1977

Predictive validity of the Smile Early Screening Test

Allan Howard Solomon
The University of Montana

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation
https://scholarworks.umt.edu/etd/7860

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
PREDICTIVE VALIDITY OF THE SMILE
EARLY SCREENING TEST

By

Allan Solomon
B.A., Columbia University, 1973

Presented in partial fulfillment of
the requirements for the degree of

Master of Arts
UNIVERSITY OF MONTANA
1977

Approved by:

[Signatures]
Chairman, Board of Examiners
Dean, Graduate School

May 17, 1977

Date
ABSTRACT

Solomon, Allan, M.A., Spring 1977 Psychology

Predictive Validity of the Smile Early Screening Test (92 pp.)

Director: Dr. Herman A. Walters

The predictive validity of the Smile Early Screening Test (SEST; revised edition, 1975) was investigated by correlating the above and below average scores of two groups of kindergarten children (N=62) on this test with their scores on the Metropolitan Readiness Tests (MRT; revised edition, 1965). Results indicated high predictive validity for the SEST with 52 out of 56 overall and individual subtest correlations being significant at least at the .05 level. The data indicated that the SEST can predict children's achievement in global cognitive-intellectual areas as well as in several specific academic subjects. In addition, intercorrelations among SEST subtests were generally moderate suggesting each subtest measure contributes uniquely to the overall screening test score. Suggestions for revising the SEST were offered and the need for further research on the SEST including a factor analysis was discussed. The present study supported using the SEST as a precise screening device in kindergarten and first grade after using the MRT and teachers' observations as initial screening devices.
# TABLE OF CONTENTS

ABSTRACT ................................................................. ii
LIST OF TABLES ......................................................... iv

Chapter

I  INTRODUCTION ......................................................... 1
   Research Goals .................................................... 3
   Brief History ....................................................... 6
   Theories of Early Identification ................................. 12
   Battery Approaches ................................................. 22
   Intelligence and Achievement Tests ............................ 27
   Tests Developed Specifically for Early Identification .... 30
   Miscellaneous Identification and Screening Procedures .... 35

II  METHOD ................................................................. 40
   Subjects .............................................................. 40
   Procedure .......................................................... 43

III  RESULTS ............................................................... 44

IV  DISCUSSION ............................................................ 48

V  SUMMARY ................................................................. 62

REFERENCES ............................................................... 65

APPENDIXES
   Appendix A: Smile Early Screening Test ......................... 73
   Appendix B: Smile Early Screening Test Record Form ........... 81
   Appendix C: Smile Early Screening Test Normative Data ....... 83
   Appendix D: Smile Early Screening Test Reliability Coefficients . 86
   Appendix E: Metropolitan Readiness Tests ....................... 88

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Socioeconomic Status Based on Parents' Occupation for Missoula and Anaconda</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Samples</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Summary of Correlations Between Smile Early Screening Test and Metropolitan</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Readiness Tests Subtest and Total Scores (N=62)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Summary of Within-Test Correlations of the Smile Early Screening Test (N=62)</td>
<td>47</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Much attention, particularly in the last decade, has been focused on the need to prevent and remedy school learning difficulties. As a result of this nationwide effort, there has been a greater emphasis on early identification of those children who need special educational planning and resources to learn successfully (Maitland, Nabdeau, & Nabdeau, 1974). In addition, many educators and psychologists have recognized the lack of reliable data concerning those intellectual factors which might affect academic achievement in the early school years (Asbury, 1974). Professionals concerned with identifying learning problems early have, therefore, used a wide variety of screening procedures to pinpoint potential sources of learning difficulty in children about to enter kindergarten and first grade. These procedures range from attempts to use number of permanent teeth as a criterion (Goll, 1930) to the use of well standardized intelligence tests (Vingoe, Birney, & Kordinak, 1969). In recent years, however, investigators have been increasingly concerned with developing specific tests for early identification. Examples include the Metropolitan Readiness Tests (revised edition, 1965) and the Smile Early Screening Test (revised edition,
1975) which will be the focus of the present investigation. Developers of these screening devices intend them to: 1) pinpoint deficiencies more accurately than is possible with standardized achievement tests, and 2) provide the teacher with more specific information about which particular problems require remediation. Since this trend in test development is a potentially important contribution to both the process of early identification and the basic literature in early intellectual growth, more information on the validity and reliability of these screening instruments is clearly needed. The Smile Early Screening Test (revised edition, 1975) is similar to several of these tests in practical intent and theoretical foundation. Investigating its predictive validity should, therefore, yield data of practical and theoretical significance for both more precise identification and contributions to basic developmental literature.

Three of the main purposes of early identification are: 1) to facilitate placement of children in classes providing appropriate preventative or remedial procedures for evident deficits; 2) to prevent unfavorable socio-emotional consequences which may result from an undiagnosed learning problem; and 3) to give parents information on their children's learning potential and particular problems to facilitate their involvement in the educational process (Rogolosky, 1969; Planz, 1972; Evans, 1973). In addition, being able
to predict learning difficulties may be the major rationale for the existence of current remedial programs, many of which are expensive and tainted with ideological controversy (Katz, 1975). Three major reasons often noted for not using screening are: 1) the additional funds and professional staff which many early identification procedures require; 2) the fact that kindergarten itself may serve as an effective screening experience for many children; and 3) the danger that certain children will be labeled "learning disabled" and thus be treated poorly by teachers and other children (Maitland et al., 1974). The research goals of the current investigation will be discussed followed by a brief historical and theoretical background for identifying learning problems early; in addition, a review of the literature on various types of screening procedures will be presented.

Research Goals

In general, the purpose of the present investigation was to explore the predictive validity of the Smile Early Screening Test (SEST; revised edition, 1975) by correlating the scores of two groups of children on this test with their scores on the Metropolitan Readiness Tests (MRT; revised edition, 1965). The first group of children had scored at or above the mean on their final SEST Total, while the second group had scored below the mean. The SEST was viewed as a more efficient screening device than the MRT because:
1) it can be used with entering kindergarten children, and
2) it is offered in conjunction with a specific treatment program providing precise remediation for problems that require educational attention. As the SEST's authors declare in their introduction:

The test is not designed to identify (i.e., label) children who are slow learners, learning disabled, etc. However, it should be helpful in pinpointing the educational needs and the starting points for beginning instruction with each child. (p. ii).

Thus the SEST was viewed as functionally more effective for screening and remediation than the MRT; however, the MRT was used as a device for confirming the predictive validity of the SEST because it is also a reliable screening instrument (see Appendix E). In particular, the major hypotheses of the present research were:

1. For both groups of children correlations between their Total MRT scores and their scores on three overall SEST measures would be significant. These overall SEST measures were Total, Learning Aptitude, and Achievement.

2. The Language Arts score on the MRT which includes the composite of four subtests was viewed as a central predictor of academic achievement and would correlate significantly with SEST Achievement and Learning Aptitude scores for both groups. The four MRT subtests included in the Language Arts score were Word Meaning, Listening, Matching, and Alphabet.

3. Alphabet on the MRT would correlate highly with Alphabet on the SEST for both groups; in addition, Arithmetic on the MRT would correlate highly with Arithmetic on the SEST for both groups. Finally, Copying on the MRT would correlate highly with Copying on the SEST for both groups. These three correlations
represent pairs of scores hypothesized to measure similar abilities besides representing central predictors of academic achievement on the SEST and the MRT.

The primary limitation of the present study was that its subject population consisted of two groups of children from two separate communities, Missoula and Anaconda, Montana. Although socioeconomic classification revealed no significant differences between the two groups, it is possible that other differences exist between Missoula and Anaconda children which could have biased the results of the investigation. For example, there is a generally greater emphasis in Missoula public schools on children learning the alphabet early in the school year; however, as noted in the discussion section, teachers even within the two school systems vary widely in how much they stress this skill. On the other hand, using children from separate communities may have increased the generalizability of the conclusions of the investigation. The present research was primarily exploratory in nature and data gained should be of value in further revisions of the SEST as well as providing useful information for other researchers interested in designing screening procedures for early identification. Moreover, the present study should stimulate further predictive validity investigations of the instruments being used to screen learning difficulties in young children. Finally, this investigation contributed to the literature in early childhood
development by attempting to demonstrate the plausibility and utility of measuring early intellectual growth in relation to later school achievement—an issue which has concerned developmental theorists for decades.

Brief History

The development of tests intended partially for early identification of school learning disabilities began in 1891 with the work of Munsterberg in Germany (Peterson, 1925). Unfortunately, Munsterberg did not publish the results of his tests, but some of the tasks included reading aloud as rapidly as possible, stating the colors of ten objects named on a sheet, adding ten single-digit numbers, and constructing a square and equilateral triangle out of puzzle parts. While Munsterberg was sharply criticized by Binet and others for emphasizing "trivial mental processes" and devoting too much attention to speed of task performance (Peterson, 1925), his tests were employed in German elementary schools for diagnostic purposes. Other researchers who developed similar diagnostic tests during this early period were Boas and Bolton who tested vision, hearing, and memory in 1500 Massachusetts school children in 1891, and Gilbert who administered several mental tests to about 1200 children in New Haven in 1893 (Peterson, 1925). While the former were primarily interested in the relationship between memory and teachers' ratings of "intellectual acuteness," the
latter correlated performance on a variety of perceptual-motor tasks with "general ability" as estimated by teachers. Interestingly enough, Boas and Bolton as well as Gilbert found slight positive relationships between their measures and teachers' ratings.

All the above researchers designed tests which had identification of children's potential problems as one goal; however, the first test which was used specifically for identification and screening was Binet and Simon's 1905 Intelligence Scale (Peterson, 1925). In 1904 Binet had been commissioned by the Minister of Public Instruction in Paris to develop a test which would "select the subnormal children from the normal" (quoted in Peterson, 1925, p. 168). The objective of the testing procedure was to have the "subnormal" or mentally retarded children placed in special schools since they were unable to benefit from regular school instruction. The criteria for admission to the remedial schools included scores on the Intelligence Scale and a physical examination to determine degree of organic or sensory involvement. The Intelligence Scale consisted of 30 tests arranged according to a "metrical scale of intelligence" (Peterson, 1925, p. 169).

As Binet explained:

This scale is composed of a series of tests arranged in order of increasing difficulty, beginning at one end with the lowest intellectual level that one can observe and extending at the other to the level of average and normal intelligence (Peterson, 1925, p. 169).
The Intelligence Scale assessed competence in areas similar to those assessed by current preschool tests including: "visual coordination" by observing eye movements as a match was passed before the subject's eyes; "verbal knowledge of objects" by asking the subject to touch his head, nose, ear, and three common objects; "memory" by having the subject repeat sentences after the examiner and draw two different diagrams shown simultaneously for ten seconds; and "knowledge of differences" by asking the subject to explain the difference between, for example, wood and glass. Binet clearly recognized that the 1905 Scale was inadequately standardized and overemphasized "logical judgment" as the primary criterion for intellectual assessment. Accordingly, he worked with Simon to make two revisions of the Scale in 1908 and 1911. Binet felt, however, that his 1905 Scale had clearly demonstrated the possibility of determining in a precise and truly scientific way the mental level of an intelligence, to compare this level with a normal level, and consequently to determine by how many years a child is retarded (Peterson, 1925, p. 185).

As testimony to Binet's claim, his scales were translated into English by H. Goddard as well as into many other languages by psychologists throughout the world. In the United States they were first used to identify problems at the Training School in Vineland, New Jersey in 1910 (Robb, Bernadoni, & Johnson, 1972). His 1905 Scale, which could be
used with children as young as three, had also implied an important idea: the possibility of obtaining measures of the child's intellectual strengths and weaknesses in the earliest preschool years.

As the Binet Intelligence Scales were used more widely in Europe and the United States, a new idea gained acceptance in education. The importance of measuring children's intellectual capabilities in first grade, kindergarten, or even earlier was recognized by many authorities. Terman, who published the Stanford Revision of the Binet-Simon Scale at Stanford University in 1916, clearly saw the usefulness of early measurement and was convincing in presenting his views:

There is one reason why tests are more necessary in the kindergarten than anywhere else, if the intellectual differences which exist among pupils are to be discovered. In other school grades the work itself constitutes a kind of intelligence test. The first grade child who cannot learn to read, or the fourth grade child who cannot learn long division is readily recognized as inferior. The work of the average kindergarten offers no such clear-cut criterion of intellectual abnormality. The games, drawing, sand-pile activities and cardboard construction may disclose certain differences, but these are vague and lack meaning.

The first grade is the most critical in the school system. It is the place above all others where the raw material with which the school is to work should be correctly evaluated. It is in the first grade that retardation scores its worst record.

Schools for backward children ordinarily do not draw from classes below the third grade. By this time the dull pupil is already a lost cause. The first task of the school
when it gathers its newcomers together should be to give each child a mental test to determine the nature of his endowment. The test should then be checked up by a large amount of supplementary data and by an annual appraise-ment of progress (Terman, 1919, pp. 37-38, 42).

Terman's rationales for early assessment through testing, offered nearly 60 years ago, do not sound very different from the reasons offered by contemporary proponents of early identification. The inadequacy of conventional kindergartens for precise assessment and the need for appraisal while the child still has access to and can effectively use remedial programs were clearly present in his thinking. As the idea of early identification of those children who might have learning problems became more widespread, studies appeared attempting to delineate factors which might affect academic success in first grade and intellectual competence in the preschool years.

Zornow and Pechstein (1922) concluded that chronological age was not a reliable indicator of readiness to do first grade work, and believed mental age as measured by the Stanford-Binet was far more reliable for screening. They tested 55 six-year-olds who had not been promoted in first grade and found 89 percent had mental ages below six years while 67 percent had mental ages below 5 1/2 years. Other studies found chronological age was an extremely poor predictor of reading readiness in first grade while mental age on the Binet was a much better indicator (Thiesen, 1921;
Brooks, 1924). Goll (1930) studied the relative value of chronological age, mental age as measured by the Detroit First Grade Intelligence Test, kindergarten training, and number of permanent teeth (an index of "anatomical age") for predicting first grade promotion. He concluded that mental age was the best predictor, and that the addition of any of the other factors to mental age, by means of the multiple correlation technique, did not improve the prediction of promotion obtained through the use of mental age alone (p. 69).

Wellman (1937) attempted to assess the intelligence of preschool children using the Merrill-Palmer Scale of Performance Tests. She found positive correlations between IQ on the Merrill-Palmer and length of nursery school attendance, measures of physical activity level, chronological age, and Stanford-Binet IQ. No significant correlations were found with father's occupational class, measures of leadership, and measures of introversion-extraversion. Moreover, significant negative correlations were found between Merrill-Palmer IQ and elevations on scales measuring compliance and likelihood of behavior problems. Other studies (Driscoll, 1933; Barrett and Koch, 1930) obtained similar results, but Goodenough (1930) found that preschool girls did better than preschool boys on many intellectual measures at comparable age and training levels.

A thorough search of the literature revealed no studies on early identification and screening procedures throughout
the 1940s and most of the 1950s. The dearth of published research during these decades is interesting, since many important developmental theories such as Gesell's (1938) flourished during this period. Interest in this area may have declined during this period, or perhaps researchers conducted small-scale investigations which did not enter the psychological literature. In any case, there was a tremendous resurgence of interest in the theories and procedures of early identification in the early and middle 1960s as preschool programs proliferated. In fact, volume of publication increased dramatically throughout the late 1960s and has continued to increase up to the present.

Theories of Early Identification and Screening

While no developmental theory relates strictly to early identification and screening procedures, the approaches of a number of theorists provide substantive support for attempts to identify cognitive-intellectual problems and assets in the preschool and early school years. The major theorists considered in this section have all delineated stages or intellectual processes which children go through in intellectual growth and adaptation; these theorists are Gesell, Piaget, Werner, and Newland. Moreover, each has used his theoretical conceptualizations to design tests and procedures assessing various aspects of the intellectual growth process. Since many of these tests are used in intellectual screening and
since their underlying conceptualizations provide a well thought out basis for such procedures, it appears worthwhile to examine these theories as they relate to the concept of early identification.

Gesell became interested in the intellectual and motoric growth of normal children after developing a manual for teaching the "deficient" child (1918). He viewed intelligence as "the most objective and measurable aspect of mental phenomena," and believed the preschool period was most important for intellectual growth (Gesell, 1925, p. 17). Accordingly, he devised a series of nine "developmental schedules" following the neonatal period and arranged as follows: 1) 3 to 4 months; 2) 6 months; 3) 9 months; 4) 1 year; 5) 1 1/2 years; 6) 2 years; 7) 3 years; 8) 4 years; 9) 5 to 6 years (Gesell, 1925, p. 21). The term, "developmental schedule," refers to a normative grouping of tasks within each age level which children should be able to perform if they are developing within average limits. Gesell based his norms on data gathered at the Yale Clinic of Child Development beginning in 1919. Fifty middle-class children examined at 4, 6, 9, 12, 18, 24, 36, 48, and 60 months of age constituted the population sample.

Gesell called his schedules part of a "system of developmental diagnosis," viewing them as one step in the identification and screening process:
The developmental schedules as drawn up are designed to serve as instrumental aids in arriving at comparative judgments. These judgments may be of a broad and approximate character or they may be analytical and refined, depending upon the time which is extended in the examination and depending upon the experience of the examiner. The normative items are not to be used in a mechanical or purely psychometric manner. We must not simply measure the child; we should try to apperceive him in an interpretive manner and the items on the developmental schedules should be considered as so many tools for sharpening perception (Gesell, 1925, pp. 408-409).

Gesell, in fact, recommended supplementing the schedules with play observation, behavior records, and parental interviews. In addition, his tests were devised for young infants as well as the preschool child, leading him to define intellectual development in a broad fashion as consisting of motor, adaptive, language, and personal/social components.

"Motor development" refers to children's ability to control gross body movements and finer motor coordination, an index of the maturity of the neuromuscular system; it is identified in tasks such as reaching for and grasping objects as well as manipulating toys. "Adaptive behavior" relates to children's problem solving ability, eye-hand coordination, and ability to remove obstacles; identification procedures include, for example, cube manipulation as well as addition and subtraction. "Language development" generally refers to children's ability to communicate verbally; it is evaluated by observing the child's obedience to directions.
and naming of colors. "Personal-social" behavior refers directly to a child's reaction to social training; tasks to assess this intellectual aspect include observations of the child's ability to dress and feed himself as well as play cooperatively. Thus Gesell's theory is clearly oriented towards early identification and provides a basis for screening procedures. His cautions in the use of the developmental schedules and flexible definition of intellectual growth, moreover, suggest excellent guidelines for current diagnostic programs.

A second theory of development having important implications for early identification and screening practices is that of Jean Piaget (1950, 1952). He defines intelligence broadly as "the superior forms of organization or equilibrium of cognitive structurings" (1950, p. 7). These cognitive structures refer to "operations," actions performed by the child which are internalized and reversible (Piaget, 1950, p. 32). An example would be addition which has the reverse operation of subtraction. Piaget divides the child's growth into developmental stages: the "sensorimotor stage" (0 to 18 months); the "preoperational stage" (18 months to 7 years); the "concrete operations stage" (7 to 12 years); and finally, the "formal operations stage" (age 12 and up). During each stage children develop processes and functions which help in their overall adaptation to the environment. Examples are "assimilation" which, in the simplest sense,
refers to children's use of their present capabilities to handle new problems and "accommodation" which refers to children's changing in order to manage more difficult situations (Piaget, 1950, p. 7). Piaget and other investigators (Fogleman, 1970) have developed tests to ascertain a child's level of development within the various stages. His theory, therefore, provides a base for identification and screening procedures by pinpointing the developing child's strengths and weaknesses within the adaptation process.

A typical Piagetian test is used to assess a child's grasp of the operation of "conservation," the notion that liquids and solids can be transformed in shape without changing their quantitative features (Piaget, 1950, p. 140). Two identical glasses are filled to equal heights with liquid; the child generally agrees that the amount of liquid in the two containers is equal. If the liquid in one of the containers is then poured into another which is differently shaped, perhaps taller and narrower, a child at the preoperational level will state that the new container has more or less liquid. He thus illustrates the lack of reversibility in his "schemas" or cognitive structures which refer to a class of action sequences having a strong interrelationship (Piaget, 1952, p. 210). A child who has reached the concrete operations stage, on the other hand, will acknowledge the conservation of liquid in spite of the change in container shape. In a sense, it may seem that identification and
screening tests are not appropriate for the preoperational or preschool child since, by definition, he has not mastered many of the skills needed to perform intellectual operations. However, Piaget asserts that much of the basis of future intellectual growth is formed during this period, especially intuitive understanding and a gradually more "decentered" approach to problem solving (1950, p. 122). Thus Piagetian tests have been devised to assess cognitive and perceptual growth during this period, and Piaget offers a strong theoretical rationale for such identification procedures.

A third theory providing a strong basis for early identification and screening is the organismic-developmental approach of Heinz Werner (1957, 1963). Werner does not differentiate the child's growth into relatively discrete stages, but feels developmental change follows the "orthogenetic principle." In other words, development occurs in the direction of an increasing differentiation of the components of symbolizations and of increasing integrative systematization (autonomization) of symbolic forms (Werner, 1963, p. 40).

Thus Werner, following in the Gestalt tradition, is interested in global problems of development which he relates to biological processes. He concentrates directly on the child's mental functioning which moves from a state of relative globality and undifferentiatedness towards states of increasing differentiation and hierarchic integration (Werner, 1963, p. 7).
Moreover, this movement takes place in accordance with environmental demands.

Perhaps the best way to illustrate Werner's conceptualization of the developmental process and to show its applicability to screening procedures is to examine the acquisition of word meaning (Werner & Kaplan, 1952). Werner has devised a series of tests to determine children's proficiency in various kinds of word usage. Essentially, the tests involve asking a child to discover the meaning of a nonsense word from its usage in a series of twelve sentences. Since language development occurs "microgenetically," beginning with an undifferentiated perception and ending in consensually validated organized verbal forms, Werner assesses various levels of competence in children's use of language. These levels include: "rigidity of meaning" or inability to revise meanings; "concrete symbolism" or not choosing meanings on the basis of similarity of sound; "word sentence fusion and holophrasis" or the ability to recognize a word's uniqueness apart from the rest of a sentence; "integration of word meanings" or fitting different definitions to a single word; and finally, "lability of word usage" or the ability to distinguish between various functions a word may have in the same situation (Werner & Kaplan, 1952). Such a group of tests, especially when applied to Werner's overall theoretical framework, clearly provides the basis for a diagnostic approach to identifying language
problems in the preschool and early school years.

In addition to identifying processes of word acquisition, Werner's theory provides a basis for testing children's general level of intellectual development in terms of "syncretic" thinking. Werner explains "syncretic phenomena" as follows:

If several mental functions or phenomena, which would appear as distinct from each other in a mature state of consciousness, are merged without differentiation into one activity or into one phenomenon, we may speak of a syncretic function or a syncretic phenomenon (Werner, 1957, p. 53).

The opposite pole of syncretic is "discrete," involving an ability to treat the objects of experience as separate entities. Werner believed it was possible to identify a child's level of development on this broad psychological dimension through examination of his drawings of particular objects. The drawings were analyzed for "diffuseness" versus "articulation," referring essentially to the vagueness or clarity of the forms (Werner, 1957, p. 159). By assessing children's mental processes in this fashion, Werner again provided a potential basis for screening procedures. In fact, his ideas on evolving "differentiation" in children's perceptions provided an important basis for cognitive tests involving embedded figures as well as the rod and frame test (Baldwin, 1967).

A final theoretical conceptualization which is relevant to screening tests in general and to the SEST in particular
is T. Ernest Newland's "product-process" skills model (1972). "Product" refers to skills involving primarily memory and little or no conceptual or reasoning ability; these skills are extremely dependent on a child's prior learning experience. Examples of product skills are counting and reciting the alphabet. The SEST assesses product skills by having a child print his name, copy designs, and draw a person as well as by having him take several vocabulary and arithmetic tests. "Process," on the other hand, refers to children's developing ability to conceptualize items as being similar or different. For example, being able to classify a horse, a dog, and a cat as animals indicates process development. The SEST assesses process development through visual matching, auditory discrimination, opposites/analogies, identification of missing elements, and several other tests. Newland believes that process skills represent the basic psychological operations which enable children to learn and acquire information. In other words, "process begets product" and the two broad areas are viewed on a continuum (Newland, 1972). Since no task is purely product or process, Newland speaks of "product-dominant" and "process-dominant" items. As the child develops beyond the age of five, the most important process skills begin to require more and more prerequisite product skills such as recognizing letters, reading, and knowing vocabulary words. Thus, process and product skills are very interdependent and continuously interact in children's
intellectual development. Further information on how the product-process conceptualization relates to the SEST is provided in Appendix A.

To summarize, four theories have been presented which have in common the attempt to identify potential learning problems and assets in children about to enter kindergarten and first grade. Gesell presents a maturational perspective by outlining groups of tasks children should be able to perform within specific age limits given normal physiological development. His theory emphasizes the child's biological growth in developing intellectual awareness. Piaget, on the other hand, believes that the child's experience plays a central role in intellectual development as he tries to adapt to the environment; thus Piagetian tests for the preschool child assess experiential growth and resulting intuitive understanding of problems. Piaget and Gesell are similar, however, in outlining stages of development even though their criteria for mastery within these stages are quite different. Werner differs from Piaget and Gesell primarily by emphasizing increasing differentiation of perception as the child ages; he does not divide children's intellectual growth into relatively separate stages, but asserts that growth follows this broad principle of differentiation and continued integration. Werner is similar to Gesell, however, in emphasizing biological processes as central to intellectual growth. His preschool tests assess the child's overall intellectual develop-
ment on the syncretic-discrete dimension, referring to his ability to differentiate the objects of experience. Newland, the last theorist considered, conceptualizes growth on the product-process continuum. Product simply refers to skills involving primarily memory and little reasoning ability, while process relates to children's ability to note similarities, differences, and to solve problems through acquisition of information. Process development sounds similar to Werner's syncretic thinking, both involving increasing differentiation and classification of perceptions. Finally, the SEST is an example of a test using Newland's product-process model to assess the preschool child's intellectual development.

Battery Approaches

One of the most common approaches to early identification and screening in recent years has been the use of a test battery. For the purposes of this study a battery approach will be considered any attempt to screen learning problems based on the results of three or more separate tests. Generally these tests assess different aspects of cognitive-intellectual development and their results are often combined in an overall "predictive index." Authors using a battery approach usually estimate the relative contributions of each test to the battery's predictive power through multiple regression techniques. The primary advan-
tage of the battery may be the amount of comprehensive information obtained on each child, while the main disad­

One of the first studies to employ a battery approach to prediction was undertaken by Deputy (1930). His sample consisted of 165 entering first graders and his battery included the Pinter-Cunningham Primary Mental Test along with four other tests of cognitive and perceptual functions. He found the Pinter was the single best predictor of reading achievement at the end of first grade, accounting for 62 percent of the variance in prediction. However, using the other four tests raised the predictive power of the mental test to 75 percent. Deputy concluded that "intelligence" was the most significant factor in predicting reading success in first grade, but advocated using other measures to supplement intelligence test scores.

One of the best known contemporary studies using a battery approach was that of Jansky, de Hirsch, and Langford (1966). They created a predictive screening index consisting of 37 perceptual-motor and cognitive tests which was used to predict reading achievement at the end of second grade for a sample of 53 kindergarten children. Intellectual functions tested included: "language comprehension," "auditory-motor integration," and "word recognition" among other skills. In a later follow-up investigation
the authors found their index predicted reading failure correctly in 83 percent of the children in their original sample (Jansky & de Hirsch, 1972).

A somewhat later study (Dudek, Goldber, Lester, & Harris, 1969) used an unusual battery to predict general academic achievement in first and second grades. The battery consisted of the Wechsler Intelligence Scale for Children, the Lorge-Thorndike Group Intelligence Scale, the Lincoln-Oseretzky Motor Development Scale, and the Goodenough-Harris Draw-a-Man Test; it was administered to 103 middle-class kindergarten children in two separate schools. In addition, "personality adequacy" was rated by a trained child psychiatrist in a home interview with parent and child. While the intelligence tests (WISC and Lorge-Thorndike) were the single best predictors, each accounting for about 60 percent of the variance, adding the results of the perceptual-motor tests increased the predictive power of the entire battery to about 75 percent to 85 percent of the variance. Personality ratings, however, did not significantly improve the battery's overall predictive power.

Lessler, Schoeninger, and Bridges (1970) used a battery consisting of the Lee-Clark Readiness Test, the Bender Gestalt Test, and the Peabody Picture Vocabulary Test to predict reading achievement and teachers' ratings of overall ability at the end of first grade. The battery was adminis-
tered to 216 children in three separate schools representing a range of socioeconomic backgrounds. The authors found the Lee-Clark was the single best predictor of poor ratings on both criteria in 73 percent to 89 percent of the sample, regardless of subjects' socioeconomic level. Moreover, neither the Peabody nor the Bender improved the predictive power of the Lee-Clark alone. The authors speculated that the results may have been related to the Lee-Clark's having several sub-tests measuring vocabulary and perceptual-motor accuracy.

A similar study (Ferinden & Jacobson, 1970) used a battery consisting of the Wide Range Achievement Test (WRAT), the Bender Gestalt Test, the Evanston Early Identification Scale (EEIS), and the Metropolitan Reading Test to predict reading difficulties in 67 kindergarten children. Like Lessler et al. the authors found that the Bender contributed little to predictive accuracy, while the Metropolitan Reading Test was an effective predictor only if total test scores fell below the 30th percentile. The WRAT and the EEIS were extremely good predictors, accurately screening 97 percent of children who experienced problems in reading at first grade level one year later. Telegdy (1975) also found the Bender to be an ineffective predictor of reading and overall academic achievement in his group of 56 children who were administered a similar battery. Miller (1972), however, found that the Bender was a very good predictor of first grade handwriting in his group of 55 kindergarten children.
and stated that it should be part of any screening battery.

A much more extensive investigation than most of those described above was undertaken by Satz and Friel (1972) who were concerned with prediction of reading achievement. Their sample consisted of 497 white male kindergarten pupils who were administered a battery consisting of 14 motor and intellectual tests; in addition, several other predictive measures including age, socioeconomic, and activity levels were considered. The authors factor analyzed their results and found that tests of sensory, perceptual-motor, and mnemonic abilities accounted for the highest percentage of the variance in prediction (30.7); teachers' evaluations and socioeconomic factors accounted for 16 percent, verbal and conceptual tests for 13.4 percent, and measures of handedness and motor skills for only 7.7 percent. The authors concluded that their battery could identify several "predictive antecedents" of later learning disability, and advocated "linear multivariate models" in early prediction. A follow-up study (Satz & Friel, 1974) using the same group presented similar results and conclusions.

A more recent study (Eaves, Kendall, & Crichton, 1974) had success using a perceptual-motor screening battery with 228 kindergarten children, but warned that the predictive index did not carry any specific educational implications. The authors advocated using the battery as a screening device, but felt children identified as "high risk" should be
given individualized diagnostic testing to determine needed remediation. Book (1974), on the other hand, stated that his battery, which was composed of the Slosson Intelligence Test, the Bender Gestalt Test, and the Metropolitan Readiness Tests, provided predictive accuracy as well as classification for program placement for his sample of 725 kindergarten children.

Finally, a recent study (Wallbrown, Wallbrown, Engin, & Blaha, 1975) substantially similar to that of Jansky, de Hirsch, and Langford (1966) and Jansky and de Hirsch (1972) used ten separate tests to predict first grade achievement. The authors found the IQ score obtained from the Slosson Intelligence Test and Bender Gestalt results accounted for 49 percent of the variance in predicting reading achievement in their 120 kindergarten subjects. The authors suggested including a test measuring general intelligence and tests of visual-motor integration in future batteries.

Intelligence and Achievement Tests

In a sense, any standardized intelligence or achievement test is a predictive device and often may be used for identification and screening. Only a few researchers, however, have systematically studied the utility of these tests for such purposes. The major advantages of using these tests lie in their often extensive standardization and attempts to establish overall validity and reliability, both
tending to be more satisfactory than with most screening devices created specifically for prediction. In addition, one does not have to expend the time and effort to create a screening device if a standardized test is employed. Using these tests for identification and screening, however, has numerous drawbacks. First, many of these tests must be administered and interpreted by professional personnel. Second, they are often more time consuming than other screening devices. Third, the scores obtained tend to provide broad measures of abilities, making specific remedial recommendations difficult; and, finally, few of these tests provide specific remedial programs keyed to measured deficits.

Edwards and Kirby (1964) performed an early predictive validity study of the Lorge-Thorndike Intelligence Test (LTIT, 1957). The authors correlated the scores of 336 entering first graders on the LTIT with scores on the SRA Achievement Series administered when the subjects entered grade 3. None of the predictive validity coefficients exceeded .50 and many were around .43; the composite correlation between LTIT IQ and SRA total scores was .50. The authors believed these correlations did not justify using the LTIT to predict achievement of first graders, but admitted that their study had several methodological problems: first, sample selection excluded school failures and obviously gifted children; second, there was a two-year lapse in mea-
surement between LTIT and SRA administration; and finally, it was difficult to compare the "non-verbal" LTIT with the highly verbal SRA series.

Indeed, a second predictive validity study of the LTIT (Mendels, 1973) suggested that the test was valid for predicting academic achievement. Mendels administered the test to 118 kindergarten children representing a range of abilities based on scores on the California Test of Mental Maturity, and correlated their scores with three Metropolitan Achievement Test subtests given in grade 1. Predictive validity correlations between .46 and .62 for the LTIT and the three subtests were obtained. Adding the demographic variables of sex, age, and father's occupation did not significantly increase these correlations. Mendels (1973) concluded that the LTIT could be used alone as a valid instrument for assessing the intellectual abilities of kindergarten children.

Vingoe et al. (1969) advised using an abbreviated version of the Wechsler Preschool and Primary Scale of Intelligence for screening, but they did not present any validity data. Scott (1965) reported on the predictive validity of the Detroit Beginning First-Grade Intelligence Examination using the Stanford Achievement Test (SAT) as the criterion measure for a group of 905 entering first-graders. The author found a wide range in scores on the Intelligence Examination and on the SAT administered in grade 2.8 for the group; however, significant positive correlations between
Intelligence Examination scores and scores on all SAT sub-tests were also found. In addition, the author correlated the scores of 15 children who had scored average or above average on the Intelligence Examination and SAT with the following criterion measures administered in grade 4.8: the Stanford-Binet Intelligence Scale, the Wechsler Intelligence Scale for Children, and the California Test of Mental Maturity. For 14 of the 15 children correlations between the intelligence and achievement test scores and the criterion measures were significant, suggesting some constancy in intellectual growth throughout the elementary grades. The authors concluded, nevertheless, that school success cannot be predicted solely from intelligence tests without considering other relevant variables.

Tests Developed Specifically for Early Identification

A variety of tests have been developed over the past two decades for the specific purpose of identification of children likely to have learning problems in the early school years. The SEST is a prime example of such a procedure. Many of these so-called "readiness" or "screening" tests incorporate items and methods from existing achievement and intelligence tests, while others present somewhat more experimental approaches. Moreover, many studies of these devices attempt to gain data on overall predictive validity and most useful test items by correlating screening
test scores with later criterion measures of academic achievement. The primary advantage of these screening tests is their efficient administration by nonprofessional personnel, while their major disadvantages are often inadequate standardization and lack of information on overall validity and reliability (Rogolosky, 1969).

In an illustrative study Hopkins and Sitkei (1969) attempted to determine the efficacy of intelligence versus reading readiness tests in predicting first grade achievement in their sample of 157 lower middle-class children entering first grade. The authors used the Lee-Clark Reading Readiness Test (LCRR; revised edition, 1962) and the California Test of Mental Maturity (CTMM) as predictors; scores on the Lee-Clark Reading Test Primer and end-of-year teacher ratings were used as criterion measures. The major findings of the study were: 1) the LCRR attained an overall significant predictive validity coefficient of .61 for both criteria; 2) adding CTMM IQ to reading readiness scores raised the multiple correlation coefficient to .67; 3) adding father's occupation, sex, and age at time of initial testing failed to significantly increase predictive accuracy, raising the coefficient to .68. The authors suggested using the readiness instead of the intelligence test in predicting reading achievement, citing savings in time, expense, and more meaningful interpretation for remedial purposes.

Pate and Webb (1970) studied the predictive validity of
the First Grade Screening Test (FGST, 1966), a simple perceptual-motor group test, by correlating the scores of 205 entering first graders with teachers' ratings and Stanford Achievement Test (SAT, 1964) scores at the end of third grade. The authors found that an FGST cutting score of 19 correctly identified 84 percent of children who failed in the first three years of school according to the two criterion measures; moreover, SAT achievement scores for those above the cutting scores were on the average 1.1 grade levels higher than for those below the cutting score. The authors concluded that the FGST can predict success and failure during the three primary grades with some accuracy. A similar study (Seitz, Johnson, & Kenney, 1973) employing a perceptual-motor screening procedure named The Johnson-Kenney Screening Test came to comparable conclusions regarding their device. It was administered to 171 entering first graders and correlated significantly with teachers' ratings at the end of grade 1.

A more comprehensive device than any presented above was studied by Tebiessen, Duckworth, and Conrad (1971). The authors wished to determine the efficacy of the Schenectady Kindergarten Rating Scales (SKRS) for predicting overall achievement and adjustment in approximately 300 entering kindergarten children divided into seven diagnostic categories on the basis of SKRS profiles. The SKRS consists of 13 scales rating a variety of behavioral, cognitive, and
motor areas. While SKRS profile ratings were made by the authors, criterion measures included teachers' ratings of intelligence and adjustment at the end of first grade. Overall, the SKRS predicted teachers' diagnoses accurately for 63 percent of 152 boys and 73 percent of 142 girls with classifiable SKRS profiles. When specific diagnoses were discarded and SKRS profiles and teachers' first grade diagnoses categorized simply as problem or no problem, the predictive accuracy of the SKRS was 79 percent for the boys and 83 percent for the girls. Although these overall figures are impressive, the predictive value of the scales was much higher for certain diagnoses, especially those involving impulse control, hyperactivity, and language skills.

Smith and Solanto (1971) described a screening test they had developed which drew its items from standardized intelligence tests:

To assess Vocabulary Skills, both the picture vocabulary and the first eight words on the Stanford-Binet Intelligence Scale were employed. Numerical Skills were assessed by means of the Arithmetic subtest of the WPPSI. Visual-motor Skills and Immediate Recall Ability were assessed by means of the Block Design and Sentence subtests of the WPPSI, respectively. New Learning Ability was assessed by means of the Animal House on the WPPSI. Intelligence Level was determined by overall performance on the above objective measures (Smith & Solanto, 1971, p. 143).

Unfortunately, the authors did not include any data on the validity or reliability of their instrument. However, the validity of the CIRCUS series, a group of screening tests
measuring similar abilities in a comparable manner was strongly questioned by Raths and Katz (1975).

McKnab and Fine (1972) assessed the predictive validity of the Vane Kindergarten Test (Vane, 1968) by correlating the scores of 168 kindergarten children with scores on the Stanford Achievement Test (SAT) administered one year later. The Vane consists of three subtests: "Man" requiring the child to draw a man; "Perceptual-Motor" requiring the child to copy a square, cross, and hexagon; and "Vocabulary" which asks the child to define 11 words presented orally. The authors found the Vane correlated .52 with total achievement as measured by the SAT, accounting for about 25 percent of the variance in prediction. The sample used was above average in ability as measured by the Vane and the SAT in addition to being older than the Vane's standardization sample (72.1 months versus 65.6 months). The authors concluded, however, that the Vane was generally not useful for individual assessment, but might be more effective if administered in a preschool setting or very early in the school year. A follow-up study (Powers, 1974) found significant correlations between the Vane and scores on the SAT and Metropolitan Readiness Tests in separate groups of preschool and kindergarten children. However, the author concluded that the correlations remained too low to permit the Vane to be used in individual decisions, either at the kindergarten or preschool levels.

Kapelis (1975) examined the validity of the Meeting
School Screening Test (MSST) and the Prereading Screening Procedures (PRSP) for predicting the reading achievement of 110 entering first graders. In addition, the author compared the predictive power of the screening devices with teachers' forecasts of reading achievement at the end of first grade. The MSST consists of three subtests named "Motor Patterning," "Visual-Perceptual-Motor," and "Language," while the PRSP is composed of seven subtests measuring a variety of visual-motor and auditory skills. The criterion measures consisted of three reading subtests from the Metropolitan Achievement Tests (MAT) administered at the end of the school year. While all three predictors produced significant correlations with the MAT subjects, the PRSP was most powerful, producing correlations between .66 and .68 with MAT scores. Teachers' forecasts were the weakest predictors, producing correlations between .46 and .48. The authors went a step further in using multiple regressions to find the best composite predictor: the PRSP and Language subtest of the MSST correlating .77 with overall MAT reading achievement.

Miscellaneous Identification and Screening Procedures

This section will briefly cover devices and procedures which are not primarily oriented toward direct psychometric assessment of intellectual ability, but which have been used in screening and early identification. Perhaps foremost
among such devices is the Bender Gestalt Test which has been extensively studied by Koppitz (1963, 1973, 1975). Although it is often used clinically to assess neural impairment and brain damage, Koppitz has found the Bender to be a highly effective device for predicting academic achievement, especially if administered in the first three grades. The author recognizes, however, that the Bender does not measure many factors needed for school success and recommends it as part of a screening battery supplemented by other brief tests (Koppitz, 1975). Auxter (1971) describes such a battery which was used successfully to screen learning disabilities in a sample of 18 preschool children. It has also been found that below average Bender scores are generally poorer predictors of academic achievement than average or above average scores (Koppitz, 1963, 1975). Keogh and Smith (1970) found the Bender correlated highly with teachers' evaluations and subsequent school achievement; however, their study suggested the Bender was more accurate for identification of high potential than high risk children.

Gross (1970) used the school records of 43 kindergarten children to predict achievement in first and second grades. The measures which correlated highly with SAT achievement included educational level of the home, Stanford-Binet IQ, and Gross' own five-point personality assessment derived from recorded impressions of "drive, social maturity, independence, stability, and adjustment" (Gross, 1970, p. 278).
Wilborn and Smith (1974) suggested going even further back in the child's records to obtain data on "perinatal and developmental" events which may be used as learning problem indicators. The authors conducted a validity study on the Learning Problem Indication Index (LPII), essentially a parental checklist of problems during pregnancy and the child's early infancy, and found significant correlations between LPII measures and later incidence of learning disabilities in a sample of 432 children of diverse socio-economic backgrounds. An earlier study (Klanderman & Stone, 1973), however, suggested parental checklists tend to be inaccurate and need to be supplemented by professional judgment.

A final group of studies deals with teachers' ratings of children for the purpose of early identification. Haring and Ridgway (1967) conducted an extensive investigation in which kindergarten teachers screened over 1200 children on the basis of gross muscle coordination, verbal fluency, speech development, auditory memory, auditory discrimination, visual memory, visual discrimination, visual-motor performance, directionality, and laterality. While subsequent intellectual testing suggested teachers' judgments were highly accurate in choosing children with learning problems, few common identifiable "learning patterns" were found in the children selected. Another study (Atwell, Orpet, & Meyers, 1967) suggested that teachers were generally accurate
in predicting achievement in a sample of 100 kindergarten children on the basis of behavioral observations of cooperation, motor-speed dexterity, hyperactivity, and confidence in various situations. A more recent study (Fesbach, Adelman, & Fuller, 1974) compared the Jansky and de Hirsch Predictive Index, WPPSI IQ, and teachers' ratings based on the Student Rating Scale (SRS), a composite measure of the child's cognitive and social functioning, for predicting reading achievement in 32 kindergarten classes. Using a stepwise multiple regression analysis, the authors found the SRS was the single best predictor of first grade reading achievement, while the Jansky and de Hirsch battery and WPPSI ranked second and third. All measures, however, correlated significantly with later reading achievement. Finally, a recent review of programs geared toward early intervention (Schaer & Krump, 1976) suggested teachers' observations were the most important and trustworthy components in early identification and diagnosis.

The purpose of the present study was to examine the predictive validity of the Smile Early Screening Test (SEST; revised edition, 1975). To accomplish this goal the scores of two groups of children on the SEST were correlated with their scores on the Metropolitan Readiness Tests (MRT; revised edition, 1965). One group of children had scored at or above the mean on their final SEST Total,
while the second group had scored below the mean. For the convenience of the reader the three major hypotheses of the present investigation are briefly restated below:

1) For both groups of children correlations between their Total MRT scores and their scores on three overall SEST measures would be significant. These overall SEST measures were Total, Learning Aptitude, and Achievement.

2) The Language Arts score on the MRT which includes the composite of the Word Meaning, Listening, Matching, and Alphabet subtests would correlate significantly with SEST Achievement and SEST Learning Aptitude for both groups.

3) Alphabet on the MRT would correlate highly with Alphabet on the SEST for both groups; in addition, Arithmetic on the MRT would correlate highly with Arithmetic on the SEST for both groups. Finally, Copying on the MRT would correlate highly with Copying on the SEST for both groups.
CHAPTER II

METHOD

Subjects

Subjects (Ss) used for the present study were selected from the population of kindergarten students enrolled in public schools in Missoula and Anaconda, Montana. The subject population chosen sampled the full range of scores (from lowest to highest) based on the normative data of the Smile Early Screening Test. The children were divided into two groups based on their test performance. One group included 33 children, randomly chosen from four Missoula public schools, who had scored at or above the mean on their final test total. The second group included 29 children, randomly chosen from four Anaconda public schools, who had been found below the mean on their total test score; these students were considered "high risk" based on test performance. The two groups of children did not receive any special educational treatment based on their test performance. The second group had to be selected from Anaconda rather than Missoula public schools because all Missoula school children scoring below average on the SEST receive special remedial treatment. Such remediation would certainly have biased the results of the present inves-
tigation. The Missoula group contained 17 boys and 16 girls; the Anaconda group contained 21 boys and 8 girls. Data on father's and mother's occupation was gathered for 29 children in the Missoula group and 29 children in the Anaconda group. This information was used to classify each group socioeconomically on the basis of Warner's revised scale of occupational ratings (Warner, Meeker, & Eells, 1960, pp. 131-132).

The socioeconomic data for both groups is presented in Table 1. Warner's scale is divided into seven occupational categories based on community social value, business size, salary, and training requirements. Category 1 represents the highest status which includes professional, managerial, and business employment; Category 7 refers to the lowest status which includes heavy and unskilled labor. Intermediate categories represent various gradations of occupational status based on the above criteria. For a more detailed description of each occupational category refer to Warner et al. (1960). The average occupational rating for Missoula was 3.89 and for Anaconda 4.70; the difference between the means was non-significant, $t_{(57)} = 2.046, p > .05$. The average occupational rating for the entire sample was 4.31, indicating that the subject population is predominantly middle to lower middle class.
## TABLE 1

**SOCIOECONOMIC STATUS BASED ON PARENTS' OCCUPATION**

**FOR MISSOULA AND ANACONDA SAMPLES**

<table>
<thead>
<tr>
<th>Warner's Occupational Category</th>
<th>Number in Each Category for Missoula Sample</th>
<th>Number in Each Category for Anaconda Sample</th>
<th>Total Number in Each Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1:</strong> Professionals and large business proprietors</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><strong>Category 2:</strong> Semi-professionals and large business officials</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Category 3:</strong> Clerks and secretaries; minor business officials</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td><strong>Category 4:</strong> Skilled workers; sales personnel</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td><strong>Category 5:</strong> Small business proprietors and workers</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td><strong>Category 6:</strong> Semi-skilled workers</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td><strong>Category 7:</strong> Unskilled workers and laborers</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Socioeconomic Status $\bar{X}$ for Missoula Sample = 3.89  
Socioeconomic Status $\bar{X}$ for Anaconda Sample = 4.70  
Socioeconomic Status $\bar{X}$ for Total Sample = 4.31
Procedure

Children were taken from class and administered the Metropolitan Readiness Tests in groups of six. The MRT requires approximately 10 to 20 minutes to administer. The author served as the sole examiner for all subjects.

The scores of the entire sample on the six MRT subtests including Word Meaning, Listening, Matching, Alphabet, Numbers, and Copying as well as their Total and Language Arts scores were correlated with the following overall scores on the Smile Early Screening Test: Achievement; Learning Aptitude; and Total. In addition, correlations were made with scores on four specific Smile Early subtests including Arithmetic, Copying, Alphabet, and Memory for a total of 56 correlations between the MRT and SEST.
CHAPTER III

RESULTS

Upon completion of the testing, the MRT performance of each child from the Missoula and Anaconda samples was scored according to standard scoring procedures in the MRT manual. Pearson product-moment correlations were calculated between all SEST and MRT subtest scores and each test's total score using the FORTRAN LNLZQ4.F4 program on a DEC-PDP 10 computer. These correlations are presented in Table 2. The correlations between each individual SEST subtest and SEST Total score include that particular subtest within the Total score.

Correlations related directly to the hypotheses of the present investigation will be presented first. Hypothesis 1 predicted that for the entire sample correlations between Total MRT scores and scores on three overall SEST measures would be significant. The three SEST measures were Total, Learning Aptitude, and Achievement. The following correlations were obtained: .79 between MRT Total and SEST Total; .68 between MRT Total and SEST Learning Aptitude; .74 between MRT Total and SEST Achievement, df = 61, p < .001. Thus, Hypothesis 1 was confirmed. Hypothesis 2 predicted that for the entire group the composite MRT Language Arts
### TABLE 2
SUMMARY OF CORRELATIONS BETWEEN SMILE EARLY SCREENING TEST AND METROPOLITAN READINESS TESTS SUBTEST AND TOTAL SCORES (N=62)

<table>
<thead>
<tr>
<th></th>
<th>Metropolitan Readiness Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word Meaning</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>.41*</td>
</tr>
<tr>
<td>Copying</td>
<td>.28*</td>
</tr>
<tr>
<td>Alphabet</td>
<td>.11</td>
</tr>
<tr>
<td>Memory</td>
<td>.37*</td>
</tr>
<tr>
<td>Learning Aptitude</td>
<td>.56*</td>
</tr>
<tr>
<td>Achievement</td>
<td>.42*</td>
</tr>
<tr>
<td>TOTAL</td>
<td>.55*</td>
</tr>
</tbody>
</table>

*significant correlation, p < .05
score would correlate significantly with SEST Achievement and SEST Learning Aptitude. Results showed correlations of .59 between MRT Language Arts and SEST Achievement and .61 between MRT Language Arts and SEST Learning Aptitude, \( df = 61, p < .001 \). Thus, Hypothesis 2 was also confirmed.

Finally, Hypothesis 3 predicted significant correlations between MRT Alphabet and SEST Alphabet, between MRT Numbers and SEST Arithmetic, and between MRT Copying and SEST Copying for the entire group. The correlations obtained were as follows: .52 between MRT Alphabet and SEST Alphabet; .66 between MRT Numbers and SEST Arithmetic; and .53 between MRT Copying and SEST Copying, \( df = 61, p < .001 \). Based on these results, Hypothesis 3 was also confirmed. A number of other interesting findings emerged from examination of the intertest matrix and are considered in the discussion.

A summary of within-test correlations for the Smile Early Screening Test is presented in Table 3. Important correlations include those between Achievement and Learning Aptitude (.75), between Learning Aptitude and Total (.88), and between Achievement and Total (.93), \( df = 61, p < .001 \). In addition, SEST Copying correlated almost equally with Achievement (.65) and Learning Aptitude (.62), \( df = 61, p < .001 \). This unexpected result along with selected additional within-test correlations will be considered in the discussion.
### TABLE 3
SUMMARY OF WITHIN-TEST CORRELATIONS OF THE SMILE EARLY SCREENING TEST (N=62)

<table>
<thead>
<tr>
<th>Test</th>
<th>Arithmetic</th>
<th>Copying</th>
<th>Alphabet</th>
<th>Memory</th>
<th>Learning Aptitude</th>
<th>Achievement</th>
<th>Total</th>
<th>Total Achieve-</th>
<th>Learning Aptitude</th>
<th>Memory</th>
<th>Alphabet</th>
<th>Copying</th>
<th>Arithmetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copying</td>
<td></td>
<td>1.00</td>
<td>0.55*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alphabet</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.19</td>
<td>0.46*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.28* 0.39* 0.52*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Aptitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00 0.38* 0.30* 0.62* 0.63*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00 0.75* 0.57* 0.68* 0.65* 0.87*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00 0.93* 0.88* 0.69* 0.51* 0.67* 0.81*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant correlation, p<.05.*
CHAPTER IV

DISCUSSION

The results of the present investigation indicate that the SEST correlates highly with the MRT when the latter is used as a criterion measure of achievement. Particularly notable is the correlation between the Total scores on both tests which emerged as the highest in the matrix and represents approximately 62 percent of the variance in prediction, \( r(61) = .79, p < .001 \). The magnitude of this correlation suggests a substantial overlap between the two tests in predicting academic achievement; however, this trend in the data was expected since both tests' total scores represent gross measures of the child's general intellectual capability. The main point is that the highly significant correlation with MRT Total provides support for the SEST as an overall screening measure. In addition, correlations between total MRT scores and the two central SEST summary scores, Achievement (.74) and Learning Aptitude (.68), emerged as the second and fourth highest correlations in the matrix, \( df = 61, p < .001 \). Moreover, correlations between the MRT Language Arts score and the two SEST summary scores, .61 for Achievement and .59 for Learning Aptitude, were also highly
significant, $df = 61, p < .001$. All of the above correlations refer to overall measures of ability and indicate that the SEST can predict achievement in these broad cognitive-intellectual types of achievement.

Besides predicting achievement in this global fashion, the SEST can apparently predict achievement of more specific types as revealed in the correlations between specific SEST and MRT subtests shown in Table 2. The first row of the table shows correlations between SEST Arithmetic and all MRT subtests. Note that SEST Arithmetic correlated .66 with MRT Numbers, while correlations between SEST Arithmetic and other MRT subtests were consistently lower than .66, ranging from .27 to .49. These results indicate that the SEST Arithmetic subtest does predict achievement in numerical skills. Although the Arithmetic subtest correlates significantly with theoretically unrelated areas, it is a somewhat weaker predictor of these types of achievement. The second row of Table 2 shows correlations between SEST Copying and the MRT subtests. The correlation between SEST Copying and MRT Copying is .53, $df = 61, p < .001$; however, SEST Copying showed less relationship to other MRT subtests with correlations ranging from .12 to .43. These results indicate that SEST Copying also predicts achievement in its target area while showing lower correlations with theoretically unrelated areas. The third row of Table 2 shows correlations between SEST Alphabet and all MRT subtests. The correlation between SEST Alphabet and MRT
Alphabet is .52, df = 61, p < .001. As with the previous SEST subtests, SEST Alphabet appears to predict achievement in the target area while being a somewhat weaker predictor of achievement in hypothetically unrelated areas. Correlations of SEST Alphabet with the remaining MRT subtests ranged from .11 to .41.

The specific subtest correlations referred to above are generally smaller than those for the total and summary scores for both tests. This trend in the data seems reasonable since SEST and MRT subtests measuring similar abilities often do so in quite different ways. For example, the SEST Arithmetic subtest requires children to answer numerical questions orally, to count aloud, and to identify orally numbers written on a sheet of paper. The MRT Numbers subtest only requires the child to place written marks on the correct answers to numerical problems and to identify numbers in the same way. Thus the smaller correlations are probably, in part, the result of the different methods each test uses to measure comparable abilities.

Several of the subtest correlations were significant at the .05 level and even at higher levels, even though they were smaller than the total and summary score correlations. Moreover, several of the significant subtest correlations were between skills in which a significant overlap was not necessarily expected; an example is the correlation between SEST Copying and MRT Numbers, r (61) = .43, p < .001. This
trend in the data indicates that none of the SEST subtests predicts achievement solely within its own intended areas without overlapping with theoretically unrelated areas. While this trend certainly indicates the need for reappraisal of these subtests, it may not indicate the need for massive subtest revision. Instead, the significant overlap between subtests measuring hypothetically unrelated skills may simply be the result of a general intellectual factor needed for effective performance on several of these subtests. If such were the case, the SEST might be an example of a test illustrating Spearman's two-factor theory of intelligence (1932). A general or "g" factor might be prerequisite for efficient performance on many subtests, while a variety of "s" or specific factors might be important in performance on individual subtests. The only way to isolate such factors would be to perform a factor analysis on the SEST which is one of the recommendations of the current investigation.

The main point for the predictive validity of the SEST, however, is that the magnitude of the correlations for three SEST subtests with comparable MRT subtests were all higher than the other subtest correlations in their respective rows. Although there certainly was significant overlap between theoretically unrelated subtest scores so that the subtests need reexamination, the overall trends in the data indicate that discrete patterns of ability formation can be measured by the SEST.
Although no specific predictions were made regarding correlations of SEST Memory with other subtests, it is interesting to note the more restricted range of correlations for this subtest area. Correlations of SEST Memory with specific MRT subtests are generally low to moderate, ranging from .29 to .50. Since memory is a wide-ranging ability playing a role in many types of intellectual performance, this trend in the data is not completely unexpected. However, the correlation between SEST Memory and MRT Listening which similarly emphasizes short-term memory functions, is somewhat lower than the correlations between specific MRT and SEST subtest areas discussed above, $r (61) = .42$, $p < .001$. A possible explanation for the smaller overlap may again be that the subtests are measuring a similar ability—short-term memory—in a quite distinct manner. The MRT emphasizes auditory memory, while the SEST assesses auditory and visual memory functions in conjunction with the child's ability to place items in sequences. Thus MRT Listening primarily tests the child's auditory comprehension of passages read aloud, while SEST Memory places greater emphasis on "pure" memory functions. For example, the SEST requires the child to recognize shapes and sequences of objects after brief exposure as well as to repeat words and numbers after the examiner. Thus the SEST more directly assesses children's ability to recall, while the MRT assesses ability to recall as well as understand passages.
Only four of the fifty-six correlations presented in Table 2 did not attain statistical significance at the .05 level. The four correlations included: SEST Alphabet x MRT Word Meaning (.11); SEST Alphabet x MRT Matching (.19); SEST Alphabet x MRT Copying (.21); and finally, MRT Alphabet x SEST Copying (.12), df = 61. Moreover, correlations between MRT Alphabet and SEST Learning Aptitude (.27; p < .05) and between MRT Alphabet and SEST Total (.40; p < .005) were among the lowest of any specific MRT subtest with these two SEST summary scores, df = 61. The non-significant and relatively small correlations presented may be partially explained by the highly specific nature of the visual and auditory skills involved in learning the alphabet. This particular type of rote memory performance probably does not overlap with many other basic cognitive-intellectual abilities. In Newland's (1972) terminology learning to say the alphabet is a highly "product" dominant skill very dependent upon the child's prior learning experience. In addition, interviews with teachers suggested wide variation in emphasis on learning alphabet in various classrooms which may also account for the lower correlation. SEST assessment of alphabet requires the child to recite all 26 letters and to visually identify 8 of them; MRT assessment requires the child to visually identify 16 letters. Thus a significant overlap between the two alphabet measures was expected and found \( r (71) = .52, p < .001 \). However, in view of the limited
range of skills involved in task performance and the differential emphasis on learning the alphabet in various kindergarten classrooms, the non-significant correlations presented seem reasonable. Because alphabet appears to overlap very little with other skills and, therefore, performance on SEST Alphabet may be more related to classroom emphasis than to basic ability, the non-significant correlations suggest eliminating this subtest in future SEST revisions. Besides shortening the SEST, eliminating the Alphabet subtest would probably increase the accuracy of the SEST for screening decisions by eliminating an apparently irrelevant measure of academic potential.

Other interesting correlations in Table 2 include those between MRT Numbers and SEST Achievement and between MRT Numbers and SEST Total, $r_{(61)} = .67$ and $.70$ respectively, $p < .001$. The magnitude of these correlations suggests numerical skills may be another central predictor of academic achievement. Indeed, all correlations between MRT Numbers and SEST scores are significant at the .001 level except for the correlation with SEST Alphabet which is significant at the .01. Comparing the size of the SEST Arithmetic correlations with the size of the MRT Numbers correlations leads to a similar conclusion concerning the predictive value of mathematical skills. The correlation between SEST Arithmetic and MRT Total is the highest of any specific subtest with total MRT scores, $r_{(61)} = .64$, $p < .001$. Moreover, all other
correlations between SEST Arithmetic and MRT subtests are significant at the .001 level except for correlations with MRT Listening and MRT Matching, significant at the .02 and .05 levels respectively.

Table 3 presents a summary of within-test correlations for the SEST. Particularly notable is the correlation between Achievement and Learning Aptitude which suggests a strong overlap between the two summary scores, \( r(61) = .75, p < .001 \). Moreover, this correlation is remarkably close to the correlation between Achievement and Learning Aptitude for the SEST normative population, which was .74 (Cook & Rudio, 1976). Both these correlations, accounting for 56 percent and 55 percent of the variance respectively, provide support for Newland's assertion that product-process skills reside on a continuum. These data are also consistent with the total score correlation between the MRT and the Pinter-Cunningham Primary Mental Ability Test (.74); the MRT is a product-oriented test while the Pinter-Cunningham presents tasks more relevant to process skills. Finally, Table 3 also shows Learning Aptitude and Achievement correlating almost equally and highly with Total. The correlation between Learning Aptitude and Total is .88, while the correlation between Achievement and Total is .93, \( df = 61, p < .001 \). These results are expected based on SEST content. The test is designed to assess the child's competence in the "product" dominated Achievement and "process" dominated Learning
Aptitude areas on a more or less equal basis.

An important consideration in developing a screening device is to assess how well the individual subtests merge together to form a meaningful composite measure of a child's intellectual ability. Each subtest should contribute uniquely to this composite measure as indicated by the magnitude of the intercorrelations between the subtests. The authors of the MRT, for example, found moderate positive intercorrelations among the six MRT subtests and thus concluded that each contributed uniquely to their composite readiness measure (Hildreth et al., 1965). In general, the intercorrelations among SEST subtests shown in Table 3 are low to moderate, ranging from .19 to .55. None of the intercorrelations is large enough to suggest that any two of the subtests are measuring identical or nearly identical functions. Thus it appears that each subtest measure contributes in a unique fashion to the overall screening test score. Before leaving the intertest correlations, it should be noted that the correlation between Alphabet and Copying was the only non-significant correlation in the within-test matrix, again pointing to the highly specialized nature of the alphabet skill and the possibility of shortening the SEST by eliminating this test, \( r (61) = .19, p > .05 \).

Table 3 shows higher within-test correlations for Achievement than for Learning Aptitude. The Achievement correlations apart from that with SEST Total range from .87
for Arithmetic to .57 for Memory; the Learning Aptitude correlations apart from that with Total and with Achievement range from .30 for Alphabet to .63 for Arithmetic. This trend in the data is expected since the subtests considered are Achievement subtests on the SEST. Thus the three major subtest areas of Memory, Alphabet, and Arithmetic are product-oriented and should correlate more highly with Achievement than with Learning Aptitude. These correlations should, nevertheless, be viewed with caution because of the continuous nature of the product-process conceptualization. Memory, for example, appears to be fairly close to the middle of the continuum, correlating .38 with Learning Aptitude and .57 with Achievement, \( df = 61, p < .005 \). These correlations make sense if memory is viewed primarily as a product skill involving little conceptual or reasoning ability. An example of this type of memory would be simply remembering the words of a song. In addition, memory may be viewed as partially a process function since it underlies many conceptual skills and abilities to abstract. An example would be distinguishing between "b" and "d" based on visual memory. The product-oriented nature of memory accounts for the large correlation with Achievement, while the process characteristics may account for the smaller but still sizable correlation with Learning Aptitude.

Of the four subtests shown on Table 3 only SEST Copying correlated almost equally with Achievement (.65) and
Learning Aptitude (.62), df = 61, p < .001. The other three subtest areas showed higher correlations with Achievement than with Learning Aptitude as discussed above. The almost identical correlations were unexpected since Copying, like the other three subtests, is viewed as a product-oriented skill on the SEST. There are two possible explanations for this unexpected result. First, the manner in which the child's ability to copy is assessed on the SEST may be more process-oriented than the authors intended. In other words, the Copying subtest may be only minimally based on the prior experience of the child and may require revision. The Copying subtest currently requires the child to duplicate 12 forms of increasing complexity. The forms might be made simpler or more comparable to objects the child encounters in daily life. Admittedly such revision would be quite difficult since most of the forms are already very simple. To make them comparable to common objects would probably make these basically simple forms quite complex. A second explanation is that the ability to copy may be close to the middle of the product-process continuum. Based on this reasoning, it may not be appropriate to classify copying as primarily a product skill based mainly on the child's previous experience. This latter explanation seems plausible in view of the visual-motor coordination skills which copying requires. These skills may be an important part of a child's evolving style of learning, and therefore at least partially representative.
of process development.

There are strong correlations between the SEST and the MRT for individual subtest, summary, and total scores. These correlations provide evidence that the SEST can predict overall academic achievement as well as achievement in specific subareas. Significant correlations were, moreover, obtained in an intellectually diverse group of children and between tests measuring abilities in a quite different manner. These factors strengthen conclusions supporting the predictive validity of the SEST. In addition to providing data on the SEST's predictive validity, the present research offered evidence supporting the theoretical validity of the product-process distinction, especially when particular skills are viewed on a continuum representing these two broad intellectual areas. Thus certain SEST subtest skills such as Memory and Copying appear to lie close to the middle of the continuum, while others such as Arithmetic and Alphabet are more clearly product-oriented. Of course, the significant overlap between the product and process areas illustrated in the current investigation and referred to by Newland (1972) indicates the need for further research to clarify the limitations within which these concepts can be applied to screening test development.

In general, the present study supports using the SEST for screening and remediation decisions. The SEST has a number of advantages over readiness tests such as the MRT
including: a lower age range of children with which it can be effectively employed; a more precise treatment program for measured deficits; a more comprehensive assessment of the child's ability in particular subareas; a greater chance for relevant behavioral observations on the child since it is individually administered; and finally, the impressive validity and reliability data which have been gathered on this relatively new test (see Appendix A). However, the SEST does take considerably longer to administer than the MRT and cannot be administered by one examiner in groups. It is therefore recommended that the MRT and teachers' observations be used as initial screening devices to identify children with potential learning problems. The SEST can then be administered to those children who appear to need help, thus providing more precise prescriptive data for treatment if specific learning problems are identified. This procedure should result in more accurate screening decisions, more effective treatment strategies, and great savings of time and effort for teachers.

Further studies of the SEST are needed to provide more data on the test's overall screening power as well as constructive modifications in test format. In particular, a factor analysis of the SEST is recommended to clearly establish the central factors in a correlational matrix of the SEST with other tests which measure abilities hypothesized to be important in SEST performance. Being able to identify
these factors would have important implications not only for SEST development and revision, but also for theories of early intellectual development. For example, it would be interesting to know whether Spearman's (1932) two-factor theory of intelligence is applicable at the preschool and early school levels. In addition, further validity studies on the SEST are clearly needed. These studies could make predictions about correlations between SEST subtests and thus provide more information on the construct validity of the SEST. For example, predictions could be made about correlations between particular SEST subtests based on their hypothesized position on the product-process continuum. These correlations were dealt with only post hoc in the present study. To further illustrate, a researcher might predict a significant correlation between the process-oriented Visual Matching and Opposites/Analogies subtests or between the product-oriented Alphabet and Arithmetic subtests while predicting non-significant correlations between unrelated subtests. Besides providing data on overall construct validity, such research could establish whether discrete abilities are measured in the Learning Aptitude subtests as they appear to be in the Achievement subtests. Finally, further SEST studies such as those described might establish more clearly where conceptually difficult areas such as Copying and Memory reside on the product-process continuum.
CHAPTER V

SUMMARY

The predictive validity of the Smile Early Screening Test (SEST; revised edition, 1975) was investigated by correlating the above and below average scores of two groups of kindergarten children (N = 62) on this test with their scores on the Metropolitan Readiness Tests (MRT; revised edition, 1965). Results indicated high predictive validity for the SEST: 52 out of 56 overall and individual subtest correlations were significant at least at the .05 level including a correlation of .79 between the total scores on both tests. The data indicated that the SEST can predict academic achievement in the broad "product" and "process" cognitive-intellectual areas hypothesized by Newland (1972) as well as in the following specific subareas: Arithmetic, Copying, Alphabet, and Memory. There was, however, significant overlap between the SEST and the MRT for theoretically unrelated subtest areas. Moreover, correlations between the alphabet subtests on both the SEST and the MRT and other subtests were the only non-significant correlations obtained, leading the author to suggest elimination of the Alphabet subtest in future SEST revisions.
Within-test correlations for the SEST showed a strong overlap between the product-oriented Achievement and process-oriented Learning Aptitude summary scores as predicted by Newland's (1972) theory, $r (61) = .75, p < .001$. In general, intercorrelations among SEST subtests were low to moderate, suggesting that each subtest measure contributes uniquely to the overall screening test score. Although higher within-test correlations were obtained for Achievement than for Learning Aptitude as expected, the SEST Copying subtest unexpectedly correlated about equally with both scores. The author suggested either revising the Copying subtest or reclassifying it as partially a process and primarily a product skill.

Suggestions for further research on the SEST were offered including the need for a factor analysis. In general, the study supported using the SEST as a precise screening device and in remedial decisions but suggested using the MRT and teachers' observations as initial screening devices.
REFERENCES

Arthur, G. An application of intelligence tests to the problem of school retardation. School and Society, 1919, 15, 89.


Brooks, F. Chronological age as a factor in reading achievement. School and Society, 1924, 20, 82.


Deputy, E. C. Predicting First Grade Reading Achievement. New York: Columbia University Teachers College, 1930.


Gesell, A. What Can the Teacher Do for the Deficient Child? Hartford, Conn.: State Board of Education, 1918.


Gross, M. B. Pre-school prediction of academic achievement. Psychological Reports, 1970, 26, 278.


Telegdy, G. A. The effectiveness of four readiness tests as predictors of first grade academic development. *Psychology in the Schools*, 1975, 12, 4-11.


Wallbrown, J. D., Wallbrown, F. H., Engin, A. W. and Blaha, J. The prediction of first grade reading achievement with selected perceptual-cognitive tests. Psychology in the Schools, 1975, 12(2), 140-149.


APPENDIX A

SMILE EARLY SCREENING TEST
Smile Early Screening Test

The predictive validity of the Smile Early Screening Test (revised edition, 1975) was the subject of the present investigation. The SEST consists of 24 subtests which are grouped into three broad areas: Achievement, Learning Aptitude, and Memory. These broad areas are further broken down into specific subareas as shown on the Smile Early Screening Test Record Form (see Appendix B). The subtests under Achievement are closely related to the child's prior learning experiences including tests of vocabulary, alphabet, arithmetic, and being able to print one's name. The subtests related to Learning Aptitude are considerably less related to prior experiences, although the test authors readily admit that this effect can never be completely eliminated. The Learning Aptitude subtests sample tasks such as identifying likenesses, discriminating differences, noting similarities, detecting patterns, and classifying pictures; these tasks are hypothesized to sample the major building blocks of learning (Newland, 1972). Finally, the Memory subtests are closely related to the child's ability to pay attention and acquire new information.

The SEST is designed to 1) identify high risk students in kindergarten and first grade, and 2) to provide teachers with information useful in planning an appropriate educa-
tional program for a particular child. The 1975 revision of the SEST was standardized on a sample of 96 kindergarten students and 70 first grade students. The kindergarten sample included 19 students who were referred by their teacher for screening purposes and 77 non-referred students, which approximated the percentages for referred and non-referred students in School District #1 in Missoula, Montana. The first grade norm group included only non-referred students, since referral decisions had not been made for this group at the time of standardization. Appendix C presents summaries of normative information for both the kindergarten and first grade samples.

The reliability of the SEST was established through the Kuder-Richardson method using the standardization samples. Reliability coefficients of .86 were obtained for Total Achievement and .91 for Total Learning Aptitude using entering kindergarten students (Cook & Rudio, 1976). The reliability coefficients for the entering first graders were .69 for Total Achievement and .65 for Total Learning Aptitude (Cook & Rudio, 1976). The authors attributed the somewhat depressed correlations for the first grade population to two factors. First, the norm group for first grade students may have been too homogeneous since it contained no referred students. Second, the test tended to be too easy for first grade students as can be seen in the mean scores for the normative data in Appendix C. Several of these first grade
scores are close to the maximum as a result of revising the SEST for entering kindergarten students. The SEST is currently being revised to provide more difficult items for first grade students. The Total Memory reliability coefficient was .77 for the two samples combined (Cook & Rudio, 1976). The highest individual subtest reliability coefficient was .86 for Alphabet using the kindergarten sample, while the lowest was .34 for Arithmetic using the first graders (Cook & Rudio, 1976). Appendix D presents the remaining individual subtest reliability coefficients. In general, the data provide very good support for SEST reliability.

The authors of the SEST have attempted to maximize the overall validity of their test. Content validity was established by carefully analyzing the demands of the kindergarten and first grade curricula. Individual subtests were then developed which would adequately assess the child's skills in language arts, arithmetic, and also his/her ability to use a pencil. These tests, which became part of the Achievement area, were designed to assess the child's previous learning experience. In addition, the SEST's authors wished to sample the child's ability to learn. The tests to assess this Learning Aptitude were based on the five process skills identified by T. Ernest Newland (1972): discrimination of differences; identification of likenesses; determination of sequential progress; identification of analogies; and
identification of missing elements. Newland (1972) bases his conceptualization on his work with the Blind Learning Aptitude Test. Finally, the authors assessed the child's ability to listen, to pay attention, and to comprehend auditory and visual information in five Memory tests which constituted the third major area of the SEST.

The construct validity of the SEST was assessed through five separate procedures using a sample of 50 kindergarten students. Half of these students were referred children who had scored below 180 on the Smile Early Screening Test; an additional 25 students were selected who scored in the average range or higher on the SEST. In the first procedure, correlations were made between the SEST and the Columbia Mental Maturity Scale (CMMS), a recognized test of the child's ability to see relationships and to discriminate conceptual differences. A correlation of .77 was obtained between CMMS Total and SEST Learning Aptitude, while the correlation between the CMMS and SEST Conceptual development area was .75 (Cook & Rudio, 1976). In the second procedure, correlations were made between the SEST and the Peabody Picture Vocabulary Test (PPVT); the PPVT samples the child's achievement in vocabulary development as opposed to discriminating conceptual relationships. A correlation of .70 was obtained between the PPVT and SEST Achievement, while the correlation between the PPVT and SEST Vocabulary was .75 (Cook & Rudio, 1976). The third construct validity procedure used only the referred
population (N = 24) for correlations between the SEST and the CMMS. A correlation of .77 was obtained between CMMS Total and SEST Learning Aptitude, while the CMMS x SEST Conceptual Development correlation was .68 (Cook & Rudio, 1976). The fourth procedure again utilized the referred population (N = 20) for correlations between the SEST and the PPVT. The correlation between PPVT Total and SEST Achievement was .52, while the correlation between PPVT Total and SEST Vocabulary was .57 for the referred group (Cook & Rudio, 1976). The final construct validity procedure looked at correlations between the Achievement and Learning Aptitude subtest areas of the SEST for both the normative and the referred kindergarten populations. The Learning Aptitude x Achievement correlation was .74 for the normative population and .59 for the referred population (Cook & Rudio, 1976).

In general, the SEST does correlate highly with important criterion variables related to the construct validity of the screening test. Thus, highly significant correlations were obtained between the CMMS, a recognized test of learning aptitude, and SEST Learning Aptitude and SEST Conceptual Development for referred as well as normal populations. Moreover, significant correlations were obtained between the PPVT, a known test of vocabulary achievement, and SEST Achievement and SEST Vocabulary for the same populations. However, the correlations between the PPVT and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
SEST Achievement and Vocabulary for the referred population are somewhat depressed (.52 and .57 respectively). These correlations may be a result of the generally less consistent intellectual performance of this group as shown in the fifth procedure involving correlations between Total Learning Aptitude and Total Achievement for each group. There was a depressed correlation between Learning Aptitude and Achievement for the referred population (.59) and a higher correlation for the normative population (.74). The SEST's authors predicted a high relationship between these two broad intellectual areas for a normal population; however, for children with learning disabilities or behavioral problems, it was expected that the relationship between Achievement and Learning Aptitude would be much lower. This prediction was proved correct in the fifth procedure, and the Learning Aptitude x Achievement correlations for the two groups provide further support for the construct validity of the SEST.

The overall results of research up to the present show impressive reliability and validity coefficients for the SEST. The figures are more impressive for a test which is just over two years old and has already undergone one major revision. The present predictive validity study was part of an on-going research project to evaluate the effectiveness of the SEST for screening and remediation decisions. The SEST will continue to undergo revision to make it a more
reliable and efficient screening instrument. The current edition of the SEST is nicely laid out in a red looseleaf folder with concise and easily understood directions for the administrator and child. The artwork is particularly attractive and most teachers report it very effective in capturing the child's attention. Scoring is almost entirely objective on a one point per correct item basis with a maximum possible score of 250. Perhaps the major disadvantage of the SEST is that it takes 45 minutes to an hour to administer in individual sessions. Future revisions will aim to reduce this time to administer somewhat without sacrificing the comprehensive assessment of ability which the current edition so well provides.
APPENDIX B

SMILE EARLY SCREENING TEST RECORD FORM
SMILE EARLY SCREENING TEST RECORD FORM

Name: ___________________________ Age: ___ Sex: ___ School: ________ Grade ___
Address: ___________________________ Parent's Name: _____________________ Examiner: ___________

Date Tested                      Yr.       Mo.       Day
Date of Birth                    ————      ————      ————
Age                              ————      ————      ————

Summary Scores
Achievement Total
Learning Aptitude Total
Memory Total
Entire Test Total

SUBTEST SCORES

Achievement
Physical Development
1. Draw-a-person (14) ________
2. Printing Name (2) ________
3. Copying designs (12) ________
TOTAL (28) ________

Academic Skills
16. Vocabulary (34) ________
18. Arithmetic (25) ________
19. Alphabet (13) ________
TOTAL (72) ________
ACHIEVEMENT TOTAL (100) ________

Learning Aptitude
Perceptual Development
4. Visual Matching (10) ________
6. Visual Discrimination (5) ________
8. Auditory Discrimination (10) ________
TOTAL (25) ________

Sequencing Skills
5. Match Sequences (5) ________
23. Picture Arrangement (5) ________
24. Detecting Patterns (10) ________
TOTAL (20) ________

Conceptual Development
7. Discrimination (5) ________
20. Opposites/ Analogies (10) ________
21. Similarities (5) ________
22. Classification (10) ________
TOTAL (30) ________

Reasoning Skills
9. Visual Absurdities (10) ________
10. Cause-effect (5) ________
11. Critical Thinking (10) ________
TOTAL (25) ________

LEARNING APTITUDE TOTAL (100) ________
APPENDIX C

SMILE EARLY SCREENING TEST NORMATIVE DATA
### SUMMARY OF NORMATIVE INFORMATION - KINDERGARTEN

**SMILE EARLY SCREENING TEST**

#### Summary Scores - Mean S.D.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement Total</td>
<td>75.65</td>
<td>11.37</td>
</tr>
<tr>
<td>Learning Aptitude Total</td>
<td>80.64</td>
<td>11.98</td>
</tr>
<tr>
<td>Memory Total</td>
<td>34.01</td>
<td>6.32</td>
</tr>
<tr>
<td>ENTIRE TEST TOTAL</td>
<td>190.52</td>
<td>26.50</td>
</tr>
</tbody>
</table>

#### Sub Test Scores

<table>
<thead>
<tr>
<th>Sub Test</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Achievement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Draw-a-person (14)*</td>
<td>8.40</td>
<td>2.37</td>
</tr>
<tr>
<td>2. Printing Name (2)</td>
<td>1.67</td>
<td>0.66</td>
</tr>
<tr>
<td>3. Copying designs (12)</td>
<td>6.31</td>
<td>2.58</td>
</tr>
<tr>
<td>TOTAL (28)</td>
<td>16.38</td>
<td>4.42</td>
</tr>
<tr>
<td>Academic Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Vocabulary (34)</td>
<td>29.64</td>
<td>3.21</td>
</tr>
<tr>
<td>18. Arithmetic (25)</td>
<td>21.89</td>
<td>3.53</td>
</tr>
<tr>
<td>19. Alphabet (13)</td>
<td>7.75</td>
<td>4.00</td>
</tr>
<tr>
<td>TOTAL (72)</td>
<td>59.27</td>
<td>8.29</td>
</tr>
<tr>
<td>ACHIEVEMENT TOTAL (100)</td>
<td>75.65</td>
<td>11.37</td>
</tr>
<tr>
<td><strong>Learning Aptitude</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Visual Matching (10)</td>
<td>9.44</td>
<td>0.98</td>
</tr>
<tr>
<td>6. Visual Discrimination (5)</td>
<td>4.45</td>
<td>0.85</td>
</tr>
<tr>
<td>8. Auditory Discrimination (10)</td>
<td>9.08</td>
<td>1.28</td>
</tr>
<tr>
<td>TOTAL (25)</td>
<td>22.97</td>
<td>2.14</td>
</tr>
<tr>
<td>Sequencing Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Match Sequences (5)</td>
<td>3.93</td>
<td>1.06</td>
</tr>
<tr>
<td>23. Picture Arrangement (5)</td>
<td>3.12</td>
<td>1.67</td>
</tr>
<tr>
<td>24. Detecting Patterns (10)</td>
<td>6.96</td>
<td>2.65</td>
</tr>
<tr>
<td>TOTAL (20)</td>
<td>14.00</td>
<td>4.17</td>
</tr>
<tr>
<td>Conceptual Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Discrimination (5)</td>
<td>3.14</td>
<td>1.43</td>
</tr>
<tr>
<td>20. Opposites/Analogies (10)</td>
<td>8.44</td>
<td>1.78</td>
</tr>
<tr>
<td>21. Similarities (5)</td>
<td>4.22</td>
<td>1.36</td>
</tr>
<tr>
<td>22. Classification (10)</td>
<td>8.77</td>
<td>1.43</td>
</tr>
<tr>
<td>TOTAL (30)</td>
<td>24.56</td>
<td>4.76</td>
</tr>
<tr>
<td>Reasoning Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Visual Absurdities (10)</td>
<td>8.45</td>
<td>1.53</td>
</tr>
<tr>
<td>10. Cause-effect (5)</td>
<td>3.98</td>
<td>0.94</td>
</tr>
<tr>
<td>11. Critical Thinking (10)</td>
<td>6.68</td>
<td>1.76</td>
</tr>
<tr>
<td>TOTAL (25)</td>
<td>19.10</td>
<td>3.16</td>
</tr>
<tr>
<td>LEARNING APTITUDE TOTAL (100)</td>
<td>80.64</td>
<td>11.98</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### SUMMARY OF NORMATIVE INFORMATION - FIRST GRADE

#### SMILE EARLY SCREENING TEST

<table>
<thead>
<tr>
<th>Summary Scores</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement Total</td>
<td>89.73</td>
<td>5.31</td>
</tr>
<tr>
<td>Learning Aptitude Total</td>
<td>91.61</td>
<td>4.35</td>
</tr>
<tr>
<td>Memory Total</td>
<td>40.53</td>
<td>5.32</td>
</tr>
<tr>
<td>ENTIRE TEST TOTAL</td>
<td>221.91</td>
<td>10.76</td>
</tr>
</tbody>
</table>

#### SUB TEST SCORES

<table>
<thead>
<tr>
<th>Achievement</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Draw-a-person (14)</td>
<td>10.06</td>
<td>1.88</td>
</tr>
<tr>
<td>2. Printing Name (2)</td>
<td>1.97</td>
<td>.16</td>
</tr>
<tr>
<td>3. Copying Designs (12)</td>
<td>9.37</td>
<td>1.71</td>
</tr>
<tr>
<td><strong>TOTAL (28)</strong></td>
<td>21.40</td>
<td>2.85</td>
</tr>
<tr>
<td><strong>Academic Skills</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Vocabulary (34)</td>
<td>32.59</td>
<td>1.36</td>
</tr>
<tr>
<td>18. Arithmetic (25)</td>
<td>23.99</td>
<td>2.07</td>
</tr>
<tr>
<td>19. Alphabet (13)</td>
<td>11.76</td>
<td>2.75</td>
</tr>
<tr>
<td><strong>TOTAL (72)</strong></td>
<td>68.33</td>
<td>3.58</td>
</tr>
<tr>
<td><strong>ACHIEVEMENT TOTAL (100)</strong></td>
<td>89.73</td>
<td>5.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Aptitude</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceptual Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Visual Matching (10)</td>
<td>9.54</td>
<td>.94</td>
</tr>
<tr>
<td>6. Visual Discrimination (5)</td>
<td>4.59</td>
<td>.69</td>
</tr>
<tr>
<td>8. Auditory Discrimination (10)</td>
<td>9.50</td>
<td>.91</td>
</tr>
<tr>
<td><strong>TOTAL (25)</strong></td>
<td>23.63</td>
<td>1.77</td>
</tr>
<tr>
<td><strong>Sequencing Skills</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Match Sequences (5)</td>
<td>4.80</td>
<td>.82</td>
</tr>
<tr>
<td>23. Picture Arrangement (5)</td>
<td>4.80</td>
<td>.58</td>
</tr>
<tr>
<td>24. Detecting Patterns (10)</td>
<td>8.89</td>
<td>1.56</td>
</tr>
<tr>
<td><strong>TOTAL (20)</strong></td>
<td>18.49</td>
<td>1.72</td>
</tr>
<tr>
<td><strong>Conceptual Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Discrimination (5)</td>
<td>4.34</td>
<td>1.01</td>
</tr>
<tr>
<td>20. Opposites/Analogies (10)</td>
<td>9.66</td>
<td>.58</td>
</tr>
<tr>
<td>21. Similarities (5)</td>
<td>4.74</td>
<td>.73</td>
</tr>
<tr>
<td>22. Classification (10)</td>
<td>9.57</td>
<td>.73</td>
</tr>
<tr>
<td><strong>TOTAL (30)</strong></td>
<td>28.31</td>
<td>1.70</td>
</tr>
<tr>
<td><strong>Reasoning Skills</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Visual Absurdities (10)</td>
<td>9.36</td>
<td>.72</td>
</tr>
<tr>
<td>10. Cause-effect (5)</td>
<td>4.49</td>
<td>.63</td>
</tr>
<tr>
<td>11. Critical Thinking (10)</td>
<td>7.34</td>
<td>1.39</td>
</tr>
<tr>
<td><strong>TOTAL (25)</strong></td>
<td>21.19</td>
<td></td>
</tr>
<tr>
<td><strong>LEARNING APTITUDE TOTAL (100)</strong></td>
<td>91.61</td>
<td>4.35</td>
</tr>
</tbody>
</table>
APPENDIX D

SMILE EARLY SCREENING TEST RELIABILITY COEFFICIENTS
## Summary of Kuder-Richardson Reliability Coefficients

<table>
<thead>
<tr>
<th>Sub-test</th>
<th>Kindergarten</th>
<th>Kindergarten and First Grade (combined)</th>
<th>First Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Copying Designs</td>
<td>.80</td>
<td></td>
<td>.57</td>
</tr>
<tr>
<td>2. Vocabulary</td>
<td>.68</td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>3. Arithmetic</td>
<td>.85</td>
<td></td>
<td>.34</td>
</tr>
<tr>
<td>4. Alphabet</td>
<td>.76</td>
<td></td>
<td>.78</td>
</tr>
<tr>
<td>5. Achievement Total</td>
<td>.86</td>
<td></td>
<td>.69</td>
</tr>
<tr>
<td>6. Visual Sequential Memory</td>
<td></td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>7. Auditory Sequential Memory</td>
<td></td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>8. Memory for Stories-Visual</td>
<td></td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>9. Memory for Stories-Auditory</td>
<td></td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>10. Auditory-Visual Memory</td>
<td></td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>11. Memory Total</td>
<td></td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>12. Perceptual Development</td>
<td>.64</td>
<td></td>
<td>.64</td>
</tr>
<tr>
<td>13. Sequencing Skills</td>
<td>.75</td>
<td></td>
<td>.39</td>
</tr>
<tr>
<td>14. Conceptual Skills</td>
<td>.85</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>15. Reasoning Skills</td>
<td>.68</td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>16. Learning Aptitude Total</td>
<td>.91</td>
<td></td>
<td>.65</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX E

METROPOLITAN READINESS TESTS
Metropolitan Readiness Tests

The Metropolitan Readiness Tests (MRT--Form A; revised edition, 1965) will be used as the criterion measure in the predictive validity study of the SEST. The MRT is, moreover, a screening device and will be briefly reviewed. The MRT consist of six subtests: "Word Meaning," a 16-item picture vocabulary test; "Listening," a 16-item test of ability to comprehend phrases and sentences instead of individual words; "Matching," a 14-item test of visual perception involving recognition of similarities; "Alphabet," a 16-item test of ability to recognize lower-case letters of the alphabet; "Numbers," a 26-item test of the child's knowledge of numbers; and "Copying," a 14-item test measuring visual perception and motor control (Hildreth, McGauvran, & Griffiths, 1965). The test is designed to provide a "quick, convenient, and dependable basis for early classification of kindergarten children and first-graders" (Hildreth et al., 1965, p. 2). Form A of the revised MRT was standardized on a sample of approximately 15,000 children in some 70 school systems during the fall of 1964. The standardization group represented all the major geographical regions of the United States and a broad range of socioeconomic backgrounds.

Reliability of the MRT was established through the split-halves and alternate forms methods in seven different
school systems. Split-halves reliability for Form A achieved correlations ranging from .90 to .95 in all seven school systems (Hildreth et al., 1965, p. 27). Alternate forms reliability utilizing Forms A and B of the MRT averaged correlations of .91 in the seven systems (Hildreth et al., 1965, p. 28). Alphabet and Matching subtests showed the highest reliability with median correlations of .88 and .82 respectively in all determinations; Word Meaning and Listening were least reliable with median values of .61 and .52 respectively (Hildreth et al., 1965, p. 28).

The authors of the MRT have devoted considerable attention to the validity of their test, discussing content, construct, and predictive validity in their manual. To maximize content validity the authors have drawn on research with earlier editions of the MRT as well as their own professional judgment in delineating characteristics most important for success in first grade work. Their list of characteristics includes: comprehension and use of language; visual perception and discrimination; auditory discrimination; richness of verbal concepts; general mental ability shown in capacity to infer and to reason; knowledge of numerical and quantitative relationships; sensory motor abilities; adequate attentiveness and ability to follow directions (Hildreth et al., 1965, p. 15). In addition, the authors present detailed explanations of how each subtest measures these characteristics as well as other skills needed in first grade work.
Data for construct validity has been obtained largely by correlating the MRT with other readiness and intelligence tests. In general, studies suggest the MRT correlates more highly with readiness tests such as the Murphy-Durrel Reading Readiness Analysis, the Pinter-Cunningham Primary Mental Ability Test, and Lee-Clark Reading Readiness Test than with tests measuring general intelligence. Correlations between the MRT and these readiness tests generally range between .70 and .85 as reported in the manual (pp. 16-17). However, the MRT has also been correlated with a number of intelligence tests including the Otis-Lennon Mental Ability Test, California Test of Mental Maturity, Van Alstyne Vocabulary Test, and the Stanford-Binet; overall correlations generally range between .50 and .70 for these general intelligence tests (Hildreth et al., 1965, pp. 16-17).

The manual presents a considerable amount of data on the predictive validity of both forms of the MRT as assessed by the Stanford Achievement Test (SAT). The sample employed in one study consisted of 9,497 first graders who were administered the MRT in October and the SAT in May. The range of correlations between the MRT and six SAT subtests measuring reading and arithmetic achievement was .57 to .67 (p. 18). However, there was a wide range of variability in May achievement for each category of October readiness.

Another group of studies reported in the manual dealt with the predictive validity of the MRT as assessed by the
Metropolitan Achievement Tests (MAT; 1959 revision). The correlations between the MRT and MAT subtests reported were again generally high in predicting overall achievement as well as achievement in specific areas such as reading and arithmetic. Generally correlations ranged between .58 and .81 for both forms in the studies presented in the manual (pp. 18-19). However, these studies used entering first graders measured over a one-year period. Other data presented suggested the predictive power of the MRT as assessed by MAT achievement drops if measured from end of grade 1 to end of grade 2, with correlations ranging between .43 and .60 (Hildreth et al., 1965, p. 23). Another illustrative study not in the manual (Mitchell, 1962) suggested the MRT was a good predictor of MAT achievement, and that no significant differences in test predictive validity existed between whites and blacks.

One other test which has been used to assess MRT predictive validity is the Gates Primary Word Recognition Test with total score correlations of .66 between both tests (Hildreth et al., 1965, p. 22). In addition, correlations between MRT scores and teachers' ratings in various subject areas for a group of 150 first graders average between .58 and .66 as reported in the manual (p. 22).

In general, the level of coefficients in all these studies strongly attest to the predictive validity of the MRT. Moreover, the data on reliability as well as construct
and content validity suggest it ranks high among readiness tests. The test itself appears well constructed with concise directions for administration and easily followed instructions for the child. Scoring is generally objective except for the Copying subtest which requires some subjective interpretation. For all these reasons the MRT appears to be an exceptionally well-tailored instrument for studying the predictive validity of the SEST.