SIT conceptual ties.

Scatter analysis, as the means by which to quantify the behavioral observations derived from the classification systems was identified as providing potentially useful qualitative information. Scatter analysis can clarify intra- or inter-test variability. However, when scatter analysis is employed for purposes of categorical diagnosis and not just as an aid to hypothesis formulation, criteria of statistically significant differences must be met. Whether scatter analysis is employed for diagnostic purposes or for hypothesis formulation, combined area and range techniques are recommended. Further, reference points should be chosen with consideration of their stability and relevancy.

Results of S-B scatter analysis studies of both children and adults have been equivocal. The only relatively consistent relationship identified is that between scatter analysis and mental age, but exceptions were found even here. The usefulness of S-B scatter as a pathognomonic sign has not be conclusively confirmed for children or adults. Five schemes for SIT classification and scatter analysis, similar in design to those proposed for the S-B, were also discussed. However, no significant empirical evidence was found to support the incorporation of any SIT classification system in the diagnostic process. Generally, there appears to be a paucity of investigative data regarding SIT classification system-scatter analysis validity and employment.

METHOD

Hypotheses

The following null hypotheses were tested:

Ho_1: There is no statistically significant relationship between the SIT classifications (Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual Motor, and Social Intelligence-Reasoning) at the two to five, six to ten, eleven to fourteen and fifteen to adult age levels.

Ho_2: There is significant ($p \leq .05$) inter-class agreement between the distribution of functions assessed by the SIT at the two to five, six to ten, eleven to fourteen, and fifteen to adult age levels.

Ho_3: There is no significant ($p \leq .05$) inter-class agreement between the distribution of functions for the SIT and S-B items included within the: two to five year level; six to ten year level; eleven to fourteen year level; fifteen to adult year level; or two to adult level.

Ho_4: There is no significant ($r \geq .70$) relationship between the test items included within the SIT categories.

Ho_5: There is no significant difference between an individual SIT category's total reliable variance and its squared multiple correlation with the rest of the SIT categories.

Ho6: There is no significant ($p \leq .05$) relationship between test-retest category scores.

Ho7: There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery A including administration of subtests of the Illinois Tests of Psycholinguistic Abilities (ITPA) and the Wide Range Achievement Test (WRAT).

Ho8: There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery B including administration of subtests of the Detroit Tests of Learning Aptitude (Detroit) and the WRAT.

Subjects

Three ex-post-facto subject samples were utilized in this study. Sample one included twenty children at each grade level one through seven and ten children at grade eight. These subjects were chosen by random sampling without replacement from a population of children attending a Chicago north side Catholic parochial school. As judged by parish administrative personnel, the community in which this school is located is of high middle socio-economic status. One hundred percent of the graduating eighth graders from this school go on to high school and approximately seventy-five percent of those students eventually attend college.

As displayed in Table 1, the average age of the Sample One children within each grade was appropriate for December - February testings. The selection of boys and girls was relatively evenly distributed throughout the sample with the greatest disporportion of girls (n=13) to boys (n=7) selected at the fourth grade, and the greatest disporportion of boys (n=14) to girls (n=6) at the seventh grade level. The average SIT IQ of Sample One on test administration one was 117.23, with mean IQ's for all grade levels but seven, falling between plus one and plus two standard deviations above the mean (assuming a SIT standard deviation of seventeen points). The average IQ standard deviation of Sample One, on test administration one was 13.43, suggesting a restriction when compared with the population standard deviation. The average IQ of Sample One on test administration two was 120.23 with mean IQ's, again for all grade levels but seven, falling between plus one and plus two standard deviations above the mean. The average IQ standard deviation of Sample One on test two was 12.71 which also suggests restriction. These high mean IQ's and narrow standard deviations may be related to the reported socio-economic status of the community.

Sample Two and Sample Three included ninety-five children, between the ages of six to fourteen, who were given psychoeducational evaluations at a Chicago university children's service center. Children between six years and eight years eleven months were included in Sample Two. Sample Three included those children between nine years and thirteen years eleven months.

As displayed in Table 2, the proportion of boys to girls in this sample was approximately two to one. Such a ratio is not atypical of other reported learning disabled samples. As noted in table 2, the mean IQ of Sample Two was 103.13 which is reasonably close to the population mean of 100. However, Sample Two's standard deviation of 12.79 does suggest a restriction in the sample as would be expected from a select group. These children, identified as having academic problems, also live on the north side of

Table 1

Sample One Sex, Age, IQ and IQ Standard Deviations by Grade

Grades	Number Boys	Number Girls	Mean Age	Mean IQ-1	Std.D. IQ-1	Mean IQ-2	Std.D. IQ-2
1	10	10	6.65	123.90	11.85	127.00	7.87
2	8	12	7.68	123.45	10.78	122.85	9.78
3	10	10	8.51	117.00	10.29	117.70	10.06
4	7	13	9.54	117.85	15.87	121.85	14.49
5	10	10	10.55	115.55	15.38	119.65	14.48
6	11	9	11.62	115.00	12.68	118.35	13.99
7	14	6	12.58	113.05	14.54	115.20	13.79
8	6	4	13.56	117.90	13.07	118.30	14.87
Total	76	74	9.85	117.97	13.43	120.23	12.71

Table 2

Sample Two Sex, Age, IQ and IQ Standard Deviations by Grade

Grades	Number Boys	Number Girls	Mean Age	Mean IQ	Std.D. IQ-1
1	9	2	6.95	97.92	11.90
2	12	6	7.70	104.28	9.85
3	10	6	8.58	105.88	15.92
Total	31	14	7.84	103.13	12.79

Chicago. However, as the children were attending a university center and as participation was not restricted by community or parish boundaries, a wider geographic and socio-economic status is represented.

As noted in Table 3 the proportion of boys to girls in Sample Three was a little less than two to one. The mean IQ of Sample Three is noted as 98.84 which is again reasonably close to the population mean of 100. However, restricted sampling, as would be expected, is again suggested by a narrow standard deviation of 13.20 for Sample Three.

The inclusion of a client of the university center in Sample Two or Three was based upon that client's enrollement in a public or parochial regular classroom, grades first through eight, and the inclusion in their university evaluation the following tests or subtests: Detroit Tests of Learning Aptitude subtests 2, 4, and 6, or the Illinois Test of Psycholinguistic Abilities; the Wide Range Achievement Test; and the SIT.

Because ex post facto subjects were involved in this study no informed consent procedures were possible. However, participation in testing at both institutions was voluntary and both institutions publicized their testing objectives as involving not only individual evaluation but also research. Permission for access to student files for diagnoses, planning and/or research purposes was inherent

Table 3

Sample Three Sex, Age, IQ and IQ Standard

Deviations by Grade

Grades	Number Boys	Number Girls	Mean Age	Mean IQ	Std.D. IQ
3	1	3	9.33	95.75	7.63
4	6	4	9.63	101.10	7.53
5	5	2	10.95	101.29	5.38
6	8	2	11.43	99.10	22.47
7	6	3	12.57	97.22	13.85
8	6	4	13.71	97.30	12.54
Total	32	18	11.50	98.84	13.20

to this investigator's work as a visiting instructor, supervising diagnostic evaluations and remedial programs conducted at the university service center, and as the Learning Enrichment Program Director of the parochial school. Permission was granted by both the clinic director and school principal for the compilation of a research sample.

Procedure

The study was a three step investigation of the use of the SIT as an aid when conducting educational evaluations. Step One involved the development of a SIT classification schema and an anlysis of the developmental character of the test items included in the SIT. Step Two involved an investigation of the reliability of the schema developed in Step One. Step Three involved an ex-post-facto investigation of the validity of the proposed SIT classification schema.

Step One

A classification system for the SIT was developed by having three independent educational psychologists with at least five years of diagnostic experience classify each SIT item from year two upward, using category definitions and a sorting procedure, as adhering to one of the Modified Sattler Categories (Language, Memory, Conceptual Thinking, Numerical Reasoning, Social Intelligence-Reasoning, and

Visual Motor). Each test item and its category assignment is presented in Appendix A. Item classifications were then compared by the investigator to determine consensus of item assignment. In the case of lack of agreement of an item's function assessment by at least two out of three judges, an additional judge was asked to assign the test item, at which time consensus was achieved for all test items. The percentage of placement agreement within each category and the total test, by three out of three judges, two out of three judges, and two out of four judges was computed. The resultant classification schema with category designations was graphically represented on a chart referred to as a SITFILE (see Appendix A).

Next, item classifications were analyzed to determine rank ordering of categories by calculating the ratio of total test items to category items. These rank orderings were compared to Sattler's (1965) rank ordering of the S-B according to the Sattler Classification System, with a modification for the combination of Social Intelligence and Reasoning classifications. SIT classifications were also rank ordered by distribution of items within the age levels: two to five; six to ten; eleven to fourteen; and fifteen to adult. The resultant SIT rank orders within age levels were compared to S-B rank orders within similar age levels. These age levels were utilized to maintain comparability with the original Sattler S-B age levels. An intra-class coefficient was calculated between the ranking of categories across SIT age levels and compared with that calculated for the S-B Sattler categories. The Kendall Coefficient of Concordance was employed for this purpose. To evaluate inter-class correlations Kendall Tau coefficients were also calculated between SIT category ranks and between S-B category ranks within and across age groups and within the total tests.

Step Two

Initial individual administrations of the SIT to one hundred fifty (150) randomly selected students enrolled in grades one through eight were conducted and completed in December of 1979 as part of a school testing program. Individual retest administrations of the SIT to the same one hundred fifty (150) students were conducted and completed in February of 1980. Each administration of the SIT, completed in a single session, was conducted at the school in special rooms set aside for that purpose. Eight examiners, all female, were employed to administer the SIT. Five of the examiners were graduate students in special education who had successfully completed relevant courses in educational and diagnostic testing. Individual instruction on and practice with the administration of the SIT was provided to the three examiners without relevant previous training. Throughout the testing, informal discussions were led by the investigator, as a member of the school faculty, to ensure adherence to standardization procedures. Each examinee protocal was scored during the administration of the test and then later rescored by the investigator. Children were excused from their regular classes for the testing. Previous to the testing, letters were sent to the parents of all children enrolled at the school informing them of the upcoming testing.

Appropriate safeguards were taken to maintain subject anonymity. A three digit code was employed when transfering information from original protocals to SITFILES and diagnostic data sheets. The first digit indicated a child's grade placement at the time of the intial testing. The other two digits ranged from 01 to 20 as identifying numbers.

To evaluate the internal consistency reliability of the proposed SIT schema, December test results were used to compute Kuder-Richardson #20 Coefficients at grade levels; first, second, third, fourth, fifth, sixth, seventh, and eighth. As it was noted that Sample One's mean IQ deviated significantly from that of the population, the computed Kuder-Richardson #20 coefficients were corrected for restricted intelligence range. This was done by computer, according to the expression (Thorndike, 1951):

$$R_{jj} = \frac{r_{jj} + r_{jk}\left(\frac{\sigma^2_k}{s^2_k} - 1\right)}{1 + r_{jk}\left(\frac{\sigma^2_k}{s^2_k} - 1\right)}$$

where R_{jj} is the reliability coefficient for the full range of intelligence,

r_{jj} is the reliability coefficient for the restricted group,

σ^2_k is the variance of IQ in the general population ($\sigma = 17$)

and s^2_k is the variance of IQ in the restricted group.

In Step Two, Language, Memory, Conceptual Thinking and Numerical Reasoning Category specificities, at grade levels one through eight, were also calculated by subtracting a category's shared variance from its total reliable variance. The remainders, reliable specific variances, were compared to the proportion of error variance for the category. Consistent with Silverstein's (1976) argument that squared multiple correlations as an estimate of common variance are objective and unique, they were calculated at grade levels one through eight using the SIT item responses from the December test administrations. Error variance was calculated by subtracting each category's internal consistency reliability (corrected for restriction of range) from unity.

To summarize the responses within each SIT category, subjects' SIT responses were transfered to SITFILES, and the standard deviation interpretive method employed with a three step scoring procedure, resulting in one hundred fifty (150) pairs of scores for each category. The SIT scoring wheel was used to find each examinee's mental age (MA) on the chronological age index that corresponded to his IQ of 117 (mean score plus one standard deviation). The noted MA then became that examinee's year and month level point for a plus one standard deviation. In a similar fashion, the wheel was used to determine the minus one standard deviation (mean score minus one standard deviation) year and month level point for each examinee test administration. Year-month level points for plus one standard deviation and minus one standard deviation were then recorded in the appropriate spaces on the corresponding SITFILES.

After determining the plus one and minus one standard deviation year-month points for a testee for a given category, category scores were computed. The computation of each SITFILE category score involved three steps. First, correct within category responses made before the plus one standard deviation year-month point were counted (including those assumed correct below the basal). Next, to this number was added two times the number of correct within category responses made after the plus one standard deviation year-month point. From this sum was subtracted two times the number of within category errors made before the minus one standard deviation year-month point. The difference was the SITFILE category score. In this manner, SITFILE category scores were calculated for each of the six SITFILE categories for each Step Two SIT administration resulting in nine hundred (900) pairs of SIT category scores.

To obtain a measure of category stability, Language, Memory, Conceptual Thinking and Numerical Reasoning paired category scores were correlated using the Pearson Product Moment statistic. The resultant correlations were then also corrected for restriction of IQ range according to the previously described expression (Thorndike, 1951). Means and standard error of measurements (SE_m) were also calculated by computer for each category to reflect consistency of performance.

Step Three

The age scores of ninety-five children, previously administered either Detroit subtests two, four and six (for children nine to fourteen) or ITPA subtests Auditory Reception, Auditory Association, Auditory Sequential Memory, Verbal Expression, and Grammatic Closure (for children six to nine) and the WRAT (for all children included in Step

Three) were correlated with SITFILE Language, Memory, Conceptual Thinking, and Numerical Reasoning Category scores. The decision as to determination of the qualifying test battery was made giving consideration to both test and norm limitations as well as diagnostic convention.

All test results utilized in Step Three were generated at a Chicago university children's educational service center as part of a parent initiated psychoeducational evaluation, necessitated by a child's school difficulties. Individual cubicles were utilized to provide privacy during testing. Children between six and eight were generally excused from their regular classes to participate in the testing which was conducted at the center between 9:00A.M. and 1:00 P.M.. Children nine and up were brought to the center between 3:30 P.M. and 7:00 P.M. after completing a regular school day.

Forty-nine examiners were involved in Step Three testing. These twelve men and thirty-seven women were all special education Master's Candidates enrolled in a Learning Disabilities Diagnostic Practicum. Consequently, all examiners had previously completed courses which provided specific training and experience in the administration of all the diagnostic tests included in this study. Additionally, all examiners were re-instructed on each test's administration and interpretation. Follow-up discussions and close supervision by educational specialists on the university's staff further assured adherence to standardization procedures. Examinees were assigned each examiner with special consideration given to an examiner's preference for testing experience with a certain age group as well as to scheduling constraints.

The battery of tests given each child began with the administration of the SIT. The order of the administration of the rest of the tests was determined by the diagnostician conducting the child's evaluation. Additional tests of cognitive ability, information processing and academic functioning were also administered, consistent with the Service Center's objectives. Testing was conducted over a minimum of four, one-and-one-half-hour to two-hour sessions over at least three weeks. Time for breaks was provided during each session to minimize fatigue and optimize validity of diagnostic test results.

Each child's test protocols were scored by the test administrator and then rescored by a member of the University staff. A record of each child's relevant scores on the specified tests and subtests was provided this investigator along with a copy of the child's SIT protocol. Each child's SIT item scores were transcribed on individual SITFILES to provide a visual representation of his/her performance within each of the categories of the proposed SIT classification schema. The standard deviation, chronological age method and three step category score computation was again utilized to quantify individual SITFILES (see earlier discussion for details). The Pearson Product Moment statistic was employed to obtain a measure of correlation between each SIT category and the administered diagnostic tests.

Instrumentation

The tests included in the present investigation included the following: The Slosson Intelligence Test (1963); The Sattler (1965) Stanford Binet Classification Schema; The Modified Sattler Classification Schema; The Detroit Tests of Learning Aptitude (Detroit) (1967); The Illinois Test of Psycholinguistic Ability (ITPA) (1968); and the Wide Range Achievement Test (WRAT) (1965).

The Slosson Intelligence Test: As previously described the SIT is an abbreviated intelligence test. It consists of one hundred and ninety-four (194) items age graded from five months to twenty-seven years. Administration of the SIT results in a ratio IQ with a mean of 100. A more complete description of the SIT can be found in Section two - Survey of the Literature.

The Sattler Stanford Binet Classification Schema: The Sattler classification was developed to assist test administrators in interpreting S-B results. It is a classification system based on categories developed with attention to face validity. Sattler's categories include:

Language: This category includes tests related to maturity of vocabulary in relation to the prekindergarten level, extent of vocabulary referring to the number of words the child can define, quality of vocabulary measured by such tests as abstract words, rhymes, word naming, and definitions, and comprehension of verbal relations.

Memory: This category contains meaningful, nonmeaningful and visual memory tests. The tests are considered to reflect rote auditory memory, ideational memory, and attention span.

Conceptual Thinking: This category, while closely associated with language ability, is primarily concerned with abstract thinking. Such functions as generalization, assuming an "as if" attitude, conceptual thinking, and utilizing a categorical attitude are subsumed.

Reasoning: This category contains verbal and non-verbal reasoning tests. The verbal absurdity tests are the prototype for the verbal reasoning tests. The pictorial and orientation problems represent a model for the nonverbal reasoning tests. Reasoning includes the perception of logical relations, discrimination ability and analysis and synthesis. Spatial reasoning may also be measured by the orientation tests.

Numerical Reasoning: This category includes tests involving arithmetic reasoning problems. The content is closely related to school learning. Numerical reasoning involves concentration and the ability to generalize from numerical data.

Visual-Motor: This category contains tests concerned with manual dexterity, eye-hand coordination, and perception of spatial relations. Constructive visual imagery may be involved in the paper folding test. Nonverbal reasoning ability may be involved in some of the visual-motor tests.

Social Intelligence: This category strongly overlaps with the reasoning category, so that consideration should be given to tests classified in the latter as also reflecting social comprehension. The area of social intelligence includes aspects of social maturity and social judgment; whereas, the items concerning obeying simple commands, response to pictures, and comparison reflect social maturity.

While Sattler did not employ either judges or factor analysis to achieve a reliability estimate of the categories, Silverstein (1965) compared the Sattler and Valett schemes and noted seventy-five percent agreement of the total test suggesting a satisfactory degree of reliability of the item assignments.

The Modified Sattler Classification Schema: The Modified Schema includes the following categories:

> Language: This category includes tests related to maturity of vocabulary in relation to the prekindergarten level, extent of vocabulary referring to the number of words the child can define, quality of vocabulary measured by such tests as abstract words, rhymes, word naming, and definition, and comprehension of verbal relations.
>
> Memory: This category contains meaningful and non-meaningful memory tests. The tests are considered to reflect rote auditory memory, ideational memory, and attention span.
>
> Conceptual Thinking: This category, while closely associated with language ability, is primarily concerned with abstract thinking. Such functions as generalization, assuming an "as if" attitude, conceptual thinking, and utilizing a categorical attitude are subsumed.
>
> Social Intelligence-Reasoning: This category contains verbal and non-verbal reasoning tests. Reasoning includes the perception of logical relations, discrimination ability, and analysis and synthesis. The area of social intelligence includes aspects of social maturity and social judgment; whereas, the items concerning obeying simple commands, response to pictures, and comparison reflect social maturity.
>
> Numerical Reasoning: This category includes tests involving arithmetic reasoning problems. The content is closely related to school learning. Numerical

reasoning involves concentration and the ability to generalize from numerical data.

Visual-Motor: This category contains tests concerned with manual dexterity, eye-hand coordination, and perception of spatial relations. Nonverbal reasoning ability may be involved in some of the visual-motor tests.

A comparison of these categories with those of the original Sattler categories reveals significant similaritites. However, on the modified classification schema the old Sattler Social Intelligence and Reasoning categories have been combined to form one category. This collapsing of the two categories into one was done in accordance with Sattler's (1965) observation that the two categories strongly overlapped. Other changes in the categories' definitions were instituted to maintain concordance between definitions and the SIT test design. For example, there are no visual memory items included in the SIT and therefore, this component of the Sattler Memory Category was deleted from the Modified Sattler Memory Category for the SIT.

Because the modified classification schema for the SIT is a new instrument, no previous measures of its reliability or validity are available. However, measures of item assignment reliability have been generated as discussed in the Step One procedure section to follow. Test-retest measures of category internal reliability will be evaluated in Step Two of this study.

<u>The Detroit Tests of Learning Aptitude (1967)</u>: The Detroit is intended to assess the learning capabilities of individuals from three years of age through adulthood by evaluating what Baker and Leland refer to as the special phases of mental facilitites. The 1967 revision of the 1938 edition is composed of nineteen subtests, each of which must be administered individually. The authors have included subtests which they feel assess eight psychological functions: comprehension and reasoning, practical judgment, verbal ability, time and space relationships, number ability, attentive ability (auditory), attentive ability (visual) and motor ability. All nineteen subtests are not intended for administration to a single individual. The authors of the Detroit recommend selecting between nine and thirteen subtests for administration depending upon the age of the individual and subtest relevancy to suspected learning difficulties. Special training is necessary both for appropriate administration of the test as well as for its interpretation. Administration of selected subtests can take variable amounts of time and is expected to result in a pattern of scores useful for diagnostic interpretation. The Detroit has been increasingly utilized to evaluate the older (nine and a half years up) child who is experiencing educational problems.

Scoring of the subtests results in mental age scores ranging from three years to nineteen years. Norms are presented in age levels by three month increments. Subtests' mental ages are to be ordered so as to determine a median MA. This median mental age is then to be inserted into the formula $\frac{MA}{CA}$ (with a constant chronological age of fifteen years zero months for all individuals at or above the fifteen year chronological age level) resulting in a Detroit-IQ.

The Detroit has been well accepted, but questions over its standarization have been raised. With students drawn from the Detroit Public Schools, fifty pupils, with IQ's between ninety and one hundred and ten from every age level, were initially included for norming purposes. Subsequently, an additional one hundred and fifty students at each age level were included. The authors report a retest reliability coefficient of .959 over a five month period and a .91 correlation between Detroit IQ's and S-B (Form LM) IQ's. On a sample with a restricted range of scores they report a Detroit IQ standard deviation of eight points.

The subtests included for the purposes of evaluating concurrent validity were chosen with primary consideration given to the assessing of the areas A) verbal ability and B) auditory attentive ability. They are:

Number 2 - Verbal Absurdities: This subtest consists of a series of absurd statements about which the examinee must state what it is that is foolish.

Number 4 - Verbal Opposites: This subtest consists of a list of ninety-six words. The examiner says a word from the list and the examinee is to say its antonym.

Number 6 - Auditory Attention Span for Unrelated Words: This subtest consists of seven sets of unrelated words ranging in length from two to eight words. The subject is to repeat them correctly after their presentation by the examiner.

The Illinois Test of Psycholinguistic Abilities: The ITPA is a content scale designed to test the cognitive skills which are involved when a form of communication transaction is necessitated. For children between two to ten years of age, its design is based upon Osgood's psychological model assessing levels of organization, psycholinguistic processes and channels of communication. The test, used primarily with children encountering learning difficulties, consists of ten main subtests and two supplementary subtests all of which must be administered individually by specially trained examiners. Administration of the total test takes approximately one hour.

The ITPA was normed on nine hundred sixty-two (962) children described as free from physical handicaps or emotional disturbances, whose average IQ's ranged between

84 and 116. Limitations of the standardization population to "normal" children is observed to have resulted in a restriction in the range of scores as well as lower reliabilities. Individual subtest reliability coefficients are subsequently presented with their subtest descriptions. The ITPA provides three types of norms for interpreting test scores: age norms, scaled scores, with a mean of 36 and a standard deviation of 6, and composite psycholinguistic age norms.

The subtests considered relevant for this investigation include:

<u>Auditory Reception</u>: This subtest is intended to evaluate an individual's ability to gain meaning from auditorily received stimuli. The test authors report high internal consistency coefficients for this subtest with a median coefficient of .95 after a correction for the restricted intelligence range. Test-retest reliability coefficients (over a five month period) are reported as ranging from .63 to .79.

<u>Auditory Association</u>: This subtest, through the use of verbal analogies, measures a child's ability to relate auditorily received stimuli in a meaningful way. Author reported corrected internal consistency coefficients range between .86 to .94, with five month test-retest reliabilities from .83 to .90.

<u>Auditory Sequential Memory</u>: Success on this subtest requires the ability to reproduce, immediately after presentation, sequences of digits ranging in length from two to eight digits. A child is allowed two trials on each sequence but more credit is given for success on the first trial than on the second. The authors report a median internal consistency coefficient of .90 (corrected for restricted range of intelligence) with five month stability coefficients between .75 and .89.

Verbal Expression: This subtest assesses, through the use of common objects, a child's ability to convey ideas in words. The test's median internal consistency is reported as .85 with stability coefficients over a five month period ranging between .63 to .74.

Grammatic Closure: This subtest is designed to measure a child's ability to make use of the redundancies of oral language to internalize syntax and grammatic inflections. Grammatic Closure subtest internal consistency coefficients are reported in the .80's for eight age groups of average intelligence children. Five month stability coefficients for three age groups, four year olds, six year olds, and eight year olds of .72, .78, and .87 are reported.

Wide Range Achievement Test (1965): The WRAT is basically an individually administered assessment device intended to measure an individual's proficiency in the basic school subjects of reading (word recognition and pronunciation), written spelling (copying marks, writing name and writing words from dictation), and arithmetic (counting, reading numbers, symbols, solving oral problems and performing written computations). Preceding the 1965 revised edition are the 1936 and 1946 editions. The 1965 edition is divided into two levels: Level One for children age five years zero months to eleven years eleven months, and Level Two for individuals age twelve years zero months to adulthood. It is a relatively easy test to administer requiring minimal examiner training. Administration of the entire test takes between twenty and thirty minutes.

Norms for the WRAT are not based on a representative national sample. However, large samples of 5,868 individuals for Level One and 5,933 individuals for Level Two, drawn from Deleware, Pennsylvania, New Jersy, Maryland, Florida, Washington.and California were utilized for standardization purposes: grade norms equivalent to mental ages; standard scores with a mean of 100 and a standard deviation of 15; and percentiles.

Investigations of WRAT reliabilities are reported by the test authors as ranging from .92 to .98 for the reading and spelling subtests and .85 to .92 for the arithmetic subtest. The authors also report WRAT validity coefficients of .81 and .93 with the California Achievement Test. Henderson, Butler and Goffing (1969) report WRAT validity coefficients between .38 and .61 with the WISC. Elliot (1969) reports validity coefficients of .56 to .79 between the Pictorial Test of Intelligence and the WRAT.

RESULTS

Step One: Development of a SIT Classification Schema

In Step One, SIT items from year two to twenty-seven were assigned to categories on the basis of content analysis. Appendix A includes all assigned SIT items and their classification assignments. A detailed analysis of the data from Appendix A is presented in Tables 4 through 9.

Table 4 indicates the cumulative sorting decisions by judges by category. The greatest within category concurrence by three judges was achieved with the assignment of four items to the Visual-Motor Category. Within the Language Category, concordance by three judges on 91.8% of the forty-nine assigned test items was achieved. One-hundred percent agreement of function assessment was not achieved on only two out of thirty-three items identified as assessing Numerical Reasoning. In order to determine final test item assignments of eighteen items to the Conceptual Thinking Category, it was necessary to seek a placement decision on one item by a fourth judge. This referral to a fourth judge is reflected in the lower concordance percentage (66.7) by three judges. Two items upon which function agreement was not reached by the initial three judges necessitated these items' submission to a fourth judge and resulted in one-hundred percent placement agreement on only

52% of the twenty-five items included within the Memory Category. Only four out of the nineteen items placed within the Social Intelligence-Reasoning Category were so placed by the concordance of three judges. Thirteen items assigned to this category were placed upon the agreement of two judges, while assignment by a fourth judge was necessary to determine the placement of two additional items. In total, 73.7% concordance by three judges determined the placement of the one hundred forty-eight (148) SIT items considered. Failure to achieve consensus by three judges necessitated the consultation of a fourth judge for five or 3.4% of the one hundred forty-eight (148) SIT items.

A comparison of the distributions of S-B and SIT items for the six classifications (Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor, and Social Intelligence-Reasoning) was undertaken as presented in Table 5. Accordingly, Language occupies rank one for the SIT (33%); while it occupies rank two for the S-B (26%). The Visual-Motor Category is least represented of the six categories for the SIT with 3% of the items; while the Numerical Reasoning Category is least represented of the S-B categories with 9% of the items. Memory items occupy rank three on both the SIT and S-B. Conceptual Thinking items constitute 13% of the S-B items and 12% of the SIT items.

Table 4

Juried SIT Item Classification Decisions*

Category	Number of SIT Items	Concordance by 3/3 Judges Number	Concordance by 3/3 Judges Percent	Concordance by 2/3 Judges Number	Concordance by 2/3 Judges Percent	Concordance by 2/4 Judges Number	Concordance by 2/4 Judges Percent	Cumulative Percent
Language	49	45	91.8	4	8.2	0	0	100.0
Memory	25	13	52.0	10	40.0	2	8.0	100.0
Conceptual Thinking	18	12	66.7	5	27.8	1	5.6	100.1
Numerical Reasoning	33	31	93.9	2	6.1	0	0	100.0
Visual-Motor	4	4	100.0	0	0	0	0	100.0
Social Intelligence-Reasoning	19	4	21.1	13	68.4	2	10.5	100.0
Total	148	109	73.7	34	23.0	5	3.4	100.1

*Based on data from Appendix A.

Table 5

Rank Order of S-B and SIT Classifications: Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor, Social Intelligence-Reasoning

	S-B**			SIT*		
Rank	Category	Number	Percent	Category	Number	Percent
1	Social Intelligence Reasoning	36	30	Language	49	33
2	Language	32	26	Numerical Reasoning	33	22
3	Memory	17	14	Memory	25	17
4	Conceptual Thinking	16	13	Social Intelligence Reasoning	19	13
5	Visual-Motor	12	10	Conceptual Thinking	18	12
6	Numerical Reasoning	9	7	Visual-Motor	4	3
Total		122	100		148	100

*Based on data from Appendix A.

**Alternate tests excluded (Sattler, 1965)

Tables 6 and 7 present rank orderings based on distribution percentages for the S-B and SIT six category classification systems by age level groupings. The four age levels were utilized to facilitate S-B and SIT comparisons. A comparison of the four age levels indicates that on the S-B, Social Intelligence-Reasoning items occupy rank one in age groups two to five, six to ten and eleven to fourteen with percentages ranging from 27 to 38. However, at the fifteen to adult S-B age level, a decrease to the 3.5 rank with only 15% representation is noted for Social Intelligence-Reasoning items. By comparison, for the two to five age level, as noted in Table 7, SIT Social Intelligence Reasoning items occupy rank one with 37% of the items. A SIT Social Intelligence-Reasoning distribution drop to the fifth rank is noted at the six to ten age level with no SIT Social Intelligence-Reasoning items presented after the six to ten year level. Language items occupy rank two at levels: six to ten, and eleven to fourteen, and rank one at level fifteen to adult, for both the S-B and SIT with S-B percentages of 20%, 30%, and 31% and SIT percentages of 23%, 29%, and 57%. Within the S-B, Language items occupy rank two at the two to five age level. Within the SIT, Language items occupy rank 3.5 at the two to five age level. Within both the S-B and SIT Visual-Motor items have the greatest frequency at the two to five age level, decrease in frequency at the six to ten age level, and are not present in either SIT or S-B age levels eleven to fourteen or fifteen to adult. Within the S-B, Numerical Reasoning items increase from not present at the two to five age level, to 10% at the six to ten age level, 9% at the eleven to fourteen age level, and then 15% at the fifteen to adult level. By contrast, Numerical Reasoning items are present within the SIT two to five age level (14%) and decrease to 13% at the six to ten level, then increase to 38% at the eleven to fourteen level, decreasing again at the fifteen to adult level to 28%. Memory items on the S-B are ranked: 4, 3.5, 3, and 5. Memory items on the SIT are ranked: 3.5, 1, 3, and 3. Conceptual Thinking items on the S-B fluctuate from 5% at the two to five level, to 17% at the six to ten level, decrease to 9% at the eleven to fourteen age level and then increase to 12% at the fifteen to adult level. Within the SIT, Conceptual Thinking items make up 14% of all items at the two to five level, 20% of the items at the six to ten level, 13% of the items at the eleven to fourteen level and 6% of the items at the fifteen to adult level.

From Tables 6 and 7 and the preceding discussion, it is noted that while there are similarities, neither the SIT nor S-B measure the same functions to the same extent at each age level. To test null hypothesis one (There is no relationship between the SIT classifications (Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor and Social Intelligence-Reasoning) at the two to five, six to ten, eleven to fourteen and fifteen to adult age levels.) a Kendall Coefficient of Concordance was calculated by computer, thus quantifying the extent to which all SIT category ranks at the different age levels tended to agree. Null hypothesis one is rejected since the resultant coefficient of .52 suggests a moderate degree of concordance among the SIT category ranks at the four age levels. The variance of the rank sums is fifty-two percent of the maximum possible. A Kendall Coefficient of Concordance was also calculated for the S-B using the data from Table 6. The resultant coefficient of .65 suggests a moderately high degree of concordance for the S-B as well.

Table 6

Rank Order of S-B Classifications**: Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor and Social Intelligence-Reasoning by Age Level Groupings

	2 to 5 Years			6 to 10 Years			11 to 14 Years			15 Years to Adult		
Rank	Category	#	%	Category	#	%	Category	#	%	Category	#	%
1	Social Intelligence Reasoning	16	38	Social Intelligence Reasoning	8	27	Social Intelligence Reasoning	8	35	Language	8	31
2	Language	11	26	Language	6	20	Language	7	30	Conceptual Thinking	7	27
3	Visual-Motor	9	21	Memory	5	17	Memory	5	17	Numerical Reasoning	4	15
4	Memory	4	10	Conceptual Thinking	5	17	Conceptual Thinking	2	9	Social Intelligence Reasoning	4	15
5	Conceptual Thinking	2	5	Numerical Reasoning	3	10	Numerical Reasoning	2	9	Memory	3	12
6	Numerical Reasoning	0	0	Visual-Motor	3	10	Visual-Motor	0	0	Visual-Motor	0	0
Total		42	100		30	101		24	100		26	100

**Based on data from Sattler S-B Classification Schema (1965)

Table 7

Rank Order of SIT Classifications*: Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor and Social Intelligence-Reasoning by Age Level Groupings

	2 to 5 Years			6 to 10 Years			11 to 14 Years			15 Years to Adult		
Rank	Category	#	%	Category	#	%	Category	#	%	Category	#	%
1	Social Intelligence Reasoning	16	37	Memory	9	30	Numerical Reasoning	9	38	Language	29	57
2	Language	6	14	Language	7	23	Language	7	29	Numerical Reasoning	14	28
3	Numerical Reasoning	6	14	Conceptual Thinking	6	20	Memory	5	21	Memory	5	10
4	Memory	6	14	Numerical Reasoning	4	13	Conceptual Thinking	3	13	Conceptual Thinking	3	6
5	Conceptual Thinking	6	14	Social Intelligence Reasoning	3	10	Social Intelligence Reasoning	0	0	Social Intelligence Reasoning	0	0
6	Visual-Motor	3	7	Visual-Motor	1	3	Visual-Motor	0	0	Visual-Motor	0	0
Total		43	100		30	99		24	101		51	101

*Based on data from Appendix A.

Next, using the data of Table 7, null hypothesis two (There is significant inter-class agreement between the distribution of functions assessed by the SIT at the two to five, six to ten, eleven to fourteen, and fifteen to adult age levels.) was tested by calculating Kendall Tau coefficients. The resultant coefficients are presented in Table 8.

The SIT tau coefficients between age levels are: two to five and six to ten, .09 ($p \geq .05$); six to ten and eleven to fourteen, .41 ($p \geq .05$); and eleven to fourteen and fifteen up, .86 ($p \leq .05$). These coefficients indicate that the distribution of category items for age levels two to five and six to ten, and six to ten and eleven to fourteen, are not significantly related; but that they are related for the eleven to fourteen and fifteen to adult age levels. There is also no significant relationship noted between the distribution of test items for age levels: two to four and eleven to fourteen, two to four and fifteen to adult, and five to ten and fifteen to adult. Consequently, null hypothesis two is rejected except between the age interval eleven to fourteen and fifteen to adult.

Table 8

SIT and S-B Inter-Class Kendall Tau Coefficients

By Age Levels

τ Probability	2-5/6-10	6-10/11-14	11-14/15 up	2-5/11-14	2-5/15 up	6-10/15 up
SIT with SIT*	.09 .82	.41 .25	.86 .02	.00 1.00	.00 1.00	.55 .13
S-B with S-B**	.65 .08	.89 .02	.36 .33	.55 .13	.00 1.00	.45 .24

*Calculations based on data from Table 7

**Calculations based on data from Table 6

The data from Tables 5, 6, and 7 was employed to test null hypothesis three (There is no significant ($p \geq .05$) interclass agreement between the distribution of functions for the SIT and S-B items included within the: two to five year level; six to ten year level; eleven to fourteen year level; fifteen to adult year level; or two to adult level). The tau coefficient for the total SIT and S-B is -.14 ($p \geq .05$). The tau coefficients between the SIT and S-B at the individual age levels are: two to five, .26 ($p \geq .05$); five to ten, .22 ($p \geq .05$); eleven to fourteen, .07 ($p \geq .05$); and fifteen to adult, .50 ($p \geq .05$). These coefficients indicate that the distribution of items within the SIT and S-B tend not to be related. Consequently, null hypothesis three is not rejected.

Table 9

Kendall Tau Coefficients Between SIT and S-B Category Rankings by Four Age Levels and by Total Tests

Age Levels	SIT with S-B τ	Probability
Two years to Five years	.26	.50
Six years to Ten years	.22	.56
Eleven years to Fourteen years	.07	.85
Fifteen years to Adult	.50	.17
Total Tests	-.14	.70

Summary of Step One Results

In Step One a SIT classification system was developed based upon content analysis concordance with a high percentage (.737) of 100% placement agreement by three judges. An intra-class coefficient of .52 lead to the rejection of null hypothesis one, suggesting that similar functions are tested within the SIT at the four specified age levels. However, failure to not reject null hypothesis two indicates that while similar functions are tested throughout the SIT, the distribution of the different functions between all age levels but eleven to fourteen and fifteen to adult are not significantly related. Finally, low tau coefficients (-.14 to .50) and the consequent failure to reject null hypothesis three suggest differences in the underlying developmental structure of the SIT and S-B.

Step Two: Reliability of the Proposed SIT Classification Schema

In Step Two the reliability of the Step One classification system was investigated. Coefficients were computed for internal reliability, category specificity, and stability reliability. However, as noted in Table 5 only four Visual-Motor items are included in the SIT, all before year eight. Nineteen Social Intelligence-Reasoning items are included in the SIT, but only three of these items occur after year five. Due to the limited number of items assessing both of these areas, their failure to span the grades one through eight, and basal and ceiling SIT design, coefficients were computed for internal reliability, category specificity, and for stability reliability, for only the Language, Memory, Conceptual Thinking and Numerical Reasoning SITFILE categories.

Internal consistency reliability refers to consistency in results throughout a test during a single administration. Using the data obtained from the December testings of one hundred-fifty (150) children, grades one through eight, in which items were scored as "passed" or "failed", Kuder-Richardson #20 coefficients were computed. Table 10 presents the obtained internal consistency coefficients for the SITFILE categories Language, Memory, Conceptual Thinking and Numerical Reasoning, as well as the coefficients corrected for a restricted range of intelligence.

For first graders, as seen in Table 10, low obtained coefficients ranging between .13 to .51 and corrected coefficients between .37 to .72 suggest inconsistencies in item performance and high variable error. Across no other grade level, for all four categories, are such low coefficients noted. Corrected coefficients for the Language Category (with the exception of that of first graders) range between .72 and .86. Corrected coefficients for the Memory Category (with the exception of that of first graders) range between .70 and .80. Corrected coefficients for the Numerical Reasoning Category (with the exception of that of first graders) range between .77 to .90. Lower corrected coefficients ranging from .54 to .79 obtained on the Conceptual Thinking classification suggest category unreliability, particularly for children at grade levels: three, four, five and seven. The data included in Table 10 indicates a rejection of null hypothesis four (There is no significant ($r \geq .70$) relationship between the test items included within the SIT categories.) for children in grades two through eight except for the Conceptual Thinking Category for children beyond second grade.

Category specificity was investigated in order to test null hypothesis five: (There is no significant difference between an individual SIT category's total reliable variance and its squared multiple correlation with the rest of the SIT categories). Table 11 presents the amount of specificity for each of the four SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Categories, along with the error variance for each category. Cohen (1959) suggested informal rules for evaluating the sufficiency of subtest or category specificity. Accordingly, a category's reliable specific variance should equal .25 or more of the total variance and should exceed its error variance. Kaufman (1976) extended Cohen's rules suggesting that a subtest or category had ample specificity if it met both of Cohen's conditions, adequate specificity if it met one of Cohen's criteria, and inadequate specificity if it met neither.

Table 10

Kuder-Richardson #20 Internal Consistency Reliability for SITFILE Categories

Language, Memory, Conceptual Thinking and Numerical Reasoning by Grade

	Language		Memory		Conceptual Thinking		Numerical Reasoning	
Grade	Obtained	Corrected	Obtained	Corrected	Obtained	Corrected	Obtained	Corrected
1	.47	.68	.44	.69	.51	.72	.13	.37
2	.42	.73	.54	.80	.55	.79	.67	.85
3	.55	.80	.54	.78	.18	.56	.64	.83
4	.81	.83	.69	.72	.49	.54	.74	.77
5	.80	.83	.74	.78	.50	.57	.77	.80
6	.55	.72	.67	.80	.44	.64	.75	.84
7	.82	.86	.63	.70	.46	.58	.73	.79
8	.69	.80	.60	.74	.52	.66	.64	.90

Accordingly, Table 11 indicates that across all grade levels with the exception of the fifth and seventh grades, the Language Category has ample specificity, and that at the fifth and seventh grade levels it has adequate specificity. Across all grade levels but two the Memory Category also evidences ample specificity. For second graders on the Memory Category only adequate Memory specificity is noted. Within the Conceptual Thinking Category ample specificity is suggested for first through third graders; however, beyond fourth grade this category's specificity appears inadequate. Within the Numerical Reasoning Category for grade levels third, fifth, sixth and eighth, ample specificity is noted; for first and second graders adequate specificity is suggested, but for fourth and seventh graders inadequate specificity is indicated. Consequently, null hypothesis six is rejected across all four categories for: first, second and third graders. It is also rejected for the Language Category for fourth through eighth graders, for the Memory Category for fourth through seventh graders, and for the Numerical Reasoning Category for fifth, sixth and eighth graders.

Table 11

Language, Memory, Conceptual Thinking and Numerical Reasoning Category

Specificity by Grade

Grade	Language		Memory		Conceptual Thinking		Numerical Reasoning	
	Specific Variance	Error Variance	Specific Variance	Error Variance	Specific Variance	Error Variance	Specific Variance	Error Variance
1	.44	.32	.47	.31	.40	.20	.32	.63
2	.29	.27	.22	.20	.28	.21	.21	.15
3	.60	.20	.64	.22	.39	.45	.52	.17
4	.53	.18	.12	.28	.06	.46	.02	.24
5	.21	.18	.32	.22	-.01	.43	.44	.20
6	.29	.28	.37	.20	.18	.36	.40	.16
7	.16	.14	.28	.30	.05	.42	.14	.21
8	.33	.20	.08	.26	.11	.34	.48	.10

Stability reliability is determined in order to evaluate how constant scores can be expected to be if testing is repeated after a specific time lapse. For the SITFILE stability study, a two to two-and-one-half month test-retest interval was employed. Pearson stability coefficients calculated for the four categories: Language, Memory, Conceptual Thinking and Numerical Reasoning, are presented in Table 12. These stability coefficients, corrected for restricted intelligence range are presented in Table 13.

Stability coefficients for children at all grade levels in all four categories except Numerical Reasoning for first graders are significant ($p \leq .05$) ranging from .54 to .92 for Language; .51 to .97 for Memory: .52 to .85 for Conceptual Thinking; and .42 to .97 for Numerical Reasoning. Thus, null hypothesis six (There is no significant ($p \leq .05$) relationship between test-retest category scores.) is rejected. Significantly, all stability coefficients in Table 13, corrected for restricted intelligence range (except that for Numerical Reasoning for first graders) are above .70. Those for the Language Category are .78 to .95: for Memory .72 to .93; for Conceptual Thinking .73 to .88; and for Numerical Reasoning .58 to .98. Consequently, all four categories, except Numerical Reasoning for first graders, meet at least minimum research standards (Nunnally, 1978).

Table 12

Obtained Stability Reliability of Language, Memory, Conceptual Thinking and Numerical Reasoning Categories by Grade

r / Probability	Language	Memory	Conceptual Thinking	Numerical Reasoning
Grade	SIT-1/SIT-2	SIT-1/SIT-2	SIT-1/SIT-2	SIT-1/SIT-2
1	.66 .01	.51 .02	.52 .02	.42 .04
2	.54 .01	.82 .01	.62 .01	.79 .01
3	.87 .01	.86 .01	.74 .01	.83 .01
4	.90 .01	.92 .01	.82 .01	.83 .01
5	.92 .01	.68 .01	.76 .01	.88 .01
6	.72 .01	.82 .01	.68 .01	.81 .01
7	.87 .01	.81 .01	.85 .01	.84 .01
8	.92 .01	.82 .01	.75 .01	.97 .01
Total	.92 .01	.84 .01	.82 .01	.92 .01

Table 13

Corrected Stability Reliability of Language, Memory, Conceptual Thinking, and Numerical Reasoning Categories by Grade

Grade	Language	Memory	Conceptual Thinking	Numerical Reasoning
1	.80	.72	.73	.58
2	.78	.92	.82	.91
3	.94	.93	.86	.92
4	.91	.93	.83	.84
5	.93	.73	.79	.89
6	.82	.89	.80	.88
7	.90	.85	.88	.88
8	.95	.88	.83	.98
Total	.93	.87	.85	.93

Means, Standard deviations, and standard error of measurements of the four categories, Language, Memory, Conceptual Thinking and Numerical Reasoning, as calculated from the six hundred (600) December SIT category scores are presented in Table 14.

A review of the data, from age to age, within the four categories reveals a systematic increase of mean scores paralleling that of grade level placement. For Language, mean scores range from 2.5 to 6.7; for Memory, from 2.8 to 4.5; for Conceptual Thinking, from 11.3 to 16.2; and for Numerical Reasoning, from 5.9 to 20.5. Language Category standard deviations range between 2.5 to 6.8; while, the Memory Category standard deviations cluster close to 3.0 and 4.0. The Conceptual Thinking Category's noted standard deviations range between .9 to 6.5. All obtained standard error of measurements, while varying from category to category and grade level to grade level, are reasonably small ranging from .20 to 2.06.

Table 14

Means, Standard Deviations (Std.D.) and Standard Error of Measurements (SEm) of December SITFILE Scores on the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories by Grade for Sample One

	Language			Memory			Conceptual Thinking			Numerical Reasoning		
Grade	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D	SE_m
1	13.7	2.5	.56	10.1	3.1	.70	11.3	2.1	.46	5.9	.9	.20
2	15.1	3.2	.71	14.0	3.4	.75	11.7	2.5	.55	8.7	3.4	.76
3	13.9	3.1	.69	14.5	3.6	.80	12.5	1.6	.37	8.7	3.0	.67
4	20.3	6.7	1.50	15.8	4.1	.91	13.0	2.3	.52	10.7	4.5	1.00
5	21.6	6.0	1.35	16.0	4.5	1.00	14.3	2.2	.48	12.6	5.0	1.12
6	24.5	4.4	.98	17.7	3.8	.86	15.0	2.4	.54	14.5	4.7	1.05
7	25.1	6.3	1.41	17.5	3.3	.73	15.7	2.1	.47	16.7	5.4	1.20
8	27.9	4.0	1.28	18.7	2.8	.90	16.2	1.6	.51	20.5	6.5	2.06
Total	19.7	6.8	.56	15.2	4.3	.35	13.5	2.7	.22	11.7	5.9	.48

Table 15 presents the means, standard deviations and standard error of measurements of the February SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Category scores. For the Language Category, mean scores ranged from 14.1 to 28.5, standard deviations from 3.1 to 7.1, and standard error of measurements from .58 to 1.48. In the Memory Category, mean scores ranged from 11.9 to 19.6, standard deviations from 3.00 to 4.1, and standard error of measurements from 2.95 to 4.05. Mean scores for the Conceptual Thinking Category ranged from 10.7 to 16.3, with standard deviations between 1.5 to 2.8 and standard error of measurements from .22 to .61. For the Numerical Reasoning Category, mean scores ranged from 6.1 to 20.5, standard deviations from 1.3 to 8.3 and standard error of measurements ranged from .28 to 1.51.

A comparison of the mean scores on the December and February testings (Tables 14 and 15) reveals small maximum mean score gains: Language 1.7; Memory 1.8; Conceptual Thinking 1.2; and Numerical Reasoning 1.1. Standard deviations and standard error of measurements are also similar for the two testings. When comparing the December obtained standard deviations and standard error of measurements of the total sample with those obtained in February, the Language standard deviation increased .3 and the standard error of measurement .02; the Memory standard deviation decreased .2 and the standard error of measurement .02; the Conceptual

Table 15

Means, Standard Deviations (Std.D.) and Standard Error of Measurements (SEm) of February SITFILE Scores on the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories by Grade for Sample One

	Language			Memory			Conceptual Thinking			Numerical Reasoning		
Grade	Mean	Std.D.	SE_m	Mean	Std.D	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m
1	14.1	3.1	.69	11.9	3.7	.82	10.7	1.7	.37	6.1	1.3	.28
2	15.1	3.3	.73	14.1	3.6	.81	12.1	1.6	.36	8.7	3.5	.78
3	13.8	3.2	.71	15.3	3.9	.86	12.9	1.6	.36	9.1	3.8	.85
4	21.4	6.0	1.34	16.3	3.8	.86	13.7	2.0	.45	11.8	4.6	1.03
5	23.4	5.8	1.30	16.9	3.1	.68	15.5	2.7	.61	13.6	4.3	.96
6	26.2	5.6	1.25	18.0	3.4	.72	15.8	1.7	.38	15.2	4.6	1.02
7	25.4	5.7	1.27	18.4	2.9	.65	16.1	1.8	.41	17.0	5.4	1.20
8	28.0	4.7	1.48	19.6	3.0	.93	16.3	1.5	.47	20.5	8.3	1.51
Total	20.4	7.1	.58	16.1	4.1	.33	14.07	2.7	.22	12.2	6.0	.49

Thinking standard deviation increased .57 while the standard error of measurement remained the same; and the Numerical Reasoning standard deviation increased .1 and the standard error of measurement .01.

Summary of Step Two Results

In Step Two results regarding SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning internal consistency, category specificity, stability reliability and accuracy of measurement were presented. An inspection of the distribution of Visual-Motor and Social Intelligence-Reasoning items reveals that no further meaningful investigation of these categories is possible at this time. The results presented for the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories reject null hypotheses four, five and six. For the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories Kuder-Richardson #20 coefficients ranged between .13 to .82 with corrected coefficients ranging between .37 to .86. Only for first graders did the four categories appear to lack sufficient internal consistency. At least adequate specificity is suggested for all four categories for all age levels one through eight, except for Conceptual Thinking grade levels four through eight and Numerical Reasoning for seventh graders. Significant stability coefficients ($p \leq .05$) for the four categories ranged between .51 to .97, with coefficients corrected for restricted intelligence range between .58 to .98. In all instances, accuracy of measurement is suggested by reasonably small standard error of measurements.

Step Three: Validity of the Proposed SIT Classification Schema

In Step Three, correlations between SITFILE categories Language, Memory, Conceptual Thinking and Numerical Reasoning and two diagnostic batteries were calculated. The results of this investigation of the SITFILE's concurrent validity are presented in Tables 16 and 17. In interpreting these correlations, consideration should be given to the restrictions of range of the two samples. Tables 2 and 3 indicate that Samples Two and Three are restricted groups with narrow IQ standard deviations. Cronbach (1970) points out that correlations will be smaller in a select group than in one containing a wide range of abilities. Further, he states that it is unusual for validity coefficients to rise above .60.

Table 16 presents the coefficients of correlation and levels of significance between SITFILE category scores and ITPA and WRAT age scores for first, second and third graders, under nine years of age. Since, each group is rather small, it is best to study the rows labeled "total". The "total" coefficients of correlation of the Language Category with the ITPA subtests Auditory Reception, Auditory Association, Auditory Sequential Memory, and Verbal Expression range between .05 to .55. Coefficients of the Language Category with the WRAT subtest Arithmetic, Reading and Spelling range between .29 to .36. Significant coefficients ($p \leq .05$) are noted between Language and ITPA Auditory Reception, Auditory Association, Grammatic Closure and WRAT Arithmetic, Reading and Spelling. Significant "total" coefficients of correlation ranging between .34 to .60 are indicated between the Memory Category and all subtests included in Diagnostic Battery A. The "total" coefficients of correlation of the Conceptual Thinking Category with the ITPA and WRAT subtests, ranging between .15 to .62 are all significant ($p \leq .05$) except that between Conceptual Thinking and Sequential Memory. Significant correlations ($p \leq .05$) are noted between Numerical Reasoning and ITPA Auditory Reception, Auditory Association, Grammatic Closure and WRAT Arithmetic, Reading and Spelling. Numerical Reasoning and Sequential Memory and Verbal Expression subtests do not correlate significantly.

Ample correlations that appear capable of adding information that can aid further diagnostic test selections and hypotheses formulations are those above .50. Such correlations are noted between the Language Category and Auditory Reception, Auditory Association, and Grammatic Closure; the Memory Category and Grammatic Closure and Arithmetic; the Conceptual Thinking Category and Grammatic Closure, Arithmetic, Reading and Spelling; and the Numerical Reason-

Table 16

Coefficients of Correlation of SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery A Age Scores for Sample Two

r Probability			ITPA					WRAT		
Category	Grade	n	Aud. Recept.	Aud. Ass.	Aud. Seq.Mem.	Verbal Express.	Grammatic Closure	Arith-metic	Read.	Spell.
Language	1	11	.70 .02	.58 .06	.02 .23	-.34 .31	.64 .03	.14 .67	.15 .65	.26 .44
	2	18	.32 .20	.63 .01	-.56 .01	.08 .75	.40 .10	-.05 .84	-.18 .48	-.25 .31
	3	16	.53 .04	.31 .24	.53 .04	.08 .77	.52 .04	.45 .08	.61 .01	.33 .22
	Total	45	.51 .01	.51 .01	.05 .75	.11 .47	.55 .01	.38 .01	.36 .01	.29 .05
Memory	1	11	.79 .01	.81 .01	.49 .13	.16 .63	.64 .03	.52 .10	.38 .25	.51 .11
	2	18	.23 .36	.05 .84	.46 .05	.22 .37	.24 .32	-.07 .76	.00 .99	.16 .52
	3	16	.54 .03	.33 .21	.54 .03	.21 .43	.50 .05	.63 .01	.63 .01	.29 .28
	Total	45	.45 .01	.44 .01	.46 .01	.34 .02	.52 .01	.60 .01	.47 .01	.48 .01

Table 16

Coefficients of Correlation of SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery A Age Scores for Sample Two

r Probability			ITPA					WRAT		
Category-Grade		n	Aud. Recept.	Aud. Ass.	Aud. Seq.Mem.	Verbal Express.	Gramatic Closure	Arith-metic	Read.	Spell.
Conceptual Thinking	1	11	.49	.54	.17	.40	.64	.44	.54	.73
			.12	.09	.61	.22	.03	.18	.09	.01
	2	18	.55	.19	-.22	.51	.47	.39	.22	.11
			.02	.46	.36	.03	.05	.11	.39	.66
	3	16	.32	.33	.52	-.14	.31	.23	.39	-.07
			.22	.22	.04	.61	.24	.38	.13	.80
	Total	45	.44	.43	.15	.38	.54	.62	.55	.55
			.01	.01	.31	.01	.01	.01	.01	.01
Numerical Reasoning	1	11	.30	.32	.64	-.05	.31	.64	.37	.32
			.38	.33	.03	.89	.36	.03	.37	.34
	2	18	-.18	.13	-.10	-.33	.06	-.24	-.51	-.03
			.48	.60	.70	.18	.82	.33	.03	.90
	3	16	.46	.38	.43	-.17	.42	.43	.54	.17
			.07	.15	.10	.54	.10	.10	.03	.53
	Total	45	.29	.38	.26	.02	.31	.52	.40	.34
			.05	.01	.08	.90	.01	.01	.01	.02

ing Category and Arithmetic. The statistically significant correlations for children in grades one through three reported in Table 16 provides evidence leading to the rejection of null hypothesis seven: There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery A including administration of subtests of the Illinois Tests of Psycholinguistic Abilitites and the Wide Range Achievement Test.

Table 17 presents the coefficients of correlation and levels of significance of SITFILE category scores and Detroit and WRAT age scores for the sample three children. Again, while correlations were computed by grade levels, since the number of children at each grade is rather small, it is best to study the rows labeled "total". The coefficients of correlation of the combined or total sample for the Language Category with Detroit and WRAT subtests range from .21 to .67. Significant correlations ($p \leq .01$) are noted between the Language category and all investigated subtests but the WRAT Reading subtest. Significant correlations are indicated between the Memory Category and Verbal Absurdities .31 ($p \leq .03$); Verbal Opposites .52 ($p \leq .01$); Auditory Attention .51 ($p \leq .01$); Arithmetic .53 ($p \leq .01$); Reading .50 ($p \leq .01$) and Spelling .51 ($p \leq .01$). The "total" coefficients of correlation of the Conceptual Thinking Category with the Detroit and WRAT subtests, all of which are significant ($p \leq .02$), range between .33 to .57. Significant ($p \leq .01$) total sample correlations are also noted between the Numerical Reasoning Category and the investigated Detroit and WRAT subtests: Verbal Absurditites .30; Verbal Opposites .53; Auditory Attention .40; Arithmetic .56; Reading .33; and Spelling .43.

Considering only resultant correlations above .50 which would identify those capable of reasonably furthering diagnostic hypotheses formulation, ample correlations are noted between the Language Category and Verbal Opposites, and the WRAT Arithmetic subtest. Correlations above .50 are also indicated between the Memory Category and Verbal Opposites, Auditory Attention, Arithmetic and Reading; the Conceptual Thinking Category and Verbal Opposites and Arithmetic; and the Numerical Reasoning Category and Verbal Opposites and Arithmetic. The statistically significant correlations found between the Sample Three test scores as reported in Table 17 reject null hypothesis eight (There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery B including administration of subtests of the Detroit Tests of Learning Aptitude and the WRAT).

Table 17

Coefficients of Correlation of Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery B Age Scores for Sample Three

r Probability			Detroit			WRAT		
Category-Grade		n	Verbal Absurdities	Verbal Opposites	Auditory Attention	Arithmetic	Reading	Spelling
	3	4	.94 .05	.93 .07	.65 .35	.86 .13	-.61 .39	.40 .60
	4	10	.43 .22	.52 .89	-.06 .88	.45 .20	-.13 .73	-.15 .67
Language	5	7	.76 .05	.66 .11	.58 .17	-.71 .07	-.94 .01	-.55 .20
	6	10	.05 .89	.65 .04	.33 .34	.77 .01	.56 .09	.56 .09
	7	9	.60 .09	.77 .02	.77 .02	.64 .07	.14 .72	.26 .50
	8	10	.44 .20	.91 .01	.68 .03	.32 .37	-.20 .58	-.24 .51
	Total	50	.42 .01	.67 .01	.47 .01	.51 .01	.21 .15	.36 .01
	3	4	.26 .73	.40 .60	.64 .36	.56 .44	.80 .20	-.75 .25

Table 17

Coefficients of Correlation of Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery B Age Scores for Sample Three

r Probability			Detroit			WRAT		
Category	Grade	n	Verbal Absurdities	Verbal Opposites	Auditory Attention	Arithmetic	Reading	Spelling
Memory	4	10	.12 .74	.46 .18	.43 .21	.39 .27	.46 .18	.51 .14
	5	7	.37 .41	.43 .34	.64 .12	.45 .32	.21 .96	.48 .28
	6	10	.01 .99	.32 .37	.50 .15	.68 .03	.58 .08	.52 .12
	7	9	.79 .01	.28 .47	.52 .15	.75 .02	.47 .21	.42 .26
	8	10	.44 .20	.33 .36	.56 .10	.39 .26	.21 .38	-.14 .71
	Total	50	.31 .03	.52 .01	.51 .01	.53 .01	.50 .01	.51 .01
	3	4	.91 .09	.93 .07	.60 .40	.85 .15	-.60 .40	.45 .55
	4	10	-.20 .59	.26 .46	.36 .31	.22 .54	.29 .42	.55 .10

Table 17

Coefficients of Correlation of Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery B Age Scores for Sample Three

r Probability			Detroit			WRAT		
Category-Grade		n	Verbal Absurdities	Verbal Opposites	Auditory Attention	Arithmetic	Reading	Spelling
	5	7	-.27 .56	-.10 .82	.42 .34	-.08 .87	-.33 .46	-.26 .58
Conceptual Thinking	6	10	-.44 .21	.30 .39	.13 .72	.77 .01	.76 .01	.65 .04
	7	9	.89 .01	.57 .10	.41 .27	.70 .04	.03 .94	.10 .80
	8	10	.76 .01	.74 .01	.51 .13	.64 .05	.32 .36	-.09 .80
	Total	50	.33 .02	.57 .01	.41 .01	.56 .01	.39 .01	.44 .01
	3	4	.15 .85	-.11 .88	-.20 .80	-.23 .77	.61 .39	.62 .38
	4	10	.45 .19	.32 .37	.23 .52	.40 .25	.35 .32	.30 .40
Numerical Reasoning	5	7	-.04 .95	-.10 .84	.01 .99	.81 .03	.57 .18	.81 .03
	6	10	-.01 .98	.43 .21	.41 .22	.70 .02	.62 .06	.59 .07

Table 17

Coefficients of Correlation of Language, Memory, Conceptual Thinking and Numerical Reasoning Category Scores with Diagnostic Battery B Age Scores for Sample Three

r Probability		Detroit			WRAT		
Category-Grade	n	Verbal Absurditites	Verbal Opposites	Auditory Attention	Arithmetic	Reading	Spelling
7	9	.54	.80	.67	.84	-.12	-.06
		.13	.01	.05	.01	.77	.89
8	10	.57	.84	.64	.30	.01	-.21
		.08	.01	.05	.40	.96	.57
Total	50	.30	.53	.40	.56	.33	.43
		.01	.01	.01	.01	.02	.01

Means, standard deviations and standard error of measurements of Language, Memory, Conceptual Thinking and Numerical Reasoning Categories from three hundred eighty (380) Step Three SITFILE category scores are presented in Tables 18 and 19. Similar to the pattern noted from the mean SITFILE category scores of Sample One, the mean SITFILE category scores of Sample Two and Sample Three also suggest a systematic increase of mean score paralleling that of grade level placement. For Sample Two, Language mean scores range from 8.91 to 11.65; for Memory from 6.0 to 11.6; for Conceptual Thinking from 6.91 to 11.31; and for Numerical Reasoning from 4.91 to 8.38. Language Category standard deviations range between 1.99 to 3.26. The Memory Category's standard deviations range from 2.75 to 4.19. Standard deviations for the Conceptual Thinking Category are: first graders 2.81, second graders 1.97 and third graders 2.36; while, the Numerical Reasoning Category's standard deviations cluster close to 1.5.

For Sample Three, Language mean scores range from 11.25 to 21.1; for Memory from 10.5 to 15.7; for Conceptual Thinking from 9.25 to 13.8; and for Numerical Reasoning from 6.75 to 13.2. Sample Three Language Category standard deviations range between 2.5 to 7.41. The Memory Category's standard deviations range from 1.41 to 4.69. Standard deviations for the Conceptual Thinking Category range between 1.20 to 4.11 while the Numerical Reasoning Category's standard deviations range from 1.25 to 8.99. All Sample Two and Sample Three obtained standard error of measurements are reasonably small, ranging from .30 to 2.84.

Summary of Step Three Results

In Step Three, results regarding the validity of the Language, Memory, Conceptual Thinking and Numerical Reasoning SITFILE categories are presented. Additional information regarding the SITFILE category scores consistency of performance is also presented. The reported coefficients between the four categories and the specified educational diagnostic tests suggest that the four SITFILE categories do correlate with other tests purported to measure specific functions relevant to educational diagnosis. Correlations ranging from .30 ($p \leq .03$) to .67 ($p \leq .01$) suggest rejection of null hypotheses seven and eight for the Language, Memory, and Conceptual Thinking and Numerical Reasoning Categories.

Table 18

Sample Two Means, Standard Deviations (Std.D.) and Standard Error of Measurements (SE_m) for SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Categories by Grade

	Language			Memory			Conceptual Thinking			Numerical Reasoning		
Grade	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m
1	8.91	2.95	.89	6.00	3.29	.99	6.91	2.81	.85	4.91	1.58	.48
2	9.78	1.99	.47	9.94	2.75	.65	9.89	1.97	.46	6.11	1.28	.30
3	11.63	3.26	.82	11.06	4.19	1.05	11.31	2.36	.59	8.38	1.58	.87
Total	10.22	2.86	.42	9.39	3.88	.57	9.70	2.82	.42	6.61	2.67	.39

Table 19

Sample Three Means, Standard Deviations (Std.D.) and Standard Error of Measurements (SE_m) for SITFILE Language, Memory, Conceptual Thinking and Numerical Reasoning Categories by Grade

Grade	Language			Memory			Conceptual Thinking			Numerical Reasoning		
	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m	Mean	Std.D.	SE_m
3	11.25	2.50	1.25	10.50	3.00	1.50	9.25	4.11	2.06	6.75	1.71	.85
4	14.70	4.72	1.49	11.20	4.69	1.48	10.70	2.00	.63	8.00	2.54	.80
5	17.57	3.21	1.21	14.00	1.41	.54	14.00	1.83	.69	8.71	1.25	.47
6	17.40	7.41	2.34	12.90	4.25	1.35	12.40	2.59	.82	12.50	8.99	2.84
7	17.44	5.34	1.78	14.33	3.08	.10	13.22	1.20	.40	12.22	5.33	1.78
8	21.10	4.63	1.46	15.70	2.31	.73	13.80	2.86	.90	13.20	5.67	1.79
Total	17.14	5.63	.80	13.34	3.70	.52	12.46	2.73	.39	10.70	5.71	.81

Recapitulation of Step One Through Step Three Results

In Step One through Three, eight null hypotheses were investigated. Null hypothesis one (There is no association between the SIT classifications Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor and Social Intelligence-Reasoning at the two to five, six to ten, eleven to fourteen and fifteen to adult age levels.) was rejected. A Kendall coefficient indicates moderate concordance among the SIT category ranks at the four age levels. Null hypothesis two (There is significant ($p \leq .05$) inter-class agreement between the distribution of functions assessed by the SIT at the two to five, six to ten, eleven to fourteen and fifteen to adult age levels.) was rejected except between age intervals eleven to fourteen and fifteen to adult. Kendall Tau coefficients indicate that the distribution of category items for age levels two to five and six to ten, and six to ten and eleven to fourteen are not significantly related; but, they are related for the eleven to fourteen and fifteen to adult age levels. Null hypothesis three (There is no significant ($p \leq .05$) interclass agreement between the distribution of functions for the SIT and S-B items included within the: two to five year level; six to ten year level; eleven to fourteen year level; fifteen to adult year level; or two to adult year level.) was not rejected. Proposed SIT and S-B categories and their item distributions tend not to be related as indicated by low tau coefficients. Limited item distributions and the basal ceiling SIT grading criteria limit the investigation and utility of the SITFILE Visual-Motor and Social Intelligence-Reasoning Categories.

Null hypothesis four (There is no significant ($r \geq .70$) relationship between the test items included within the SIT categories.) was rejected with the exception of the Conceptual Thinking Category for children in grades three, four, five and seven, and the Visual-Motor and Social Intelligence-Reasoning Categories for all children. Other SITFILE categories for children in grades two through eight appear to have sufficient internal consistency. Null hypothesis five (There is no significant difference between an individual SIT category's total reliable variance and its squared multiple correlation with the rest of the SIT categories.) was rejected for the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories. At least ample specificity is reported for these four categories for first, second and third graders and for the Language, Memory and Numerical Reasoning Categories for fifth and sixth graders. Null hypothesis six (There is no significant ($p \leq .05$) relationship between test-retest category scores.) was also rejected. All reported corrected stability coefficients, except that of Numerical Reasoning for first graders, exceeded .70. Language, Memory, Conceptual Thinking and Numerical Reasoning Category reliability is also suggested by small standard error of measurements.

Null hypothesis seven (There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery A including administration of subtests of the Illinois Tests of Psycholinguistic Abilities and the Wide Range Achievement Test.) was rejected for the Language, Memory, Conceptual Thinking, and Numerical Reasoning Categories. Significant correlations ($p \leq .05$) are reported between these four categories and the ITPA and WRAT subtests for six through eight year olds. Null hypothesis eight (There is no significant ($p \leq .05$) relationship between SIT category scores and test scores obtained on Diagnostic Battery B including administration of subtests of the Detroit Tests of Learning Aptitude and the Wide Range Achievement Test.) was rejected for nine through fourteen year olds. Significant correlations ($p \leq .05$) are reported between all four categories and the Detroit and WRAT subtests for nine through fourteen year olds.

DISCUSSION

SIT Classification Schema

Sattler (1976) states that a classification system is a convenient way of describing an individual's test performance. In Step One a SIT classification system, based upon the Sattler S-B classification, was employed by judges to categorize SIT items from year 2-0 up. Factor analysis was not performed since it may too narrowly constrict qualitative behavioral analysis. Further, the resultant schema was intended to be only an aid to hypothesis formulation. A comparison of the item classifications of three independent judges resulted in 100% agreement of item classification for 72.4% of the SIT items evaluated. Agreement by two out of three judges resulted in the placement of another 23% of the included SIT items. There was complete agreement by three judges on items included in the Visual-Motor Category and 93.9% agreement on items included in the Numerical Reasoning Category, with 91.8% agreement in the Language Category. Categories with lesser obtained placement agreement were the Memory and Social Intelligence-Reasoning Categories.

It was noted that there were no items in the Visual-Motor Category beyond year 7-4. Consequently, while the Visual-Motor Category has 100% agreement of item assignment, it lacks representativeness across age levels. In the Social Intelligence-Reasoning Category there are no items beyond year 9-2, with only two items between year 5-10 and 9-2. However, the high percentage of agreement on item assignments to the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories and their broader throughout test item distributions suggest that a classification system can be developed for the SIT with at least moderate face validity. The proposed SIT classification schema appears to provide a limited format whose reliability and validity can be investigated.

Developmental Analysis of Intelligence and SITFILE Construct Validity

As part of the further investigation of the proposed SIT classification schema, the proportion of SIT category items within four age groups (two to five, six to ten, eleven to fourteen, and fifteen to adult) was calculated. Since the SIT is reported as a general measure of intelligence, an evaluation of its category-age level intra-class relationship should reveal a significant similarity. The moderate concordance (w=.52) among the rankings of the categories at the four age levels suggests that the test tends to measure similar functions throughout the test. The association within the Sattler S-B classification schema, with the Social Intelligence and Reasoning Categories combined, at the same four age levels was also calculated. The resultant coefficient of concordance (w=.65) suggests that the S-B also tends to measure similar functions throughout the test. This is consistent with Slosson's statement that he designed his test in a manner similar to the S-B and suggests construct validity. However, both the SIT and S-B coefficients of concordance suggest only moderate intra-test similarities. If inter-class agreement is not found between age levels then systematic differences in relative weights at the four age levels can be hypothesized as reflecting developmental designs.

Kendall Tau coefficients of agreement between both SIT and S-B age levels are presented in Table 8. They indicate significant agreement for the S-B only between age levels six to ten and eleven to fourteen, and for the SIT only between age levels eleven to fourteen and fifteen to adult. Both the S-B and SIT place different weights on the functions they test between all but two levels. This is consistent with Sattler's developmental analysis of the S-B (Sattler, 1965) and argues for a developmental analysis of the SIT.

The extent to which a SIT developmental analysis, similar to that presented by Sattler for the S-B, can be articulated is dependent upon demonstration of significant inter-class agreement between the S-B and SIT classification schemes. As presented in the previous section, null hypothesis three (There is no significant ($p \leq .05$) inter-class between the distribution of functions for the SIT and S-B items included within the: two to five year level; six to ten year level; eleven to fourteen year level; fifteen to adult year level; or two to adult level.) was not rejected. No significant Kendall Tau coefficients of agreement were identified between the total tests or between the S-B and SIT at any age levels. Therefore, while the SIT may be patterned after the S-B, the tests do not appear to place equal emphasis on the measurement of similar functions at the same age levels. The S-B and SIT do not test the same things to the same extent at any age level.

Sattler, in his developmental analysis of the S-B, focused upon changing behavioral demands. Failure to identify distributive similarities between the S-B and SIT argues against a similar SIT developmental focus. However, an analysis of the SIT function weights at the four age levels does suggest a different underlying structural design.

Reference to Table 6 suggests that in the early age level two to five, the SIT primarily measures Social Intelligence-Reasoning items. This seems to reflect a belief that a young child's early cognitive development is based upon environmental awareness and social interactions. This one category contains 37% of the SIT items in the two to five age level with another 7% of the items assigned to the Visual-Motor Category and the rest of the items evenly distributed, 14% each, between the remaining categories. This early emphasis on social maturity and discrimination ability is consistent with the Piagetian cognitive developmental hypothesis that the foundation of mental activity can be traced to recognition of one's potential as an active doer rather than as a passive recipient of the wisdom of others (Schwebel and Raph, 1973). Further, it is consistent with the cognitive developmental view that as a result of successful interactions with people and objects, links will be formed, facilitating assimilation and accommodation and the movement from one stage of mental development to a higher one (ie. increased intelligence).

Between six and ten, an age grouping that corresponds roughly to Piaget's Concrete Operational Stage, SIT item distributions, to be consistent with cognitive developmental theory, would be expected to shift from an emphasis on interactions to an emphasis on internalized actions. Accordingly, an increase in the SIT distribution of Memory items to 30% and a decrease in the SIT distribution of Social Intelligence-Reasoning items to 10% and Visual-Motor items to 3% is noted. This distributive shift away from Social Intelligence-Reasoning items to Memory items is consistent with the expectation that children at the Concrete Operational Stage need a grasp of temporal perspectives, as a prerequisite to discovering relations of reciprocity and annulment. Concomitantly, at the Concrete Operational Stage a child is expected to rely more heavily upon symbolization. An increase in the distribution of SIT Conceptual Thinking items to 20% is consistent with movement away from physical representations to symbolic ones. Also, during the Concrete Operational Period a child's ability to control logical quantifiers such as one, some and all begin to emerge, providing a broader base for numerical operations. Thirteen percent of the SIT items between six and ten involve numerical reasoning.

As noted in Table 8 a significant positive relationship exists between the SIT item distributions in the eleven to fourteen and fifteen to adult age levels. Consequently, the SIT items included in both these age groupings appear to be testing significantly similar functions with equal weight. These adolescent age levels correspond to Piaget's Formal Operational Period. During the Formal Operational Period previous concrete operations are expected to be extended making possible greater application of mathematical laws. An increase in the proportion of numerical reasoning items to 38% at the eleven to fourteen age level with 28% at the fifteen to adult age level suggests a concomitant increase in the SIT emphasis on this ability. Piaget (1970) notes that cooperation as an objectively conducted discussion also emerges during this adolescent period. The extent to which discussion can give rise to internalized conversations appears to derive from an individual's ability to symbolize information, retain that information and deal with a wide variety of complex relations. SIT Language items during the period between eleven to fourteen increase to 29% and from fifteen to adult to 57%, centering around quality of vocabulary and comprehension of verbal relations. Included proportionately high Memory items, eleven to fourteen - 21% and fifteen to adult - 10%, are both of an ideational and attention span nature. Conceptual Thinking items account for 13% of the test items between eleven to fourteen and 6% of the test items at the fifteen to adult level, assessing an individual's ability to employ a categorical attitude. The absence of Social Intelligence-Reasoning and Visual-Motor items at both the eleven to fourteen and fifteen to adult levels is consistent with the Piagetian hypothesis that at the Formal Operational Stage thought becomes hypothetical in nature.

In sum, it appears that while a developmental analysis based upon changing behavioral demands, similar to that postulated for the S-B, can not be articulated for the SIT, a cognitive developmental basis can be hypothesized. The SIT function distributions are significantly different from one another as well as from those of the S-B, at three different age levels. Between the third and fourth age levels similarities are suggested justifying the combined discussion of items within the eleven to fourteen and fifteen to adult age levels. The SIT from year two to adult can be viewed as composed of three developmental levels. The distribution of functions within these levels can be related to requisite skills at each of the three later Piagetian Cognitive Developmental Stages.

Reliability

In Step Two a system for quantifying SIT category responses was presented and SITFILE reliability as self-consistency and as accuracy of measurement was investigated for the Language, Memory, Numerical Reasoning and Conceptual Thinking Categories. Narrow Visual-Motor and Social Intelligence-Reasoning item distributions and a basal-ceiling test design make the Visual-Motor and Social Intelligence-Reasoning Categories useless to educational hypotheses formulation. Therefore, their reliability was not evaluated and no description of an individual's intra-test variability should be derived from an analysis of responses to items within these categories. For the other four categories, two aspects of self consistency were considered: internal consistency and subtest specificity. Two aspects of accuracy of measurement were also considered: stability and consistency of performance.

Low previously presented Kuder-Richardson #20 coefficients (see Table 10) for the Conceptual Thinking Category for children beyond second grade, and for the Language, Memory and Numerical Reasoning Categories for first graders, indicate inconsistency in item performance and variable error. Therefore, no description of an individual's intra-test variability should be derived from an analysis of responses to items within the Conceptual Thinking Category, nor should interpretation be made of any category score of a first grader. However, for children in second through eighth grade, the Language, Memory and Numerical Reasoning Categories do appear to have sufficient internal consistency of results throughout the test to rule out sampling of content as a major source of measurement error. It should be noted that even the reported corrected coefficients may still be suppressed since other criteria such as achievement, sensory and motor integrity and socio-economic status also have restricted ranges within the experimental sample.

The second measure of self consistency involved investigating whether a category's variance was both reliable and unique to that particular category. SITFILE specificities for the Language, Memory and Numerical Reasoning Categories (see Table 11) are optimistic in encouraging specific function interpretation for children in grades two through eight. For these three categories, inadequate specificity was only found for the Memory Category for eighth graders and for the Numerical Reasoning Category for fourth and seventh graders. The Conceptual Thinking Category possesses adequate specificty only for first, second and third graders. However, since it has been demonstrated generally to lack internal consistency, it is not appropriate to think of the Conceptual Thinking Category as being reliable even in this limited sense for these few grade levels.

To evaluate the Language, Memory, Numerical Reasoning and Conceptual thinking Categories, accurary of measurement individual category scores were specified statstically, so as to reduce their ambiguity. The quantification system employed involved combined area and range intra-category scatter analysis. Thus, the advantages of both consideration of consistency of performance and number of levels over which success or failure occurred were incorporated within the scoring system. A subject's chronological age was chosen as the reference point from which to measure scatter. Chronological age was chosen because it met both criteria of stability and psychological relevance. An individual's chronological age is not dependent upon his IQ, but at the same time IQ and academic expectations are not independent of chronological age. Choosing chronological age as a reference point also avoids mean-score reference point complications. Further, by employing standard deviation units the age-scale problem of distance between chronological age and year levels suggesting relative differences at different levels is also minimized.

Stability reflects the extent to which similar scores are achieved from testing to testing. Since SITFILE category scores are to be used as an aid to educational hypothesis formulation and facilitation of remedial planning, it is necessary to determine that changes in test scores are not due to measurement error alone. Obtained stability coefficients for the Language, Memory, Conceptual Thinking and Numerical Reasoning Categories are presented in Table 12. The obtained coefficients were corrected for restricted range of intelligence. These corrected coefficients are presented in Table 13. Only two corrected coefficients for children in grades two through eight for the Language, Memory and Numerical Reasoning Categories fall below the .80 point considered adequate for basic research. While, eleven out of the twenty-one presented Language, Memory and Numerical Reasoning coefficients for the children in grades two through eight were above the .90 level considered requisite in applied settings. These coefficients give evidence of the potential reliability of SITFILE Language, Memory and Numerical Reasoning Category scores for children in grades two through eight. They suggest that the precision of the SITFILE for the identified categories is relatively high, that daily fluctuations in the examinee or test environment do not significantly affect category scores for a period up to two and one-half months and that the functions measured are reasonably stable over time. However, a comparison of mean SITFILE scores on the first and second testings (see Tables 14 and 15) reveals gains ranging between .1 and 1.8. This practice effect should be considered when retesting a child within a relatively short time interval.

The second aspect of accuracy of measurement considered is consistency of performance. Consistency of performance is reflected as an estimate of the standard deviation of a set of obtained scores from its "true" score. Expressed as standard error of measurement, it is dependent upon the standard deviation of the distribution of the obtained scores and the reliability coefficients of the test. SITFILE standard error of measurements are presented in Tables 14 and 15. Since restriction of range within the samples suppressed reliability coefficients, it is also believed that this restriction of range has also affected the standard error of measurement estimates. Still, in a limited sense, these standard error of measurement estimates can be used with a known degree of certainity to establish zones within which "true" scores lie. Reasonable accuracy of measurement is suggested since, for all but eighth graders, standard error of measurement estimates were less than one-fourth as large as the the standard deviation of the category scores.

Validity

In Step Three the concurrent validity of the proposed SIT classification and SITFILE was evaluated by correlating SITFILE scores with scores obtained on three other individually administered diagnostic tests: the Illinois Tests of Psycholinguistic Abilities, the Detroit Tests of Learning Aptitude and the Wide Range Achievement Test. The resultant correlation coefficients between Language, Memory, Conceptual

Thinking and Numerical Reasoning Category scores and ITPA and WRAT age scores for six through eight year olds are presented in Table 16. Correlations between Language, Memory, Conceptual Thinking and Numerical Reasoning Category scores and Detroit and WRAT age scores for children nine through fourteen years of age are presented in Table 17. Significant coefficients ($p \leq .05$) establish a relationship between the functions assessed by the Language, Memory, Conceptual Thining and Numerical Reasoning Categories and the aptitudes measured by these diagnostic tests.

The six through eight year old total sample correlations above .50 suggest diagnostically useful correspondences between the Language Category and ITPA Auditory Reception, Auditory Association and Grammatic Closure subtest scores. These correlations suggest that Language Category performances can provide information pertinent to the evaluation of a child's internalization of semantics and his capacity to relate meaningfully auditorily received stimuli. Correlations greater than .50 between the Memory Category and the ITPA Grammatic Closure subtest and the WRAT Arithmetic subtest, for the total six through eight year old sample, suggest that the SITFILE measure of attention span, rote and ideational memory may also provide information related to a child's ability to make use of the redundancies of language as well as to his ability to perform arithmetic operations. While significant Conceptual Thinking validity coefficients are presented in Table 16, no interpretation of the diagnostic usefulness of this category is recommended as this category has been found to be unreliable. Significant correlations, greater than .50 are noted between total Sample Two's SITFILE Arithmetic Category scores and WRAT Arithmetic scores. Consequently, a child's performance on the SIT Arithmetic Category may provide insight into that child's probable performance on other computation tasks.

For children nine to fourteen years of age, Sample Three correlations between the SITFILE categories and Detroit and WRAT subtests also suggest diagnostic usefulness. For these older children, information regarding performances on the Language Category may provide insight into one's internalization of semantics and to one's arithmetic problem solving ability (Detroit Verbal Opposites and WRAT Arithmetic subtests). Interpretation of a nine through fourteen year old's performance on the Memory Category may further understanding of that individual's auditory reception, arithmetic proficiency, reading attack skills and spelling recall, as well as short term recall (Detroit Verbal Opposites, Auditory Attention Span and WRAT Arithmetic, Reading and Spelling). Conceptual Thinking coefficients, for these older children, again should not be interpreted due to the category's unreliability. Significant correlations, greater than .50, between the Numerical Reasoning Category and the Detroit Verbal Opposites subtest and the WRAT Arithmetic subtest suggest a relationship between arithmetic reasoning ability and verbal ability, as well as between SITFILE computation skills and WRAT computation skills.

SITFILE Diagnostic Value

The development of a SIT classification schema as described in the preceding Method and Results Sections was conducted and is graphically presented as the SITFILE. Both null hypotheses one and two were rejected. The resultant schema, based upon a juried placement decision, is similar in design to the S-B Binetgram. Both schemes identify similar functions with proportionately different emphases at different age levels. This enables one to maintain a global view of intelligence, while at the same time focusing on patterns of test performance.

However, the relative emphasis of the different functions within the S-B and SIT are not significantly similar. Null hypothesis three was not rejected. Consequently, while behavior demands can be articulated in conjunction with changes in proportional function assessments at different age levels by the S-B, attention to maturation of developmental schema can be postulated in conjunction with changes in proportional function assessments at different age levels by the SIT.

These proportional differences are consistent with a developmental theory of intelligence, but they are not consistent with an assortative theory of diagnostic assessment. The cognitive developmental theory of intelligence has a hierarchical basis, while assortative testing theory implies that thorough measurement of each function is necessary for the full understanding of cognitive organization. Consonance with developmental theory prevents the inclusion within the SIT of visual-motor items and social intelligence items, respectively beyond the 7-4 and 9-2 age levels. It should also be recalled that modifications of the Sattler category definitions were necessary prior to their application to the SIT. At no age level within the SITFILE are visual memory or non-verbal reasoning items included. Consequently, the diagnostic information to be gained from an interpretation of SITFILE category scores is first limited by the scope and design of the SIT itself.

The SITFILE Visual-Motor and Social Intelligence-Reasoning Categories were found to lack sufficient representativeness, prohibiting any evaluation of their reliability. The Conceptual Thinking Category was found to lack internal consistency (see Table 10), and therefore, no further investigation or discussion of its reliability or validity is warranted. Consequently, item responses within these three categories, Visual-Motor, Social Intelligence-Reasoning and Conceptual Thinking, should not be interpreted for diagnostic purposes.

However, null hypotheses four, five and six were rejected for the Language, Memory and Numerical Reasoning Categories for second through eighth graders. There does appear to be a significant relationship between the test items included within these categories for children at these grade levels (negation of null hypothesis four). This indicates that the Language, Memory and Numerical Reasoning Categories possess adequate internal consistency to justify their incorporation in the diagnostic evaluation of second through eighth graders, if other measures of the categories' reliability and validity are equally sufficient.

Significant differences were found between each of these three categories' total reliable variance and their squared multiple correlation with the other SITFILE categories, except for the Memory Category for eighth graders and for the Numerical Reasoning Category for fourth and seventh graders (negation of null hypothesis five). Thus, except for the three previously noted exclusions, there appears to be some empirical sanction for limited unique interpretation of the individual Language, Memory and Numerical Reasoning Categories for second through eighth graders. However, caution must be excercised in interpreting these categories in relative isolation as the proportion of common variance (1.00 minus the sum of the specific and error variance) exceeds the proportion of specific variance for the three categories at most grade levels. The fact that the categories have more common variance than specific variance strongly argues against their independent use for daignosis and necessitates consideration of possible interactions if these category scores are to be considered as providing implications for further testing or hypothesis formulation.

There also appears to be a significant relationship between the test-retest Language, Memory and Numerical Reasoning Category scores (negation of null hypothesis six). Corrected stability coefficients in the .80's and .90's (see Table 13) provide further evidence of the reliability of the Language, Memory and Numerical Reasoning Categories with children in grades two through eight. Low standard errors of measurement are also presented (see Tables 14 and 15) suggesting accuracy of measurement.

The rejection of null hypotheses four, five and six for the Language, Memory and Numerical Reasoning Categories for second through eighth graders suggests that consistent with the limitations previously noted, the scores from these categories, consisting of homogeneous items, measuring relatively independent functions, are reasonably accurate and stable over time. Therefore, they seem to possess adequate reliability justifying their diagnostic employment. However, reliability is only one criterion used when judging the adequacy of a test. Test administrators are also concerned with a test's predictive ability, particularly when test results are used to form hypotheses regarding behaviors external to the measuring instrument itself.

Significant correspondences between SITFILE categories and ITPA, Detroit and WRAT subtests indicate a rejection of null hypotheses seven and eight for the Language, Memory and Numerical Reasoning Categories. Performances on these categories do correlate significantly with performances on other tests purported to measure functions relevant to educational diagnosis (negation of null hypotheses seven and eight). Consequently, the Language, Memory and Numerical Reasoning Categories do possess predictive validity (see Tables 16 and 17) for children in grades two through eight. Interpretation of Language, Memory and Numerical Reasoning Category performances can provide information, respectively, regarding a second through eighth grade student's internalization of semantics and understanding of verbal relationships, his ability to store and retrieve auditorily received stimuli, and his arithmetic computation. For children beyond second grade, performances on the Memory Category may also provide information regarding their reading attack and spelling proficiency.

However, no diagnosis or diagnostic hypothesis should be based upon a category score interpreted in isolation. The fact that multiple correlations were found between each of the Language, Memory and Numerical Reasoning Categories and the different diagnostic tests support the previously stated hypothesis of both function and category interactions. An awareness of these multiple correlations strongly cautions against employing individual categories in an independent fashion. Also, while interpretation of these three category scores can provide specific information, it must be remembered that a relatively small sample of behavior in a short period of time has been examined. The bandwidth-fidelity dilemma (Cronbach, 1970) is resolved for the SITFILE only when the SIT, for the purpose of quickly assessing intelligence, is administered as but one part of a diagnostic battery. Under such circumstances no extra testing time is required. Additional diagnostic tests are administered and interpretation of SITFILE Language, Memory and Numerical Reasoning Category scores may provide supportive information by which to aid the formulation of diagnostic hypotheses.

The SITFILE is not a definitive diagnostic instrument. It is neither all encompassing nor comprehensive. Its Conceptual Thinking, Visual-Motor and Social Intelligence-Reasoning Categories do not provide any reliable information. The Language, Memory and Numerical Reasoning Categories do appear to represent a reliable and valid extension of the SIT's utility. Consequently, there appears to be at least partial empirical justification for the development of a SIT classification system and the employment of scatter analysis for its interpretation. But further investigation is necessary. The relatively small samples involved in this study were not broadly representative. If in follow-up studies SITFILE categories continue to prove reliable and valid, with larger and more broadly representative samples, then results derived from their interpretation can be considered when formulating diagnostic hypotheses regarding a child's specific learning aptitudes.

Follow-up studies should also investigate the reliability and validity of employing the SITFILE with "special" populations. Patterns of SITFILE scores could also be investigated to evaluate whether SIT scatter analysis can serve as a pathognomic sign. If such further investigations of SITFILE reliability and validity prove encouraging, standardization studies could be undertaken to develop norms for both "mainstream" as well as "special" populations.

SUMMARY

Previous studies of the SIT have dealt with its reliability and validity as a quick individual measure of intelligence. The present study was not concerned with further establishing these criteria or with evaluating the usefulness of the SIT as an instrument for screening special populations. Rather, this study was designed to explore the validity of extending the SIT's usefulness through the development of a classification system and application of standard deviation scatter analysis.

Many schemes for extending the usefulness of the S-B, a parent instrument of the SIT, have been suggested. Some of the more recent ones have been proposed by Meeker (1969), Sattler (1965) and Valett (1964). A degree of acceptance of these schemes, as well as increased diagnostic demands, have led to wider interest in classification systems and scatter analysis for the SIT. However, none of the published SIT schemes (Boyd, 1974; Stone, 1975) have presented empirical results to support their schemes' employment.

In the present three step investigation a SIT classification system, patterned after Sattler's (1965) S-B schema, was developed by having at least two out of four judges, using content analysis and a sorting technique, agree on the assignment of each SIT item from year two up to either

the Language, Memory, Conceptual Thinking, Numerical Reasoning, Visual-Motor or Social Intelligence-Reasoning Categories. A high percentage of agreement by three out of three judges on 73.7% of the items suggests that the resultant SITFILE has face validity. Analysis of the distribution of the items within both the Sattler S-B Binetgram and the SITFILE categories at four different age levels (two to five, six to tem, eleven to fourteen, and fifteen to adult) suggest that both the S-B and SIT share similar function assessments but different developmental designs. Thus, a degree of construct validity is indicated. However, the SIT's suggested underlying cognitive developmental design appears to be limited in its assortative diagnostic usefulness.

One hundred-fifty (150) Chicago parochial school students, grades two through eight, participated in an exploration of the SITFILES's reliability. Ninety-five (95) students attending a university diagnostic service center participated in the study of the SITFILE's validity. Individual category scores were calculated by using chronological age as the reference point for standard deviation scatter analysis. Although these samples are not broadly representative, they do provide a subtantial amount of data from which limited generalizations are possible.

Accordingly, the SITFILE Conceptual Thinking, Visual-Motor and Social Intelligence-Reasoning Categories were found to lack sufficient reliability, and performances on them should not be interpreted. The Language, Memory and Numerical Reasoning Categories were found generally to possess sufficient reliability for middle class white students in grades two through eight. For these categories, corrected internal consistency coefficients are reported ranging between .70 and .90. Adequate specificity, while somewhat lower for the Memory Category than for the Language and Numerical Reasoning Categories, is reported. Corrected Pearson stability coefficients between .73 and .98 for the Language, Memory and Numerical Reasoning Categories are also reported, as are small standard error of measurements.

A measure of each Language, Memory and Numerical Reasoning Categories' validity was obtained by correlating SITFILE category scores with age scores achieved by the ninety five (95) service center students on either the ITPA or the Detroit and the WRAT. Significant correlations ($p \leq .05$) suggest that the SITFILE Language, Memory and Numerical Reasoning Categories, for children six to fourteen, measure functions related to those measured by these frequently employed diagnostic instruments.

Considered in toto, the results of this investigation suggest that a classification system for the SIT can be articulated, but that only three of the six included categories have the requisite reliability and validity to justify their interpretation with even a narrowly defined middle class population of second through eight grade students. Furthermore, interactions suggested by large amounts of common variance and multiple correlations between the Language, Memory and Numerical Reasoning Categories and identified diagnostic tests argue against any independent interpretation of isolated SITFILE category scores. When used in conjunction with other diagnostic instruments, it does appear, however, that the proposed Language, Memory and Numerical Reasoning Categories and standard deviation scatter analysis interpretation of them may provide specific information facilitating the diagnostic process.

REFERENCES

Ackerman, P.T., Dykman, R.A. and Peters, J.E. Hierarchical factor patterns on the WISC as related to areas of learning deficit. Perceptual and Motor Skills, 1976, 42, 583-615.

Anderson, M., Kaufman, A.S. and Kaufman, N.L. Use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 13, 381-386.

Armstrong, R.J. and Jensen, J.A. The validity of an abbreviated form of the Stanford-Binet Intelligence Scale, Form L-M. Educational and Psychogical Measurement. 1972, 32, 463-467.

Armstrong, R.J. and Mooney, R.F. The Slosson Intelligence Test: implications for reading specialists. Reading Teacher, 1971, 24, 336-340.

Babcock, H. The level-efficiency theory of intelligence. Journal of Psychology, 1941, 11, 261-270.

Baker, H.J. and Leland, B. Examiners' Handbook Detroit Tests of Learning Ability, rev. ed. Indianapolis: Bobbs-Merrill, 1967.

Berko, M.J. Psychometric scatter: its application in the clinical prediction of future mental development in cases of childhood brain injury. Cerebral Palsy Review, 1955, 16, 16-18.

Bijou, S.W. A genetic study of diagnostic significant behavior patterns. American Journal of Mental Deficiency, 1942, 47, 171-177.

Boyd, J.E. Use of the Slosson Intelligence Test in reading diagnosis. Academic Therapy, 1974, 6 (6) 441-444.

Bradway, K.P. An experiemental study of factors associated with Stanford-Binet IQ changes from the preschool to junior high school. Journal of Genetic Psychology, 1945, 66, 107-128.

Bradway, K.P. and Thompson, C.W. Intelligence at adulthood: a twenty-five year follow-up. Journal of Educational Psychology, 1962, 53, 1-14.

Brighman, C.C. Two studies in mental test (1) variable factors in the Binet test. (11) The diagnostic value of some mental tests. Psychological Monographs, 1917, 24, 1 whole No. 102).

Carlisle, A., Shinedling, M and Weaver, R. Note on the use of the Slosson Intelligence Test with mentally retarded residents. Psychological Reports, 1970, 26, 865-866.

Cohen, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. Journal of Consul~~ta~~ting Psychology, 1959, 23, 285-299.

Cronbach, L.J. Essentials of Psychological Testing, (3rd ed.) New York: Harper and Row Publishers, 1970.

Delapa, G. Correlates of Slosson Intelligence Test, Stanford-Binet, Form L-M, and achievement indices (Doctoral dissertation, West Virginia University, 1967). Dissertation Abstracts International, 1973, 3498A. (University Microfilms No. 68-2678).

Doll, E.A. "Scattering" in the Binet-Simon Tests. Training School Bulletin, 1919, 16, 96-103.

Fromm, E., Hartman, L.D. and Marschak, M. A contribution to a dynamic theory of intelligence testing of children. Journal of Clinical and Experimental Psychology, 1954, 15, 73-95.

Garfield, S.L. An evaluation of Wechsler-Bellevue patterns in schizophrenia. Journal of Consulting Psychology, 1949, 13, 279-287.

Gilliland, A.R., Whittman, P. and Goldman, M. Patterns and scatter of mental abilities in various psychoses. Journal of Genetic Psychology, 1967, 17, 16-25.

Gittleman, M. and Birch, H.G. Childhood schizophrenia: Intellect, neurologic status, perinatal risk, prognosis, and family pathology. Archives of General Psychiatry, 1967, 17, 16-25.

Gurvitz, M.S. The Wechsler-Bellevue test and the diagnosis of psychopathic personality. Journal of Clinical Psychology, 1950, 6, 397-401.

Hammill, D. The Slosson Intelligence Test as a quick estimate of mental ability. Journal of School Psychology, 1969, 7, 33-37.

Hammill, D. Crancell, J.M. Jr. and Colarusso, R. The Slosson Intelligence Test adapted for visually limited children. Exceptional Children, 1970, 36, 535-536.

Harper, A.E. Discrimination of the types of schizphrenia by the Wechsler-Bellevue Scale. Journal of Consulting Psychology, 1950, 14, 290-296.

Harris, A.J. and Shakow, D. Scatter on the Stanford-Binet in schizophrenic, normal and delinquent adults. Journal of Abnormal Social Psychology, 1938, 134-150.

Harris, A.J. and Shakow, D. The clinical significance of numerical measures of scatter on the Stanford-Binet. Psychological Bulletin, 1937, 34, 134-150.

Heyer, A.W. "Scatter analysis" techniques applied to anxiety neurotics from a restricted cultural-educational environment. Journal of Genetic Psychology, 1949, 40, 155-166.

Houston, C., and Otto, W. Poor readers' functioning on the WISC, Slosson Intelligence Test and Quick Test. Journal of Educational Research, 1968, 62, 157-159.

Hutton, J.B. Relationships between teacher judgement, screening test data and academic performance for disadvantaged children. Training School Bulletin, 1972, 68, 197-201.

Jastak, J. A plan for the objective measurement of character. Journal of Clinical Psychology, 1948, 4, 170-178.

Jastak, J. Problems of psychometric scatter analysis. Psychological Bulletin, 1949, 49, 177-197.

Jastak, J.F. and Jastak, S.R. Manual of Instructions: Wide Range Achievement Test. Delaware: Guidance Association of Delaware, Inc., 1965.

Jerrolds, B.W., Callaway, B., and Gwaltney, W.K. Comparison of the Slosson Intelligence Test and WISC scores of subjects referred to a reading clinic. Psychology in the Schools, 1972, 9, 409-410.

Johnson, D.L. and Johnson, C.A. Comparison of four intelligence tests used with culturally disadvantaged children. Psychological Reports, 1971, 28, 209-210.

Jongeward, P.A. A validity study of the Slosson Intelligence Test for use with educable mentally retarded students. (Doctoral dissertation, University of Oregon, 1969). Dissertation Abstracts International, 30, 3323A. (University Microfilms No 70-25, 12).

Kaufman, A.S. Factor analysis of the WISC-R at eleven age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 1975, 43, 1350147.

Kaufman, A.S. Do normal children have "flat" ability profiles? Psychology in the Schools, 1976, 13, 28,284-285.

Kaufman, A.S. and Van Hagen, J. Investigation 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, 1977, 14, 10-14.

Kaufman, A.S. and Waterstreet, M.A. Determining a child's strong and weak areas of functioning on the Stanford-Binet: A simplification of Sattler's Standard deviation method. Journal of School Psychology, 1978, 16, 72-78.

Kaufman, H and Ivanhoff, J. The Slosson Intelligence Test as a screening instrument with a rehabilitation population. Exceptional Children, 1969, 35, 745.

Kendig, I, and Richmond, W.V. Psychological Studies in Dementia Praecox. Ann Arbor, Michigan: Edwards Bros., 1940.

Kirk, S.A., McCarthy, J.J. and Kirk, W.D. Examiner's Manual: Illinois Test of Psycholinguistic Abilities, rev. ed. Urbana; University of Illinois Press, 1968.

Lamp, R.E., and Traxler, A.J. The validity of the Slosson Intelligence Test for use with disadvantaged head start and first grade school children. Journal of Community Psychology, 1973, 1, 27-30.

Lamp, R.E., Traxler, A.J. and Gustafson, P.P. Predicting academic achievement of disadvantaged fourth grade children using the Slosson Intelligence Test. Journal of Community Psychology, 1973, 1, 339-341.

Lessler, K., and Galinsky, M.D. Relationship between Slosson Intelligence Test and WISC scores in special education candidates. Psychology in the Schools, 1971, 8, 341-344.

Levine, L.S. The utility of Wechsler's patterns in the diagnosis of schizophrenia. Journal of Consulting Psychology, 1949, 13. 28-31.

McFadden, J.H. Differential responses of normal and feebleminded subjects of equal mental ages on the Kent-Rosanoff Free Association Test and the Stanford Revision of the Binet-

Simon Intelligence Test. Mental Measurement Monographs, 1931, No. 7, 85.

McNemar, Q. The Revision of the S-B Scale. Boston: Houghton-Mufflin, 1942.

Machen, L.H. A validity and reliability study of the Slosson Intelligence Test with an atypical population - gifted children. (Doctoral dissertation, Catholic University of America, 1970) Dissertation Abstracts International, 1973, 2191-A (University Microfilms No. 70-22 140).

Magaret, A. Parallels in the behavior of schizophrenics, paretics, and pre-senile non-psychotics. Journal of Abnormal Social Psychology, 1942, 37, 511-528.

Magaret, A and Wright, C. Limitations in the use of intelligence test performance to detect mental disturbance. Journal of Applied Psychology, 1943, 27, 387-398.

Martin, J.D. and Rudolph, L. Correlates of the WAIS, SIT, ACT and grade point average. Educational and Psychological Measurement, 1972, 32, 459-462.

Mateer, F., The future of clinical Psychology. Journal of Delinquency, 1921, 6, 283-293.

Mathews, J. Irregularity in intelligence tests of delinquents. Journal of Delinquency, 1921, 6, 355-361.

Maxwell, M.T. The relationship between the Wechsler Intelligence Scale for Children and the Slosson Intelligence Test. Child Study Journal, 1971, 1 164-172.

Meeker, M.N. The Structure of Intellect. Columbus, Ohio: Charles E. Merrill, 1969.

Meissler, G.R. A correlation of the Slosson Intelligence Test and the Wechsler Intelligence Scale when administered to atypical children. (Doctoral dissertation, Catholic University of America, 1970). Dissertation Abstracts International, 1973, 2191-A (University Microfilms No. 70-22 140).

Merrill, M.A. On the relation of intelligence to achievement in the case of mentally retarded children. Psychological Monographs, 1924, 2, No 10, 100.

Nunnally, J.C. Psychometic Theory, New York: McGraw-Hill Book Co., 1978, 245.

Olch, Doris R. Psychometic pattern of schizophrenics, paretics, and pre-senile non-psychotics. Journal of Consultanting Psychology, 1948, 12, 127-136.

Paraskevokoulow, J.N. and Kirk, S.A. The Development and Psychometric Characteristics of the Revised Test of Psycholinguistic Abilities. Urbana: University of Illinois Press, 1969.

Piaget, J. Science of Education and the Psychology of the Child. New York: Orion Press, 1970.

Pignatelli, M.L. A comparative study of mental functioning patterns of problem and non-problem children seven, eight, and nine years of age. Genetic Psychology Monographs, 1943, 27, 69-162.

Piotrowski, Z.A. Objective signs of invalidity of Stanford-Binet tests. Psychiatric Quarterly, 1937, 11, 623-636.

Pressey, S.L. and Cole, L.W. Irregularity in a psychological examination as a Measure of Mental Deterioration. Journal of Abnormal Psychology, 1919, 13. 285-294.

Rabin, A.I. Test-score patterns in schizophrenia and non-psychotic states. Journal of Psychology, 1941, 12, 91-100.

Rabin, A.I. Differentiating psychometric patterns in schizophrenia and manic-depressive psychosis. Journal of Abnormal Social Psychology, 1942, 37, 270-272.

Rapaport, D. Diagnostic Psychological Testing, Volume I. Chicago: Year Book Publishers, 1945.

Ritter, D., Duffey, J. and Fischman, R. Comparability of Slosson and Stanford-Binet estimates of intelligence. Journal of School Psychology, 1973, 11, 224-227.

Roe, A. and Shakow, D. Intelligence in mental disorders. Annals of the New York Academy of Sciences, 1942, 361-490.

Rugel, R.P. WISC subtest scores of disabled readers: A review with respect to Bannatyne's recategorization. Journal of Consulting Psychology, 1951, 15, 62-67.

Sattler, J.M. Analysis of functions of 1960 S-B intelligence Scale Form L-M. Journal of Clinical Psychology, 1965, 21 173-179.

Sattler, J.M. Assessment of Children's Intelligence. Philadephia: W.B. Saunders Co., 1974, 129-.38.

Schafer, R. The significance of scatter in research and practice of clinical psychology. Journal of Psychology, 1944, 18, 119-124.

Schafer, R. and Rapaport, D. The scatter in diagnostic intelligence testing. Character and Personality, 1944, 12, 275-284.

Schneider, J. and Smillie, D. The use of scatter on the Stanford-Binet. Psychological Service Center Journal, 1959, 11, 73-75.

Schofield, W. Critique of scatter and profile analysis of psychometric data. Journal of Clinical Psychology, 1952, 8, 16-22.

Schwebel, M. and Raph, J. Piaget in the Classroom, New York: Basic Books, Inc., 1973.

Seashore, H.G. Differences between verbal and performance IQs on the WISC. Journal of Consulting Psychology, 1951, 15, 62-67.

Shakow, D. and Millard, M.S. A psychometric study of 150 adult delinquents. Journal of Social Psychology, 1935, 6, 437-457.

Silverstein, A.B. Comparison of two item-classification schemes for the S-B. Psychological Reports, 1965, 17, 964.

Silverstein, A.B. WISC subtest patterns of retardates. Psychological Reports, 1968, 23, 1061-1062.

Silverstein, A.B. Variance components in the subtests of the WISC-R. Psychological Reports, 1976, 39, 1109-1110.

Sloan W. and Cutts, R.A. Test patterns of defective delinquents on the Wechsler-Bellevue test. American Journal of Mental Deficiency, 145, 50, 95-97.

Slosson, R.L. Slosson Intelligence Test for Children and Adults. East Aurora, New York: Slosson Educational Publications, 1963.

Slutzky, J.E., Justman, J. and Wrightston, J.W. Screening children for placement in special classes for the mentally retarded: A preliminary report. American Journal of Mental Deficiency, 1953, 57, 687-690.

Smith, M.D., Coleman, J.M., Dokecki, P.R., and Davis E.E. Recategorized WISC-R scores of learningdisabled children. Exceptional Children, 1977, 10, 444-449.

Stewart, K., and Jones, E. Validity of the Slosson Intelligence Test: a ten-year review. Psychology in the Schools, 1976, 10, 13(4), 372-380.

Stewart, K.D. and Myers, D.G. Long-term validity of the Slosson Intelligence Test. Journal of Clinical Psychology, 1974, 30, 180-181.

Stewart K.D., Wood, D.Z. and Gallman, W.A. Concurrent validity of the Slosson Intelligence Test. Journal of Clinical Psychology, 1971, 27, 218-220.

Stone, M. An interpretive profile for the Slosson Intelligence Test. Psychology in the Schools, 1975, 12, 3, 330-333.

Strother, C.R. The performance of psychopaths on the Wechsler-Bellevue test. Academy of Science, 1944, 51, 397-400.

Swanson, M.S. and Jacobson, A. Evaluation of the SIT for screening children with learning disabilities. Journal of Learning Disabilities, 1970, 3, 318-320.

Thomson, G.H. A note on scaling tests. Journal of Educational Psychology, 1926, 17, 551-553.

Thorndike, R.L. Reliability in E.F. Lindquist (ed.) Educational Measurement. Washington, D.C.: American Council on Educationation, 1951, 560-620.

Valett, R.E. A clinical profile for the S-B. Journal of School Psychology, 1964, 2, 49-54.

Vance, H.B., Gaynor, P. and Coleman, M. Analysis of cognitive abilities for learning disabled children. Psychology in the Schools, 1976, 13, 477-483.

Vane, J.R., Weitzman, J. and Applebaum, A.P. Performance of Negro and white children and problem and nonproblem children on the Stanford Binet Scale. Journal of Clinical Psychology, 1966, 22, 431-435.

Wallin, J.E.W., Intelligence irregularity as measured by scattering in the Binet Scale. Journal of Educational Psychology, 1922, 13, 140151.

Wallin, J.E.W. A further note on scattering in the Binet Scale. Journal of Applied Psychology, 1927, 11, 143-154.

Wallin, J.E.W. A statistical study of the individual tests in ages VII and IX in the Stanford-Binet Scale. Mental Measurement Monographs, 19291 6, 58.

Warner, S.J. The Wechsler-Bellevue psychometric pattern in anxiety neurosis. Journal of Consulting Psychology, 1950, 14, 297-304.

Weisenburg, T., Roe, A. and McBride, K.E. Adult Intelligence. A Psychological Study of Test Performances. New York: Commonwealth Fund, 1936.

Wells, F.L. Mental Tests in Clinical Practice. Yonkers, New York: World Book Company, 1927, 315.

Wells, F.L. and Kelly, C.M. Intelligence and psychosis. American Journal of Insanity, 1920, 77, 17-45.

Woodworth, R.S. Computing the standard deviation and probable error of Binet mental age. Psychological Bulletin, 1928, 25, 167.

Zingale, S.A. and Smith M.D. WISC-R patterns for learning disabled children at three SES levels. Psychology in the Schools, 1978, 15, 199-204.

REFERENCE NOTES

Canfield, J.B. Item analysis for Slosson Intelligence Test. Southwest Cook County Cooperative Association for Special Education. Tinley Park, Illinois, 1975.

Hedberg, N. and Shapiro, A. Process-Function Analysis of the Slosson Intelligence Test and Diagnostic Implications. Masters Project, Northeastern Illinois University, Department of Special Education, Chicago, Illinois, 1976.

Project S.U.C.C.E.S.S., Title II - E.S.E.A., Wayne, Nebraska, 1975.

APPENDIX A

Appendix A-1: Item Classifications and Judge Concordances

		Item Classification						Concordance by		
	TEST ITEM	Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	3/3 Judges	2/3 Judges	2/4 Judges
	Produce 3 word sentence	2-0						2-0		
	Copy verticle & horizontal lines					2-0		2-0		
	Where is a chair and legs of the chair?	2-1						2-1		
	Give me the pencil						2-2		2-2	
	Give me the paper						2-3	2-3		
	What is this? (book)	2-4						2-4		
2	What do you hear with?						2-5	2-5		
	Show me your fingers and shoes						2-6		2-6	
	Where is the floor?	2-7						2-7		
	Show me the window and door	2-8						2-8		
	Say these numbers: 3, 5, or 2, 6.		2-9					2-9		
	Show me your teeth and chin						2-10		2-10	
	Are you a boy or girl?						2-11	2-11		
	Say these numbers: 8, 4, 1, or 7, 3, 5.		3-0					3-0		
	Copy a drawing of a cookie					3-1		3-1		
	Put a pencil on top of a book						3-2		3-2	
	Put a pencil under the book						3-3			3-3
	Put paper inside the book						3-4			3-4
	Say "Baby sleeps in a little bed."		3-5					3-5		
	Which square is smaller?						3-6			3-6

Appendix A-1 continued

	TEST ITEM	Item Classification: Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	Concordance by: 3/3 Judges	2/3 Judges	2/4 Judges
3	Show me your tongue and neck	3-7						3-7		
	Where is your arm and knee?						3-8		3-8	
	How many apples am I drawing?				3-9			3-9		
	Show me your thumb						3-10		3-10	
	Why do we have to take a bath?						3-11	3-11		
4	How many apples am I drawing?				4-0			4-0		
	A hat goes on your head, shoe go ? (feet)			4-1					4-1	
	Fire is hot, ice is ? (cold)			4-2				4-2		
	Where is your heel?						4-3		4-3	
	The ceiling is up, the floor is ? (down)			4-4				4-4		
	Say: "I have fun playing with ..."		4-5					4-5		
	When you are asleep your eyes are shut, awake eyes ?						4-6		4-6	
	Why do we have clocks?						4-7		4-7	
	Say these numbers: 2,9,5,3, or 8,4,1,7.		4-8					4-8		
	How many apples do you see?				4-9			4-9		
	Say: "I go to the store to buy milk ..."		4-10					4-10		
	Milk is white. Butter is ? (yellow)			4-11				4-11		
5	How many apples?				5-0			5-0		
	Copy the picture of a block					5-2		5-2		
	Which is bigger, a cat or mouse			5-4					5-4	
	What number comes after 8?				5-6				5-6	
	If I cut an apple in half, how many pieces do I have				5-8			5-8		
	How are crayon and Pencil Different and the same?			5-10				5-10		

Appendix A-1 continued

	TEST ITEM	Item Classification						Concordance by		
		Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	3/3 Judges	2/3 Judges	2/4 Judges
6	A lemon is sour, sugar is ? (sweet)			6-0				6-0		
	What is a forest made of?						6-2		6-2	
	How are milk and water different and the same?			6-4				6-4		
	How are a cat and dog differerent and the same?			6-6				6-6		
	A carrot is a vegetable. An apple is ? (fruit)			6-8				6-8		
	What does brave mean?	6-10						6-10		
7	Say these numbers: 8,5,1,9,2, or 7,3,6,4,1.		7-0					7-0		
	How many days in a week?		7-2						7-2	
	Copy drawing of a kite					7-4		7-4		
	How many eggs in a dozen?		7-6						7-6	
	How are a submarine and a fish different and alike?			7-8				7-8		
	How many months in a year?		7-10						7-10	
8	Say these numbers backwards: 4,7,3 or 6,2,9		8-0					8-0		
	Who was Christopher Columbus?		8-2						8-2	
	What do we mean by infection?	8-4						8-4		
	What does ham come from?						8-6		8-6	
	Say: "The train goes fast on ..."		8-8					8-8		
	What does destroy mean?	8-10						8-10		
9	What is a hero?	9-0						9-0		
	What is paper made of?						9-2		9-2	
	What does vacant mean?	9-4						9-4		
	Divide 28 marbles in half ...				9-6			9-6		
	What was a dugeon used for?	9-8						9-8		
	How many inches in 2 feet?				9-10			9-10		

Appendix A-1 continued

	TEST ITEM	Item Classification: Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	Concordance by: 3/3 Judges	2/3 Judges	2/4 Judges
10	How many minutes in 3/4 hour?				10-0			10-0		
	What month has only 28 days?		10-2						10-2	
	What does magnify mean?	10-4						10-4		
	How in use are a telescope and microscope different?			10-6					10-6	
	How many 5¢ in 45¢?				10-8			10-8		
	Say these numbers backwards: 8,6,9,4 or 3,1,7,5.		10-10					10-10		
11	How are a clock & calendar different and the same?			11-0				11-0		
	What does it mean to be thrifty?	11-2						11-2		
	Say these numbers: 9,3,5,8,6 or 7,4,8,1,9,2.		11-4					11-4		
	What do you do if take inventory?	11-6						11-6		
	How many days in a year?		11-8						11-8	
	Have 36 eggs. How many have if you broke half?				11-10			11-10		
12	Say numbers backwards: 7,4,2,8,1 or 5,3,6,9,2.		12-0					12-0		
	5¢ foot. How many feet buy for 65¢?				12-2			12-2		
	What does scarce mean?	12-4						12-4		
	How many inches in 2 yards?				12-6			12-6		
	Sitter charges 50¢ hour-works 5 1/2 hours - Makes?				12-8			12-8		
	What is a healthy person's temperature?		12-10						12-10	
13	How many feet in 3 yards?				13-0			13-0		
	How many pints in a gallon?				13-2				13-2	
	How many pounds in a ton?		13-4						13-4	
	Difference between contraction & expansion?			13-6					13-6	
	What does tremendous mean?	13-8						13-8		
	Difference between latitude & longitude?			13-10					13-10	

Appendix A-1 continued

Age	TEST ITEM	Item Classification: Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	Concordance by: 3/3 Judges	2/3 Judges	2/4 Judges
14	Had 40¢ - gave 10¢ away. What fraction left?				140			140		
	What does mutiny mean?	14-2						142		
	What is the principal work of a pharmacist?	14-4						144		
	Have 12 lb. turkey - cook 20 min.per lb. When start?				146			146		
	What does abundant mean?	14-8						14-8		
	Make 50¢ hour-work 1 1/2 hr. 6 Sat. How much make?				1410			1410		
15	What is principal work of an architect?	150						150		
	How many feet in a mile?		152							152
	What does fragrant mean?	154						154		
	Area of room 9X12 in sq. feet?				156			156		
	How are octave and octopus alike?		158					15-8		
	What does environment mean?	1510						1510		
16	How much change left from $5.00?				16-0			160		
	What does detain mean?	163						16-3		
	What is a deficit?	16-6						166		
	Total 300 people - men to women 5-1. How many men?				169					
17	What does mutilate mean?	17-0						170		
	What does facsimile mean?	173						173		
	What does malicious mean?	176						17-6		
	What does simultaneous mean?	179						179		
	What is a hypotenuse?	180						18-0		
	Say numbers backwards: 8,3,2,9,4, or 5,2,7,4,1,6		183					183		

Appendix A-1 continued

	TEST ITEM	Item Classification: Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	Concordance by: 3/3 Judges	2/3 Judges	2/4 Judges
18	What is the circumference of the earth in miles?		18-6							18-6
	What is a protoza?	18-9						18-9		
19	What fraction of million $ would inherit: 3/5 2/3 5/8				19-0			19-0		
	How many degrees in the interior angles of triangle?				19-3			19-3		
	What is a panorama?	19-6						19-6		
	How many min. take go 1 mile @ 40 mph?				19-9			19-9		
20	What is a plebisicite?	20-0						20-0		
	Have $40,000 with ratio 5/3. How much daughter share				20-3			20-3		
	What does prognosticate mean?	20-6						20-6		
	What was the average rate of speed on trip?				20-9			20-9		
21	Dog is canine. Cat is ? (feline)			21-0				21-0		
	What would these people carry to identify themselves	21-3						21-3		
	Who wrote the following books?		21-6						21-6	
	What was the average rate of speed on trip?				21-9			21-9		
22	What is an amulet?	22-0						22-0		
	What is the cube root of 216?				22-3			22-3		
	What is the difference between vortex and vertex?	22-6							22-6	
	How many eggs did the farmer sell as the market?				22-9			22-9		
23	What are meanings of following foreign expressions?	23-0						23-0		
	How much money did man make work 12 days?				23-3			23-3		
	Distinguish between paleontologist & philologist	23-6							23-6	

Appendix A-1 continued

	TEST ITEM	Item Classification: Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual-Motor	Social In. Reasoning	Concordance by: 3/3 Judges	2/3 Judges	2/4 Judges
	What does prevaricate mean?	23-9						23-9		
24	How many rabbits have after 4 years?				24-0			24-0		
	How far was man from his starting point?				24-3			24-3		
	Meaning of ubiquitous?	24-6						24-6		
	What are the parts of animal's body indicated?	24-9						24-9		
25	What is the meaning of prestidigitation?	25-0						25-0		
	What is uxoricide?	25-3						25-3		
	What is meaning of raiocination?	25-6						25-6		
	Who was the god of dreams?		25-9						25-9	
26	What is an amanuensis?	26-0						26-0		
	What is another name for mercy killing?	26-3						26-3		
	What is the difference between plutocracy & theocracy	26-6							26-6	
	A cow is ? A man is ? (omnivorous)			26-9				26-9		
27	Why anthropaphagite relish visit from anthropologist	27-0							27-0	

SITFILE

Age	Language	Memory	Conceptual Thinking	Numerical Reasoning	Visual Motor	Social Intelligence and Reasoning
2	0 1 4 7 8	9			0	· 2 3 5 6 7 10 11
3	7	0 5		9	1	2 3 4 6 8 10 11
4		5 8 10	1 2 4 11	0 9		3 6 7
5			4 10	0 6 8	2	
6	10		0 4 6 8			2
7		0 2 6 10	8		4	
8	4 10	0 2 8				6
9	0 4 8			6 10		2
10	4	2 10	8	0 8		
11	2 6	4 8	0	10		
12	4	0 10		2 6 8		
13	8	4	6 10	0 2		
14	2 4 8			0 6 10		
15	0 4 10	2	8	6		
16	3 6			0 9		
17	0 3 6 9					
18	0 9	3 6				
19	6			0 3 9		
20	0 6			3 9		
21	3	6	0	9		
22	0 6			3 9		
23	0 6 9			3		
24	6 9			0 3		
25	0 3 6	9				
26	0 3 6		9			
27	0					

-3SD ______
-2SD ______
-1SD ______
CA ______
+1SD ______
+2SD ______
+3SD ______

-3SD ______
-2SD ______
-1SD ______
MA ______
+1SD ______
+2SD ______
+3SD ______

Name ____________ Sex ________ Date ____________ IQ ____

APPROVAL SHEET

The dissertation submitted by Andrea Rolsky has been read and approved by the following committee:

Dr. Ronald Morgan, Director
Associate Professor, Foundations of Education, Loyola

Dr. Jack Kavanagh
Associate Professor and
Chairman, Foundations of Education, Loyola

Dr. Martha Wynne
Assistant Professor, Curriculum and Instruction, Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Education.

9/25/81
Date

Ronald R. Morgan
Director's Signature