

407290

Archives
Closed
LD
175
.A40K
Th
472

A Validity Study of the North Carolina
" Psychoeducational Screening Test

A Thesis

Presented to

The Faculty of the Department of Psychology
Appalachian State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by

Ralph W. Gorrell

May, 1978


LIBRARY
Appalachian State University
Boone, North Carolina


A Validity Study of the North Carolina
Psychoeducational Screening Test


by

Ralph W. Gorrell

Approved by:


Chairman, Thesis Committee


Professor of Psychology


Associate Professor of Psychology


Chairman, Department of Psychology

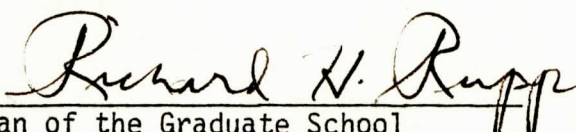

Dean of the Graduate School

Table of Contents

	<u>Page</u>
Acknowledgements.	iv
Abstract.	v
Introduction.	1
Method.	16
Results	20
Discussion.	27
Reference Notes	43
References.	44
Appendix A.	52
Appendix B.	63

Acknowledgements

I am greatly indebted to innumerable persons for their support and cooperation while I was completing my thesis. I would like to single out a few for special thanks.

Above all I want to thank my wife, Missie, who provided the encouragement and faith needed to finish this work.

My deep appreciation goes to the staff of the Boone, N.C., Developmental Evaluation Center, its director, Bill Elmore, and especially to the members of the screening team: Carolyn Hall, Sue Terrant, Saundi Mercier, Barbara Lawton, Linda Brandt, and Barbara Fish. Their support and assistance in the planning and data gathering made this project possible. In addition I am grateful to Jim Frazier, coordinator of the Statewide Prekindergarten Screening Program for his support of this independent research on a program he was deeply committed to.

I also wish to thank Hank Schneider and Bob McDonald for their help in the computer analysis of the data. Hank, in particular, provided guidance throughout this project.

Finally, my thanks go to Marty Byrd whose professional and experienced typing made this stage of the work incalculably easier.

Abstract

This paper reports the results of a concurrent validity study of the North Carolina Psychoeducational Screening Test (PET) using the McCarthy Scales of Children's Abilities and the Auditory Reception subtest of the Illinois Test of Psycholinguistic Abilities as criterion tests. Sixty-five 4-year-old children falling into three categories - delayed, average, and advanced development - were tested. The results are analyzed with traditional validity techniques of correlation and multiple regression which indicate that parts of the PET do not contribute to its overall validity although this overall validity is within acceptable limits (Pearson $r = .78$). The PET is also analyzed using screening test decision-making processes. This analysis shows that the PET has a very low under-referral rate (3.1 percent) but a high over-referral rate (35.3 percent). The predictive power of the PET is evaluated in light of the effects of varying prevalence rates. The PET is compared to several other screening tests. The PET was a part of the North Carolina Statewide Prekindergarten Screening Program (SPSP) for 4-year-old children. This program is briefly described and the implications of the results of this study for the SPSP are discussed. Suggestions for changes in the PET to improve its performance in screening are presented. The historical background and definitions of screening, along with issues which underscore the importance of evaluation of screening programs, are briefly discussed.

Introduction

Screening for developmental problems in children is on the rise in popularity. The purpose of this paper is to examine the validity of one particular screening instrument - the North Carolina Psychoeducational Screening Test (PET) - from a technical standpoint. In order to help the reader understand how this investigation fits into an overall evaluation of a screening program, a brief historical background of the development of screening is presented prior to the actual statement of the problem under investigation in this study.

The concept of mass screening is not new. It traces its roots back to public health efforts to control communicable and parasitical diseases, such as malaria, tuberculosis and syphilis. The U.S. Maritime Quarantine Service routinely screened several thousand newly-arrived immigrants in just a few hours in pre-World War I days (Smillie, 1952). It was natural to extend screening to chronic, non-communicable diseases, such as heart disease and cancer, as these diseases became the chief sources of mortality and morbidity in the more developed countries of the world. Included in the list of chronic diseases for which screening has been suggested is mental illness (Commission on Chronic Illness, 1957; Wilson and Jungner, 1968).

Pediatricians in the developed countries have also moved more towards dealing with chronic diseases rather than infectious diseases.

Frankenburg (1970) describes the changes in pediatrics as "shifts of focus from diagnosis and treatment of disease to earlier recognition and prevention of illness, or maintenance of health." Pediatricians have become more and more aware that various difficulties in intellectual, psychomotor, socio-emotional, and behavior development - all broadly covered under the label developmental difficulties - affect large numbers of children (Bakwin and Bakwin, 1972; Frankenburg and Dodds, 1967; Knobloch and Pasamanick, 1974; Knobloch, Pasamanick and Sherard, 1966; North, 1974b). Given the medical background of the concept of screening, it was logical that as pediatricians become more concerned with optimal child development they should adopt the screening concept for use in the early identification of childhood developmental problems.

The medical background of screening gives many definitions of screening an orientation towards detection of disease. Smillie (1952) defined screening as "an attempt, by mass methods, to utilize a series of clinical diagnostic procedures in early detection of chronic disease." The Commission on Chronic Illness (1957) decided screening was "the presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly." This last definition was adopted by the World Health Organization (Wilson and Jungner, 1968) and is still accepted in recent and authoritative medical literature (Frankenburg and Camp, 1975). Pediatricians have applied the disease-oriented definition of screening essentially unchanged to screening for developmental problems (Frankenburg and Camp, 1975).

Lessler (1972) reviews several definitions of screening. He suggests one which describes more completely the increase in scope and changes in emphasis which need to be made if screening is to move away from a strictly medical-disease orientation to a broader application in identifying developmental problems. This is important in screening for problems such as mental retardation, learning disabilities, and socio-emotional and behavioral disorders, which are less susceptible to analysis in the medical model. Accordingly, this investigator adopts his definition for use in this study:

Screening is the acquiring of preliminary information about characteristics which may be significant to the health, education, or well-being of the individual and which are relevant to his life tasks. The means of data collection must be appropriate and reasonable with regard to the economics of time, money, and resources for dealing with large numbers of persons (Lessler, 1972).

Essential characteristics of screening are its efficient, economical application to large population groups, and its tentativeness (the need for confirmation through further testing is essential before a final decision can be made).

Lessler (1972) also points out the difference between screening and a screening program. The former is merely the act of obtaining information through tests or other procedures with regard to identifying problems or deviations. The latter encompasses screening and, in addition, provides follow-up and treatment for those who have been identified as needing it.

The increasing popularity of developmental screening is attested to by the various recent symposia (Cutting, Haynes, Bird, Rubin, West, and Felch, 1970; North, 1974a) and government sponsored conferences (Meier, 1973; Oglesby and Sterling, 1970). Federal government involvement in developmental screening was begun through its pre-school educational program - Headstart. The Early and Periodic Screening, Diagnosis and Treatment Program (EPSDT) was the first complete expression of the Federal government's interest in developmental screening. Some type of developmental screening may also be incorporated into Federal efforts to provide a comprehensive health maintenance plan for all citizens.

The North Carolina State Government has also begun to show an interest in developmental screening. The legislature of the State of North Carolina established the Statewide Pre-Kindergarten Screening Program (SPSP) in 1975 to help children and their families while the children are in the formative stages of development. The goals of the SPSP include:

...profiling the health, psychoeducational, and social/emotional status of every child, based on screening information; initiating with the child's parents constructive action based upon the developmental plan; facilitating implementation of parental plans for benefitting their child; and monitoring with the parents the child's progress on the developmental plans (Human Resource Consultants, Note 3).

The North Carolina SPSP was designed to be applied to all four-year-olds in the state in two separate phases. The initial phase of screening consisted of: a parent questionnaire completed

through an interview with screening personnel; a brief hearing test using a screening audiometer; a vision test using the Snellen E or a picture/symbol chart; the Bean Bucket Game, designed to assess emotional status; and the North Carolina Psychoeducational Screening Test (PET).

The PET focuses on seven different areas of functioning: gross motor, visual memory, auditory memory, auditory perception, visual-motor integration, concept development, and language development. There are seven different scales on the PET, one for each area being evaluated. The PET is a screening, not diagnostic test, but it does attempt to identify both advances and delays in the child's development in all of the areas listed above except gross motor (where only delays are identified) (Division of Health Services, Note 1). Refer to Appendix A for a copy of the PET protocol.

The second phase of the screening procedure entailed several different options. A child who failed vision or hearing tests in Phase I was rescreened either by SPSP personnel or through referral to a local health department or physician. Socio-emotional or behavioral problems were referred to appropriate resources with brief counseling by SPSP personnel if needed to assure continuity of professional care. However, the largest activity in the second phase of screening was administration of the McCarthy Scales of Children's Abilities (MSCA) (McCarthy, 1972) to children who scored as advanced or delayed on the PET. Under SPSP procedures in

effect at that time, if two or more of the seven areas examined by the PET were scored as delayed or advanced (based on point scores totaled within each scale), the child was referred for individual psychological testing with the MSCA. The Division of Health Services (DHS) established cut-off points of 1.2 standard deviations above or below the mean on the MSCA for further referral, e.g., to a Developmental Evaluation Center, or intervention because of delayed or advanced development, respectively (DHS, Note 2).

As with most developmental tests, the PET was constructed by selecting items from several other previously standardized tests (Santa Clara Developmental Inventory, Stanford-Binet Intelligence Scale, Denver Developmental Screening Test, Developmental Test of Visual Motor Integration) (Zinn, Note 4). By combining and revising slightly test items which were indicative of normal four-year-olds' functioning in the original standardizations, the PET was formed. Although this is a good method of exploratory test construction, it is necessary to follow-up with thorough empirical studies of the new instrument in order to be certain it performs as expected.

A necessary, but not sufficient, condition for the success of a screening program is an accurate and efficient screening instrument. The screening instrument must be able to separate those who need to be referred for further testing (positive cases) from those who do not need follow-up (negative cases). This is a basic consideration pertaining to the screening procedure itself. A screening procedure which cannot successfully discriminate these two groups may waste limited resources following up unnecessary cases, or conversely may

fail to select those persons who are most in need of more detailed examination at the cost of their continued optimal functioning in society. Certainly determination of the acceptable social costs of either of these errors is a public policy decision rather than a scientific one. However, the information needed to make such a decision can only be found by scientific testing of screening procedures according to established standards for developmental screening tests. Other researchers (Frankenburg and Camp, 1975; Lessler, 1972; McKeown and Knox, 1968; Meier, 1973) have remarked on the general scarcity of information concerning the validity and other technical properties of the instruments used in developmental screening.

Pilot studies conducted on the PET were used to determine age norms, stability, inter-observer reliability, and validity (Zinn, Note 4). These studies, while necessary and valuable to the construction of a new instrument, need further replication and extension. Two areas in Zinn's studies which show the most need for further work are validity and age norms.

The remainder of this paper will examine the validity of the PET. First, I will briefly summarize Zinn's validity study (Note 4) of the PET, and point out reasons why further study is necessary. Then a new study of the PET's validity will be presented, followed by a discussion of some of the issues involved in validation of developmental screening instruments.

Zinn (Note 4) tested the concurrent validity of the language scale of the PET using two criterion tests, the Peabody Picture

Vocabulary Test (PPVT) and the Test of Basic Experiences (TOBE). The subjects in this study were 25 children whose background characteristics and scores on the Denver Developmental Screening Test or the Home Information Scale placed them in a high risk category for developmental delays in cognitive and language functioning. These children were enrolled in a federally-funded pre-kindergarten enrichment program designed to increase preschool readiness skills, especially language skills. The two criterion tests were given approximately three months prior to the PET. The results were evaluated using a 2 X 2 frequency table to separate the four possible outcome combinations (delay/normal) for each PET language scale - criterion test pair. A chi-square test applied to the frequency tables showed that PPVT-PET language scale correspondences in classification differed significantly from what would have occurred by chance. A similar test applied to TOBE-PET language scale correspondences showed that they did not differ significantly from what would be expected in a random distribution of test scores. Zinn concluded from his study that the PET language scale is valid. However, a more cautious conclusion about the PET's validity would seem to be called for.

Limiting examination to the results of the PET language scale validity, the evidence is inconclusive, since of the two criterion tests, only one (PPVT) produced a statistically significant relationship. Despite Zinn's assertion to the contrary one result which supported the PET language scale validity and another which

rejected it do not provide convincing evidence that the PET language scale is valid.

There are also several methodological problems which limit this study's usefulness in determining the validity of the PET. There was a relatively long interval of three months between initial testing with the criterion instruments and later administration of the PET. As Zinn acknowledged, the children were in a language enrichment program designed to accelerate cognitive, especially language, development during this interval. The change in levels of language development caused by the passage of time and by this intervention program may have reduced the degree of correspondence between the earlier and later test scores, obscuring a clear picture of the PET's validity.

Another important methodological factor to consider in evaluating Zinn's study is the situation in which the PET was validated. Rather than a wide variety of situations as recommended by the American Psychological Association (APA) in its manual on test standards (1974), only a relatively small number of children (25) with similar backgrounds (disadvantaged, high-risk) in one day care center were used as a validation sample. The fact that this sample is not representative of the entire range of children seen by the screening program reduces the confidence with which the validity data can be generalized to the entire population of four-year-olds. The small number of subjects and the restrictions placed on their selection should prompt us to use even greater caution in our

interpretation of the results (APA, 1974). In some cases the APA recommends collecting local validity data as a guide to clinically relevant referrals based on a test procedure (1974).

Perhaps the most important reason for further validation studies of the PET was that only one part of the PET was studied for its validity - the language development scale. No information on the validity of the other six scales was reported. As described above, a child was referred to Phase II of screening if he scored in the delayed (or advanced) range on two or more of the seven scales of the PET. Yet validity information is available on only one of the seven scales. The APA recommends as essential that "statements about validity should refer to the validity of particular interpretations or of particular types of decisions....Any study of test validity is pertinent to only a few of the possible uses of or inferences from the test scores" (1974). Thus the actual decision-making process used in SPSP has not been validated.

It must be pointed out that the PET has been frequently revised even while being used. Many of these changes were minor, but some, such as the addition of supplemental items to test for advanced development, were made after the studies (Zinn, Note 4) on the PET had been completed and appear to be significant. The APA recommends that after any changes in a test, new studies be carried out to revalidate the revised test (1974). Certainly the PET needs new validation studies after having a new section added to it.

Finally, it should be noted that although the chi-square statistic is a useful one in evaluating screening procedures (Thorner

and Remein, 1961), a more complete statistical treatment would help in determining the validity of the PET in comparison with other diagnostic and screening tests. Additional statistics, such as Pearson correlation coefficients, multiple regression procedures, sensitivity and specificity (Thorner and Remein, 1961), predictive efficiency (Meehl and Rosen, 1955), and the false positive and false negative rates, would provide more information and facilitate evaluation of the PET in relation to other instruments.

The appropriate choice of criterion instruments is essential if the data obtained in a validity study are to be meaningful. The APA recommends that the criterion be chosen with reference to the problem being studied (1974). Since the McCarthy Scales of Children's Abilities (MSCA) were used to examine children referred on the basis of screening by the PET, they offer the most relevant and practical criteria. The MSCA is a series of individually administered mental and motor scales designed to assess the developmental level of children aged 2 1/2 to 8 1/2 in a variety of areas. The MSCA provides standard scores on six scales: verbal, perceptual-performance, quantitative, general cognitive, memory, and motor. These six scales are made up of 18 short subtests, grouped in various ways to form each scale (McCarthy, 1972).

The APA also stipulates that a criterion instrument must be valid if the validity of another scale is to be determined against it (1974). Examining the literature on the MSCA, there is some evidence of its validity. Kaufman and Kaufman (1974) tested the

discriminatory power of the MSCA on a group of 22 five- to nine-year-old children with diagnosed minimal brain dysfunction and a contrasting group of normals matched on sex, race, age, occupation of father, and word knowledge subtest score of the MSCA. They found significant differences in favor of the normal children on 12 of the 19 subtests; these subtests fell mostly in the quantitative (Q), perceptual-performance (P), and memory (Mem) scales. MSCA General Cognitive Index (GCI) scores were better indicators of low school achievement than WISC or Stanford-Binet IQ scores. Kaufman (1973b) reported on the correlation of MSCA scores with first grade Metropolitan Achievement Test scores (given four months later) for 31 six-year-olds. He found significant correlations between GCI, P, and Q scales and achievement test scores. The MSCA was superior to the WPPSI and equal to the Stanford-Binet in overall level of significant correlations. McCarthy (1972) reports on concurrent validity using the same sample of children. The GCI was correlated .71 with WPPSI IQ and .81 with Stanford-Binet IQ. In another study of concurrent validity Davis and Rowland (1974) used a wider age range (2 1/2 to 8 1/2) of children (N = 33) than McCarthy and found higher or equal correlation between the MSCA GCI and Stanford-Binet IQ (using both 1960 and 1972 norms) than were reported in the MSCA manual.

In addition to having a significant positive relationship to other well-known tests, our criterion test must measure the same constructs which the new test - the PET - proposes to measure in

order that a valid comparison can be made between the two tests (APA, 1974). McCarthy (1972) reports that the six scales of the MSCA were devised with the help of a factor analysis to determine the different abilities tested by the MSCA. Kaufman and Hollenbeck (1973) report the results of this factor analysis using 373 children at three age levels (three, five, seven) from the standardization sample. They found three main factors which appeared at all ages: a general cognitive factor which accounted for 31-45 percent of variance in test scores; a memory factor which accounted for 16-23 percent of the variance; and a motor factor which accounted for 9-14 percent of the variance. Three other factors were found to change in importance depending on the age of the child. A verbal factor was found for ages three and seven, but these tasks loaded on the memory factor at age five. A quantitative factor was evident at age five, and a nonverbal, perceptual-performance factor was found at ages three and seven. The authors compared these six factors to factors found on the Stanford-Binet, WISC, WPPSI, and Merrill-Palmer Scale of Mental Tests. They concluded that the MSCA factors are similar to those of other tests, but more inclusive than those of any one of the other tests.

In a second factor analytic study Kaufman (1975) used the entire standardization sample of 1032 children divided into five age groups (2 1/2, 3-3 1/2, 4-4 1/2, 5-5 1/2, 6 1/2-7 1/2-8 1/2). He found verbal, motor, general cognitive, and memory factors at all age levels. A perceptual-performance factor was found in all but

the youngest age group. Additional factors which varied by age group were drawing, semantic memory, quantitative, and reasoning. Kaufman and Dicuio (1975) used separate factor analyses on a group of 688 white children and a group of 124 black children ranging in ages from 3 to 7 1/2 years who were drawn from the standardization sample. They found verbal, perceptual-performance, and motor factors for both white and black children. The black children also showed an additional memory factor.

In another paper which showed the relationship of the MSCA to a set of psychological constructs, Kaufman (1973a) analyzed the tasks of the MSCA into Guilford's structure of intellect model, and compared the operations, contents, and products obtained by analysis of the MSCA with those found on the WISC, WPPSI, and Stanford-Binet. He found general similarity between the elements of Guilford's model utilized by each test.

These studies (Kaufman, 1973a, 1975; Kaufman and Dicuio, 1975; Kaufman and Hollenbeck, 1973) support the empirical and theoretical construct validity of the MSCA as compared to more well-established intelligence tests. Other reviewers (Krichev, 1974) have remarked on the paucity of validity data available on the MSCA. The one cause for caution is the low number of children tested, never more than forty except in Kaufman's large factor analytic studies. However, no study found a lack of validity despite the use of a wide range of analyses and comparisons. The evidence available at this point, although not conclusive, supports the hypothesis that the MSCA has the validity necessary for use as a criterion instrument.

In comparing the seven abilities which the PET attempts to measure to the constructs found empirically in the factor analytic studies, there is a close correspondence in the following areas:

<u>PET scales</u>	<u>MSCA factors</u>
gross motor	motor
visual memory	memory
auditory memory	memory
visual motor performance	perceptual-performance
concept development	cognitive
language development	verbal (and cognitive)

A question remains as to whether the abilities which make up the PET auditory perception scale are measured adequately by the MSCA verbal factor. In other areas the MSCA appears to meet the requirement that it examines the same constructs as the test to be validated.

In order to obtain an appropriate criterion for the auditory perception scale of the PET, the auditory reception subtest from the Illinois Test of Psycholinguistic Abilities (ITPA) might be used. This subtest measures the ability to gain meaning from auditorily received stimuli; in this case combinations of words (Kirk and Paraskevopoulos, 1969). Its criterion-related validity has been investigated (Cicirelli, Granger, Schemmel, Cooper, and Holthouse, 1971; Hirshoren, 1969; Mueller, 1969) and found satisfactory. The auditory reception subtest has been found to load as an independent factor (Hare, Hammill, and Bartel, 1973; Newcomer, Hare, Hammill,

and McGettigan, 1975) which has been termed meaningful receptive language. This subtest appears to be a more meaningful criterion of auditory perception than the MSCA verbal factor based on the requirement of similarity of constructs previously discussed.

The validation of a screening test is a complex process. A single study, such as Zinn's (Note 4), is seldom adequate to cover all the technical aspects. Perhaps in the rush to transfer screening from the medical realm to the psychological one, the issues involved in developmental screening have not been fully explored. The rest of this paper will present a study of the validity of the PET and a discussion of the issues involved in validating screening as applied to the PET.

Method

Subjects

Sixty-five children ranging in age from 48 to 59 months (mean age = 53.4 months) participated in the study. The children tested were taken from those who came into the North Carolina Statewide Pre-Kindergarten Screening Program (SPSP) for their initial (Phase I) screening. All children were from the catchment area of the Boone, North Carolina Developmental Evaluation Center which includes seven rural, Appalachian counties in northwestern North Carolina. There were 41 boys and 24 girls included in the sample group. Out of the 65 children, 64 were white and one was black.

The frequency distribution of the major occupational categories of the heads of households, usually the father, is given in Table 1 along with the distribution of occupations found by the 1970 U.S. Census for North Carolina. Using occupational and educational status scores obtained from the Bureau of the Census (1963), a composite socio-economic status score was calculated for each family by averaging the head of household's status scores for occupation and education. Various unknown selection factors undoubtedly operated not only in those persons who consented to return for follow-up testing, but also in those who decided to bring their children for the initial screening. However, all children who were able and willing to return for follow-up were included in the study.

Instruments

The North Carolina Psychoeducational Screening Test (PET) (DHS, Note 1) is made up of seven different scales, each of which can be scored or delayed, average, or advanced based on points received for success on items within that scale. These scales, as described earlier in this paper, are: gross motor, visual memory, auditory perception, auditory memory, visual-motor performance, concept development, and language development. A total of two or more delays or advances out of the seven was the criterion for referral for further testing.

The McCarthy Scales of Children's Abilities (MSCA) (McCarthy, 1972) is a standardized, individually-administered test designed to measure skills of children, ages 2 1/2 to 8 1/2. It provides

TABLE 1
 Major Occupational Groups of
 Heads of Households

<u>Occupational Group</u>	<u>Percent Sample</u>	<u>Percent N.C. Population^a</u>
Professional	12	9.4
Managers and administrators	19	10.2
Clerical and Sales	6	16.4
Foremen and Skilled Craftsmen	34	21.4
Operatives	17	22.1
Service	5	6.0
Unskilled	4	9.4
Farmers	0	3.9
Unknown	5	5.5

^aFrom U.S. Census, 1970.

standard scores in six different areas: verbal, perceptual-performance, quantitative, general cognitive, memory, and motor skills.

Procedure

To recruit children for Phase I screening SPSP staff wrote letters to parents of four-year-olds based on birth records, contacted community organizations dealing with four-year-olds, such as day care centers and kindergartens, and used public relations methods such as posters, newspaper stories, and radio announcements.

The PET was given to all the children along with the other screening procedures used by the SPSP as a part of routine Phase I screening. The criterion tests were the MSCA and the Auditory Reception subtest of the ITPA which were given as a follow-up to the initial screening. Children who scored as positive cases of either advanced or delayed development on the PET (N = 31) were tested by members of the SPSP staff as a part of routine Phase II screening. Children who fell in the average range of development on the PET (N = 34) were tested by the investigator. In both cases parental consent for further testing was obtained after careful explanation of the reasons for further testing. See Appendix B for a copy of the permission form used for the study. Individual participants were free to terminate their involvement in the testing without penalty at any time. After the testing was completed all parents were given a full explanation of their children's performance on the test and any questions were answered.

Any children who appeared to need referral for treatment or further testing based on the follow-up test results were referred through the SPSP system in order to obtain the needed help. Thus all known standards for research with human subjects proposed by the APA (1973) have been met by this study.

The median interval between the two tests was 8.8 days. In 80 percent of the cases follow-up testing was completed within 15 days and 95 percent of the follow-up testing was completed within 30 days.

Results

An analysis of demographic data using the chi-square technique revealed that the sample population differed significantly from what would be expected of 4-year-olds in North Carolina, based on the 1970 U.S. Census, in both sex and race ($df = 1$, $\chi^2 = 5.86$, $p < .02$; $df = 2$, $\chi^2 = 36.93$, $p < .001$, respectively). Chi-square analysis of socioeconomic status scores of heads of households showed that they differed significantly from what would be expected for the U.S. population (Bureau of the Census, 1963) ($df = 9$, $\chi^2 = 35.69$, $p < .001$). These data indicate that the sample population was significantly higher in socioeconomic status, ratio of whites to blacks, and boys to girls than would be expected from a purely random sample of the population.

Pearson correlation coefficients between individual PET scales and individual MSCA scales are presented in Table 2. Only the PET

Table 2
Correlation Coefficients between Criterion Test Scores and Scores
on Each of the Seven Scales of the PET

PET Scales	McCarthy Scales							ITPA Auditory Reception
	GCI	Verbal	Perceptual- Performance	Quantitative	Memory	Motor		
Gross Motor	.31*	.27	.23	.33*	.29*	.36*	.26	
Visual Memory	.24	.27	.16	.22	.26	.12	.19	
Auditory Perception	.36*	.33*	.24	.33*	.36*	.10	.28*	
Auditory Memory	.54*	.55*	.31*	.56*	.62*	.27	.36*	
Visual-Motor Performance	.44*	.36*	.51*	.34*	.38*	.27	.30*	
Concept Development	.52*	.48*	.51*	.38*	.42*	.22	.42*	
Language Development	.76*	.69*	.63*	.59*	.68*	.39*	.41*	
Overall PET Score	.78*	.72*	.67*	.65*	.73*	.40*	.53*	

* = $p < .01$

language development scale correlated above .60 with any MSCA index. Auditory memory, visual-motor performance, and concept development were the only other PET scales which correlated above .50 with any part of the MSCA. The auditory reception subtest of the ITPA was not correlated highly with any part of the PET. Although never done in actual practice, an overall PET score was obtained by adding together all points obtained on each PET scale for each case. This overall score produced correlations only slightly higher than those obtained for the PET language development scale alone, as can be seen in Table 2.

A step-wise multiple regression analysis with a cutoff point determined by F ratio (2.8) which excluded variables accounting for less than 2 percent of criterion test score variance showed that the PET language development scale accounted for the majority of the variance in scores on the criterion tests. The percentages of variance accounted for are shown in Table 3. Concept development, auditory memory, and visual-motor performance scales accounted for additional variance on the cognitive parts of the criterion instruments. The PET gross motor scale also accounted for some variance on the motor scale of the MSCA.

Table 4 gives the frequencies of cases falling in each of the screening categories based on SPSP criteria. If a child scored as delayed or advanced in two or more of the seven PET scales, he was referred. A MSCA general cognitive index either above or below 1.2 standard deviations from the mean (above 119 or below 81) was

Table 3
 Percent Variance in Criterion Instruments
 Accounted for by Individual PET Scales

PET Scales	McCarthy Scales							ITPA Auditory Reception
	GCI	Verbal	Perceptual- Performance	Quantitative	Memory	Motor		
Gross Motor	---	---	---	---	---	7.8	---	
Visual Memory	---	---	---	---	---	---	---	
Auditory Perception	---	---	---	---	---	---	---	
Auditory Memory	2.6	4.6	---	10.2	12.1	---	---	
Visual-Motor Performance	---	---	3.3	---	---	---	---	
Concept Development	9.9	8.7	11.6	2.7	3.2	---	17.7	
Language Development	57.8	47.8	39.8	35.1	46.8	15.3	9.2	

Note: Values less than 2% not reported.

Table 4

Frequency Count of PET - McCarthy Agreement Based
on Statewide Prekindergarten Screening Program Criteria

Decisions Based on PET ^b	McCarthy GCI Classification ^a			
	Delay	Average	Advance	
Refer for Delay	5	16	0	21
No Referral	0	32	2	34
Refer for Advance	0	7	3	10
	5	55	5	65

^a ± 1.2 standard deviations. See text for full explanation.

^b Two delays or two advances. See text for full explanation.

considered confirmation of the PET decision to refer for advance or delay, respectively.

Table 5 is derived from Table 4 and provides the rates of agreement between the PET decision to refer and MSCA classification as a positive case of delayed or advanced development (also known as sensitivity), and as a negative case, i.e., average development (specificity) (Thorner and Remein, 1961). These rates of agreement indicate the probability that the PET will agree with the MSCA when applied to cases whose MSCA classification is known. Sensitivity is the chance that the screening test will be positive when applied to someone known to be a positive case. Specificity is the chance that the screening test will be negative when applied to someone known to be a negative case. Thus, sensitivity and specificity are based on the numbers of cases which fall into the respective categories as confirmed diagnoses (Feinstein, 1977). Overall agreement, over-referral, and under-referral rates are based on the total sample population of 65 children screened (Frankenburg, Goldstein, and Camp, 1971). The PET successfully identified every case of delayed development which the MSCA picked out. However the PET also classified 25 percent of the entire sample as delayed without subsequent confirmation by the MSCA. This leads to a rather low rate of accuracy for negative cases. With only five cases each falling into the advanced and delayed categories on the MSCA, caution should be used in interpreting these rates of accuracy.

The predictive accuracy (Feinstein, 1977) or efficiency (Meehl and Rosen, 1955) of a given PET decision taking into account

Table 5
Rate of PET - McCarthy Agreement

McCarthy GCI Classification	PET Agreement	PET Overreferrals ^a (%)	PET Underreferrals ^a (%)
Delay	1.00 ^b	24.6	00
Advance	.60 ^b	10.7	3.1
Average	.58 ^c	---	---
Overall	.62 ^a	35.3	3.1

^aRate based on total sample population (N = 65)

^bSensitivity, rate based on McCarthy classification (N = 5)

^cSpecificity, rate based on McCarthy classification (N = 55)

the base rate of developmental delay in the population is also of interest. That is, if a decision is made concerning any particular child based on the three possible PET outcomes (advance, delay, average), what are the chances that the decision is correct? The decision-making process is just the reverse of that used to determine sensitivity and specificity. Instead of going from known classification to screening test agreement, the starting point is screening test classification which is used to predict the confirmed classification. This probability depends on the PET outcome and the base rate of the characteristic inferred from that outcome. Table 6 shows the probability that the PET has correctly identified a given case as the prevalence rates vary. Naturally increasing prevalence rates lead to greater chances of correctly identifying positive cases and lesser chances of correctly identifying negative cases, but the basic rates of agreement between the screening test and its criterion also determine this probability. For example, given a prevalence rate of 3 percent for delayed development, out of 100 children who are referred as delayed by the PET, only ten will actually be confirmed as true cases of developmental delay. On the other hand, at the same 3 percent prevalence rate, an average PET outcome has 98.5 percent chance of being confirmed in follow-up testing.

Discussion

The question might be asked, why is such detailed technical analysis of a screening procedure necessary? The answer lies in

TABLE 6

Probability of a Correct Decision Using the PET
at Different Prevalence Rates

<u>Prevalence Rate per 100</u>	<u>PET Outcomes</u>		
	<u>Delay</u>	<u>Advance</u>	<u>Average</u>
1	.037	.049	.995
2	.071	.095	.990
3	.104	.137	.985
4	.135	.176	.980
5	.165	.213	.975
6	.194	.247	.970
7	.221	.278	.965
8	.246	.308	.960
9	.271	.337	.955
10	.293	.363	.949
11	.317	.388	.944
12	.339	.412	.939
15	.399	.475	.922
20	.484	.562	.893
25	.556	.631	.862
30	.617	.687	.829
35	.669	.734	.795
40	.715	.774	.758

several important differences between screening programs and the usual forms of clinical care provided by the helping professions.

First of all screening programs by design usually affect large numbers of people. Therefore, they have high visibility and impact in communities and may affect public attitudes towards care providers. A potentially beneficial procedure may be discredited or the image of particular care providers, institutions, or types of care may be damaged by the large-scale errors which can occur in mass programs. The public may then ignore truly beneficial programs no matter how carefully evaluated these are. As McKeown (1968) has pointed out, mass screening may also "consume vast resources before it is discovered to be ineffective or inefficient....at the expense of other medical uses of the same resources." The dangers in the application of any mass program without sufficient prior evaluation are seen in the results of the ill-fated swine-flu vaccination program, and in the debate over the benefits of the breast cancer x-ray screening program. In both of these instances a preventive measure was later questioned as evidence of increased risk associated with the measure increased.

The second major difference between most screening programs and typical clinical care is that the helping person takes the initiative in discovering (through screening) disability or deviancy, rather than waiting until the individual is motivated to seek the assistance of a care provider for a problem (McKeown, 1968). This places the care provider in a rather awkward, and at present unexplored, ethical position. Follow-up treatment is implied as a duty

of the care provider, if the individual is to receive a benefit from the screening. This problem did not arise when screening originated since its purpose was to protect the public health through control of communicable diseases. Screening for chronic disease, and developmental screening in particular, completely alters this situation.

The potential harm of labeling the person screened as either having or not having the condition being tested for seems to need careful consideration in view of the position of the care provider as initiator of the interaction. Inappropriate labeling of mental retardation or false reassurance of the absence of amblyopia are two examples of these dangers. Even though follow-up and final diagnosis may eliminate false positive cases, the anxiety and doubts caused the individual by a screening decision in the interval between screening and follow-up may not be easily removed. This is especially true in developmental screening where borderline cases occur frequently.

These special considerations mean that extra care must be used in evaluating the potential benefits and risks of screening prior to the application of any screening program to large population groups. This research has considered one aspect - the validity of a screening procedure - which is a necessary (but not sufficient) condition for a successful screening program. A full discussion of all the issues involved is beyond the scope of this paper. Several authors provide discussions of the issues involved in general health screening (Cutting et al., 1970; McKeown, 1968; McKeown and Knox, 1968;

Smillie, 1952; Thorner, 1969; Wilson, 1963; Wilson and Jungner, 1968) and developmental screening in particular (Frankenburg and Camp, 1975; Meier, 1973; Thorpe, 1975).

As McKeown and Knox (1968) pointed out, research and evaluation of screening procedures must be carried out before the evidence so urgently points to the need for screening that a program must be mounted regardless of the state of knowledge concerning the screening procedures themselves. The Early and Periodic Screening, Diagnosis, and Treatment (EPSDT) program is an example of government pressure to do something. As Dixon (1974) has stated:

EPSDT is the means the federal government has selected to force the states to seek out poor children...and to offer them preventive services. The use of periodic screening exams...is an untried method of providing health care and may be unsuccessful.

The Statewide Pre-Kindergarten Screening Program (SPSP) may have been another example of pressure to provide a program overriding the need for careful evaluation. This may have been at least partially responsible for SPSP's transformation (and effective termination as previously designed) (Conn, 1977).

The correlation and multiple regression statistics on the PET support the validity of the auditory memory, visual-motor performance, concept development, and, questionably, gross motor scales. However, by far the most effective scale on the PET is the language development scale. In every instance is correlation with the McCarthy Scales was highest, although the auditory memory scale is very close on the memory and quantitative scales.

The concept development scale had a higher correlation with the auditory reception subtest of the ITPA. Even combining all the scales of the PET into one overall score produces correlations with the MSCA that are only slightly higher than those of the language development scale alone.

The picture is similar for the multiple regression analysis where the language development scale accounts for 35 to 58 percent of the variance on the cognitive scales of the MSCA. It also accounts for more variance on the MSCA motor scale than the PET gross motor scale. Only on the auditory reception subtest did another PET scale (concept development) account for more variance.

In view of the need for economy of time and ease of use in screening tests, the language development scale of the PET might be considered for use alone, or with some supplemental measure of perceptual-motor functioning. Certainly the visual memory and auditory perception scales of the PET have not shown any results to justify their retention in the test.

Looking at the PET's performance as a screening test, its weakest area is its lack of a high level of specificity. This leads to a large number of over-referrals and a lack of efficiency in its rate of accurate prediction. However, the PET does its job well in finding cases of developmental delay, not missing a single case. A low under-referral rate is considered important in screening for developmental delay. In this case intervention begun at age four could help prevent the child from becoming caught in a cycle of failure and secondary emotional problems during his first years in

school. The PET's extremely good record in finding cases of developmental delay is achieved at the cost of many over-referrals. A change in the referral criteria might be considered to remedy this. Frequently such a change causes only a small increase in under-referrals while obtaining a large decrease in over-referrals (Frankenburg and Camp, 1975).

One problem which cannot be avoided is the relatively low rate of mental retardation in the general population. Low prevalence rates mean that even the most accurate test will have a low yield and that an apparently high rate of agreement on negative cases will produce large numbers of false positive cases.

The accepted prevalence rate for mental retardation is about 3 percent (Baroff, 1974; Robinson and Robinson, 1976). Recently some researchers (Kott, 1968; Mercer, 1973; Tarjan, Wright, Eyman, and Keeran, 1973) have suggested that the true prevalence might be lower, but 24 out of 40 states surveyed (Luckey and Neman, 1976) use a prevalence rate between 2.76 percent and 3.25 percent in planning mental retardation services.

However, there is some question as to just what range of developmental problems the PET was designed to identify. If we include learning disabilities in the range of conditions it may be able to detect, then the applicable prevalence rate will increase. The lack of clear agreement on a definition of learning disabilities makes estimation of learning disability prevalence especially difficult (Lerner, 1976; Walzer and Richmond, 1973). Lerner (1976)

reports 5 to 6 percent of Texas preschoolers were identified as learning disabled. Silverman and Metz (1973) surveyed school administrators nationwide and found 1.9 percent of elementary school children received learning disabilities special education, but a total of 3.1 percent were recognized as learning disabled.

It is often difficult to distinguish between types of developmental disorders. In reality most handicapped children display developmental problems in learning, socio-emotional adjustment, and behavior that are intricately mixed together. Miller, Hampe, Barrett, and Noble (1971) found 16 percent of the general population of children had learning or behavior problems using a parental rating scale. In North Carolina, Richardson and Higgins (1965) found 14.1 percent of children examined had difficulties with intellectual, language, or emotional development. Werner, Bierman, and French (1971) reported approximately 12 percent of two-year-olds were delayed in intellectual and social-adaptive skills. Follow-up at age ten showed 23.2 percent and 26.4 percent of children to have learning or emotional problems, respectively.

The PET is also intended to pick out cases of advanced development as well as delayed development. Estimates of the prevalence of giftedness are also confounded by a definitional problem of who should be included - academically superior, creative, or talented children. Newland (1976) estimates prevalence of giftedness at 8 percent, although more conservative estimates of 2 percent (Gibson and Channels, 1976) are also accepted. Recent federal

government reports recognize 3 percent as a conservative estimate of the prevalence rate for giftedness (Marland, 1972).

Differing prevalence estimates, definitional problems, and uncertainty over exactly which problems the PET identifies makes interpretation of probabilities which change with varying prevalence rates somewhat problematical. This investigator assumes that the PET can at a minimum identify mental retardation and perhaps may be able to pick out other developmental problems in learning and social skills.

Using North Carolina Division of Health Services (DHS) criteria (DHS, Note 2) for the MSCA, classification categories is a way of side-stepping the prevalence problem. Their criteria of ± 1.2 standard deviations from the mean will give us prevalence rates based on the normal curve. McCarthy (1972) reports that approximately 12 percent of the population will fall in each area (above a GCI of 119 and below a GCI of 81). This might be the most practical figure to use in interpreting the PET's efficiency. Given a 12 percent prevalence rate a child identified as delayed would have a 33.9 percent chance of being confirmed as delayed, and one identified as advanced would have a 41.2 percent chance of confirmation. In this study the actual prevalence rates for MSCA GCI scores above 119 and below 81 were both 8 percent leading to lower predictive accuracy rates of 24.6 percent for delays and 30.8 percent for advances. It should be noted that in interpreting rates of predictive accuracy for the average classification, the prevalence of both delays and advances must be taken into account. For example, a prevalence of 8 percent delays,

8 percent advances means a prevalence of 16 percent total positives. In this case predictive efficiency for negative cases (average) would be slightly less than 92 percent. Here the effect of low prevalence rates has the opposite effect. Since a low prevalence rate for the presence of a condition means that its absence is highly prevalent, the chance of correct prediction of its absence is increased. Simply by calling all 65 cases negative, correct prediction would occur in 84 percent of the cases.

Although the effect of prevalence rates should always be kept in mind, other researchers (Rogan and Gladen, 1978) have suggested that these predictive values are of somewhat limited value in evaluating screening procedures. They point out that screening typically selects persons who as a group have a higher proportion of positive cases than the general population. This does not guarantee that each individual in the group has a high probability of being a positive case. One way to resolve the problem of low prevalence rates is to select a population with a higher prevalence rate for screening (Frankenburg and Camp, 1975; Meehl and Rosen, 1955). A high risk register is one way of doing this for developmental problems. However, in England, where such a register has been used extensively, researchers have found that many cases of developmental delay are missed if screening is limited to cases on the high-risk register (Hooper and Alexander, 1971; Meier, 1973; Rogers, 1968; Starte, 1976).

Sensitivity and specificity do not depend on the prevalence rates for the condition being screened. They are affected by the shape of the frequency distribution curve of the characteristic being

tested for (Thorner and Remein, 1961). Thus sensitivity and specificity are important measures of a screening test because they remain relatively stable as prevalence estimates vary or, as in the case of the PET, when the applicable prevalence rate is unclear.

However sensitivity and specificity could change in a high risk population since the shape of the distribution curve of developmental delay might be different in this population (Feinstein, 1977). Revalidation studies on the high risk population would then be necessary. Since most mass screening tests are intended to be applied to the general population, whose frequency distribution for developmental delay will probably not change substantially, sensitivity and specificity should not change in this application. It also follows that prevalence estimates obtained through epidemiological surveys of the general population could help determine relatively stable predictive values for screening tests applied to the general population.

Another way to deal with the large number of over-referrals caused by low prevalence rates and low specificity, such as the PET has, is to use a series of tests. The first is designed to pick out a very high proportion of total positive cases at the cost of a large number of false positives (high sensitivity, low specificity). The second test functions in the reverse way, picking out almost all the negative cases, but mistakenly failing to identify some positive cases (low sensitivity, high specificity). All those cases which are positive on the first test are tested with the second test.

Persons who tested positive on both tests could be considered positive cases and those who tested negative on the first test could be considered negative cases (Feinstein, 1977). This solution would minimize the over-referrals occurring due to the low specificity of the first test. Some under-referrals would still occur, but these would be fewer than if the less sensitive (second) test was used alone on the general population. The problem of what to do with the group which tests positive then negative remains. If resources are limited, they could be considered negative since they would contain a lower proportion of positive cases than the group with two positive tests.

This was perhaps the rationale behind the organization of the SPSP into two phases - the PET, then the MSCA. However, the MSCA does not meet the requirements of rapidity and simplicity of administration which makes mass screening feasible. The PET, as has been shown in this study, is also longer and more complicated than is justifiable. An abbreviated PET and a second short screening test for Phase II would be more rapidly and easily administered, and more economical. This might mean that both tests could be completed in one visit instead of two used in the original system. This would certainly increase the overall economic efficiency of the program.

The balancing of the cost of follow-up on the over-referrals versus the cost of undiscovered cases (under-referrals) might be used to help determine the proper cut-off points needed on each test to assure sensitivity and specificity figures which maximize efficiency.

However, other criteria for determining cut-off points and decision-making processes are possible. Alberman and Goldstein (1970) present mathematical models for maximizing the yield of identified handicapping conditions given a fixed amount of resources in the context of use of a high risk register. The final criteria chosen depend on societal values and political and economic limits which are beyond the realm of science alone. Yet scientific input and information is necessary for a public policy decision which is most beneficial to all.

A useful method of evaluating the PET is to compare it with other tests used for screening on similar statistical indices. The major difficulty with this approach is that there are few screening tests which report their validity data in such a way that evaluation of their effectiveness using a screening-type decision-making strategy is possible. Most studies report only correlation coefficients which are not adequate evidence of screening test validity. What is needed is classification of test scores into 2 X 2 tables for referred and non-referred cases or the reporting of each individual test score. The investigator was able to find several studies which did adequately report data, and these are presented in Table 7. For further review of screening tests for developmental problems, consult Bailey, Kiehl, Akron, Loughlin, Metcalf, Jain, and Perrin (1974); Cowen, Dorr, and Orgel (1971); Frankenburg and Camp (1975); Meier (1973); Rogolsky (1968-69); and Thorpe and Werner (1974).

Table 7
Evaluative Profiles of Selected Screening Tests

Test	Denver Developmental Screening Test ^a	Bayley DQ or Stanford-Binet IQ<70	Kaiser Permanente Screening Battery ^b	Maryland Screening Battery ^c	Quick Test	Vane Kindergarten Test ^e	
Validity Criteria			Professional diagnosis of psychosocial problem	Poorest 10% by teacher rating	Stanford-Binet IQ<80	WISC IQ<80	Placement in Special education class
N	237	246	100	66	52	50	135
Sensitivity	.68	.92	.42	.78	1.00	.90	.30
Specificity	.92	.97	.98	.95	.91	.88	1.00
Under-referrals (%)	3.0	3.2	7.0	3.0	0	2.0	4.4
Over-referrals (%)	7.1	0.4	2.0	4.5	7.7	10.0	0
Predictive Validity for Positive Cases							
3% Prevalence	.21	.49	.39	.33	.26	.18	1.00
12% Prevalence	.54	.81	.74	.68	.63	.50	1.00

^aFrankenburg, Goldstein and Camp, 1971

^bAllen, Metz and Shinefield, 1971

^cArmistead and Crawford, 1974

^dPless, Snider, Eaton and Kearsley, 1965

^eColligan and O'Connell, 1974

Comparison of the statistics found for the PET (see Tables 5 and 6) shows that the PET is equal to or better than all of the tests listed in selecting without omission those children who are later confirmed as cases of developmental delay. It achieves this high sensitivity at the expense of including many children with average abilities in the referred group. Thus its accuracy in identifying negative cases is lower than any of the other tests. Some caution should be used in interpreting this table because of the differing validity criteria used and the relatively low number of children in some studies.

In conclusion, this study indicates that the PET would not be the sole screening measure of choice for a mass developmental screening program, primarily because of its high rate of over-referrals and consequent lack of efficiency. A change in the cutting score used for referral or use of the PET in combination with a test which is more accurate for negative cases (high specificity) might help remedy this problem. Another, though perhaps less desirable, way to increase efficiency would be to screen only high-risk populations which have higher prevalence rates for developmental problems. The PET does have the virtue of being highly sensitive to cases of developmental delay, though a larger sample is needed to test this sensitivity figure reliably.

The failure of several of the seven PET scales to account for additional variance on the criterion instruments also indicates that the PET is longer than necessary, another disadvantage for a screening test. These unnecessary scales may also help increase the

rate of over-referrals. A more practical screening strategy would be to abbreviate the PET, perhaps using only the language development scale and a perceptual-motor test. Revalidation would then be necessary. If the revalidation found the same characteristics of high sensitivity and low specificity on the new PET, then the strategies mentioned in the previous paragraph could be used.

A final reminder that the sample population for this study is significantly different from the North Carolina and United States population in sex ratio, ethnic composition, and socio-economic status is needed. The families who participated were willing to volunteer to return for a follow-up visit, another selection mechanism. Therefore, any generalization of this data and the conclusions based on it should be made cautiously, keeping these limitations in mind.

Reference Notes

1. Division of Health Services, North Carolina Department of Human Resources. North Carolina Psychoeducational Screening Test (DHS Form 2181, Rev. 11/75). Unpublished manuscript, 1975. (Available from Division of Health Services, Department of Human Resources, Raleigh, N.C.)
2. Division of Health Services, North Carolina Department of Human Resources. Parent Information Report (DHS Form 2316, Rev. 1/76). Unpublished manuscript, 1976. (Available from Division of Health Services, Department of Human Resources, Raleigh, N.C.)
3. Human Resource Consultants. Final Report: Consultation Services for the Statewide Prekindergarten Screening Program. Unpublished manuscript, 1976. (Available from Human Resources Consultants, 100 Eastowne Drive, Chapel Hill, N.C. 27514.)
4. Zinn, V. Final Report: The Development of the North Carolina Psychoeducational Screening Test. Unpublished manuscript, 1976. (Available from Human Resource Consultants, 100 Eastowne Drive, Chapel Hill, N.C. 27514.)

References

- Alberman, E.D. and Goldstein, H. The "at risk" register - A statistical evaluation. British Journal of Preventive and Social Medicine, 1970, 24, 129-135.
- Allen, C.M., Metz, J.R., and Shinefield, H.R. Test development in the pediatric multiphasic program. Pediatric Clinics of North America, 1971, 18(1), 169-178.
- American Psychological Association. Ethical Principles in the Conduct of Research with Human Participants. Washington, D.C.: Author, 1973.
- American Psychological Association. Standards for Educational and Psychological Tests. Washington, D.C.: Author, 1974.
- Armistead, L.M. and Crawford, E.E. An early developmental screening test for public health use. American Journal of Public Health, 1974, 64(3), 241-244.
- Bailey, E.M., Kiehl, P.S., Akron, D.S., Loughlin, H.H., Metcalf, T.J., Jain, R., and Perrin, J.M. Screening in pediatric practice. Pediatric Clinics of North America, 1974, 21(1), 123-165.
- Bakwin, H. and Bakwin, R.M. Behavior Disorders in Children (4th edition). Philadelphia: W.B. Saunders, 1972.
- Baroff, G.S. Mental Retardation: Nature, Cause, and Management. New York: John Wiley, 1974.
- Bureau of the Census, U.S. Department of Commerce. Methodology and Scores of Socioeconomic Status (Working Paper No. 15). Washington, D.C.: U.S. Government Printing Office, 1963.

- Cicirelli, V.G., Granger, R., Schemmel, D., Cooper, W., and Holthouse, N. Performance of disadvantaged primary-grade children on the revised ITPA. Psychology in the Schools, 1971, 8, 240-246.
- Colligan, R.C. and O'Connell, E.J. Should psychometric screening be made an adjunct to the pediatric preschool examination? Clinical Pediatrics, 1974, 13(1), 29-34.
- Commission on Chronic Illness. Chronic Illness in the United States, Vol. 1: Prevention of Chronic Illness. Cambridge: Harvard University Press, 1957.
- Conn, R. N.C. screening of kids' health to start at birth. Raleigh (N.C.) News and Observer, February 8, 1977, pp. 1B, 3B.
- Cowen, E.L., Dorr, D.A., and Orgel, A.R. Interrelations among screening measures for early detection of school dysfunction. Psychology in the Schools, 1971, 8, 135-139.
- Cutting, C.C., Haynes, L.L., Bird, W.W., Rubin, L., West, R.T., and Felch, W.C. Multiphasic health screening: Panel discussion at 1969 Annual Meeting, American Association of Medical Clinics. Group Practice, 1970, 19(5), pp. 21; 34-39.
- Davis, E.E. and Rowland, T. A replacement for the venerable Stanford-Binet. Journal of Clinical Psychology, 1974, 30(4), 517-521.
- Dixon, M.S. EPSDT (Early and Periodic Screening, Diagnosis and Treatment Programs). Pediatrics, 1974, 54(1), 84-90.

- Feinstein, A.R. Clinical Biostatistics (Chapter 15, On the sensitivity, specificity and discrimination of diagnostic tests). St. Louis: C.V. Mosby, 1977. (Also appeared as Clinical biostatistics - XXXI. Clinical Pharmacology and Therapeutics, 1975, 17, 104-116.)
- Frankenburg, W.K. Evaluation of screening procedures. In Oglesby and Sterling (Eds.), Proceedings, Bi-Regional Institute on Earlier Recognition of Handicapping Conditions in Childhood. Berkeley: School of Public Health, University of California, 1970.
- Frankenburg, W.K. and Camp, B.W. (Eds.) Pediatric Screening Tests. Springfield, Illinois, Charles C. Thomas, 1975.
- Frankenburg, W.K. and Dodds, J.B. The Denver Developmental Screening Test. Journal of Pediatrics, 1967, 71(2), 181-191.
- Frankenburg, W.K., Goldstein, A.D., and Camp, B.W. The revised Denver Developmental Screening Test: Its accuracy as a screening instrument. Journal of Pediatrics, 1971, 79(6), 988-995.
- Gibson, J. and Chennels, P. (Eds.) First World Conference on Gifted Children: Looking to Their Future. London: Latimer, 1976.
- Hare, B., Hammill, D., and Bartel, N. Construct validity of selected subtests of the Illinois Test of Psycholinguistic Abilities. Exceptional Children, 1973, 40(1), 13-20.
- Hirshoren, A. A comparison of the predictive validity of the Revised Stanford-Binet Intelligence Scale and the Illinois Test of Psycholinguistic Abilities. Exceptional Children, 1969, 35(7), 517-521.

- Hooper, P.D. and Alexander, E.L. Developmental assessment in general practice. Practitioner, 1971, 207, 371-376.
- Kaufman, A.S. Factor structures of the McCarthy Scales at five age levels between 2 1/2 and 8. Educational and Psychological Measurement, 1975, 35(3), 641-656.
- Kaufman, A.S. Analysis of the McCarthy Scales in terms of Guilford's structure of intellect model. Perceptual and Motor Skills, 1973, 36(3), 967-976 (a).
- Kaufman, A.S. Comparison of the WPSSI, Stanford-Binet, and the McCarthy Scales as predictors of first grade achievement. Perceptual and Motor Skills, 1973, 36(1), 67-73 (b).
- Kaufman, A.S. and Dicuio, R.F. Separate factor analyses of the McCarthy Scales for groups of black and white children. Journal of School Psychology, 1975, 13(1), 10-18.
- Kaufman, A.S. and Hollenbeck, G.P. Factor analysis of the standardization edition of the McCarthy Scales. Journal of Clinical Psychology, 1973, 29(3), 358-362.
- Kaufman, N.L. and Kaufman, A.S. Comparison of normal and minimally brain dysfunctioned children on the McCarthy Scales of children's abilities. Journal of Clinical Psychology, 1974, 30(1), 69-72.
- Kirk, S.A. and Paraskevopoulos, J.N. The Development and Psychometric Characteristics of the Revised Illinois Test of Psycholinguistic Abilities. Urbana, Illinois: University of Illinois Press, 1969.
- Knobloch, H. and Pasamanick, B. (Eds.) Gesell and Amatruda's Developmental Diagnosis (3rd ed.). New York, Harper and Row, 1974.

- Knobloch, H., Pasamanick, B., and Sherard, E.S. A developmental screening inventory for infants. Pediatrics, 1966, 38 (2, part II), 1095-1105.
- Kott, M.G. Estimating the number of retarded in New Jersey. Mental Retardation, 1968, 6(1), 28-31.
- Krichev, A. Review of McCarthy Scales of Children's Abilities. Psychology in the Schools, 1974, 11(2), 252-254.
- Lerner, J.W. Children with Learning Disabilities (2nd ed). Boston: Houghton Mifflin, 1976.
- Lessler, K. Health and educational screening of school-age children - Definition and objectives. American Journal of Public Health, 1972, 62(2), 191-198.
- Luckey, R.E. and Neman, R. Practices in estimating mental retardation prevalence. Mental Retardation, 1976, 14(1), 16-18.
- Marland, S.P. Education of the Gifted and Talented: Report to the Congress of the U.S. by the U.S. Commissioner of Education. Washington, D.C.: U.S. Government Printing Office, 1972.
- McCarthy, D. Manual for the McCarthy Scales of Children's Abilities. New York: Psychological Corporation, 1972.
- McKeown, T. Validation of screening procedures. In Lord Cohen of Birkenhead, E.T. Williams, and G. MacLackan (Eds.), Screening in Medical Care. London: Oxford University Press, 1968.
- McKeown, T. and Knox, E.G. Framework required for validation of prescriptive screening. In Lord Cohen of Birkenhead, E.T. Williams, and G. MacLachlan (Eds.), Screening in Medical Care. London: Oxford University Press, 1968.

- Meehl, P.E. and Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 1955, 52(3), 194-216.
- Meier, J. (Ed.) Screening and Assessment of Young Children at Developmental Risk (President's Committee on Mental Retardation, DHEW No. (OS) 73-90). Washington, D.C.: U.S. Government Printing Office, 1973.
- Mercer, J. The myth of 3% prevalence. In R.K. Eyman, C.E. Myers, and G. Tarjan (Eds.), Sociobehavioral Studies in Mental Retardation (Monographs of the American Association on Mental Deficiency, No. 1, pp. 1-18), 1973.
- Miller, L.C., Hampe, E., Barrett, C.L., and Noble, H. Children's deviant behavior within the general population. Journal of Consulting and Clinical Psychology, 1971, 37(1), 16-22.
- Mueller, M.W. Prediction of achievement of educable mentally retarded children. American Journal of Mental Deficiency, 1969, 73, 590-596.
- Newcomer, P., Hare, B., Hammill, D., and McGettigan, J. Construct validity of the Illinois Test of Psycholinguistic Abilities. Journal of Learning Disabilities, 1975, 8(4), 220-231.
- Newland, T.E. The Gifted in Socioeducational Perspective. Englewood Cliffs, N.J.: Prentice-Hall, 1976.
- North, A.F. Introduction: Screening in child health care. Pediatrics, 1974, 54(5), 608 (a).
- North, A.F. Screening in child health care: Where are we now and where are we going? Pediatrics, 1974, 54(5), 631-640 (b).

- Oglesby, A. and Sterling, H. (Eds.) Proceedings, Bi-Regional Institute on Earlier Recognition of Handicapping Conditions in Childhood. Berkeley, California: School of Public Health, University of California, 1970.
- Pless, I.B., Snider, M., Eaton, A.E., and Kearsley, R.B. A rapid screening test for intelligence in children. American Journal of Diseases of Children, 1965, 109(6), 533-537.
- Richardson, W.P. and Higgins, A.C. The Handicapped Children of Alamance County. Wilmington, Delaware: Nemours Foundation, 1965.
- Robinson, N.M. and Robinson, H.B. The Mentally Retarded Child (2nd ed.). New York: McGraw-Hill, 1976.
- Rogan, W.J. and Gladen, B. Estimating prevalence from the results of a screening test. American Journal of Epidemiology, 1978, 107(1), 71-76.
- Rogers, M.G.H. Risk registers and early detection of handicaps. Developmental Medicine and Child Neurology, 1968, 10, 651-661.
- Rogolsky, M.M. Screening kindergarten children: A review and recommendations. Journal of School Psychology, 1968-69, 7(2), 18-27.
- Silverman, L.J. and Metz, A.S. Numbers of pupils with specific learning disabilities in local public schools in the United States: Spring 1970. Annals of the New York Academy of Sciences, 1973, 205, 146-157.

- Smillie, W.G. Multiple screening. American Journal of Public Health, 1952, 42, 255-258.
- Starte, G.D. Results from a developmental screening clinic in general practice. Practitioner, 1976, 216(1293), 311-316.
- Tarjan, G., Wright, S.W., Eyman, R.K., and Keeran, C.V. Natural history of mental retardation: Some aspects of epidemiology. American Journal of Mental Deficiency, 1973, 77(4), 369-379.
- Thorner, R.M. Whither multiphasic screening? New England Journal of Medicine, 1969, 280(19), 1037-1042.
- Thorner, R.M. and Remein, Q.R. Principles and procedures in the evaluation of screening for disease. Public Health Monograph (No. 67), Washington, D.C.: U.S. Government Printing Office, 1961. (Public Health Service Publication No. 846.)
- Thorpe, H.S. Optimal developmental appraisal. Developmental Medicine and Child Neurology, 1973, 17(4), 481-485.
- Thorpe, H.S. and Werner, E.E. Developmental screening of preschool children: A critical review of inventories used in health and educational programs. Pediatrics, 1974, 53(3), 362-370.
- Walzer, S. and Richmond, J.B. The epidemiology of learning disorders. Pediatric Clinics of North America, 1973, 20(3), 549-565.
- Werner, E.E., Bierman, J.M., and French, F.E. The Children of Kauai. Honolulu: University of Hawaii Press, 1971.
- Wilson, J.M.G. Multiple screening. Lancet, 1963, (7298), 51-54.
- Wilson, J.M.G. and Jungner, G. Principles and practice of screening for disease. Public Health Papers (No. 34), Geneva, Switzerland: World Health Organization, 1968.

APPENDIX A

North Carolina Psychoeducational Screening
Test Protocol and Scoring Profile

Child's name: 	Telephone: _____	DOS: _____	Location: _____
	Sex:M: _____	DOB: _____	County: _____
	F: _____	Age: _____ yrs mos dys	_____
	Race: _____	Code: _____	Test: _____
	Time: _____	_____/_____/_____/_____	Retest: _____
			Examiner _____

1. Personal Information: NAME:

First _____ Middle _____ Last _____ 1N ()

AGE: Reply _____ Gesture _____ Count _____ 1A ()

2. Pencil: Record child's answers

a. What is that? _____ 2N ()

b. What color is it? _____

c. What can you do with it? _____

d. What other thing does the same as this? _____ 2R ()

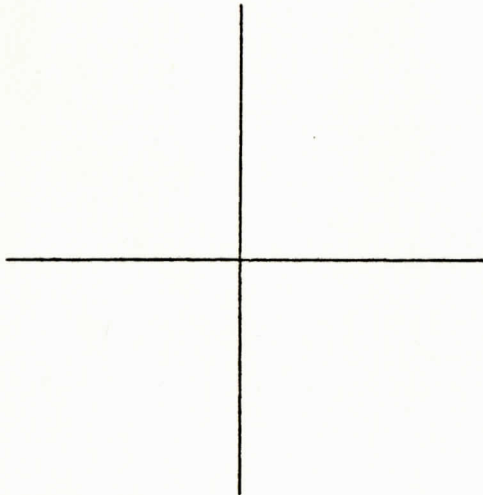
3. Copy +: UPG: yes _____ Describe: _____ 3 ()
no _____ 3UPG ()

Handedness:

right _____

left _____

both _____



4. Which one of these is most like a wheel? _____ (C or I) 4 ()
Why? _____
5. Which line is longer? 1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ (C or I) 5 ()
6. Draw a man (woman, girl, boy): Handedness: right _____ UPG: yes _____
left _____ no _____
both _____ Describe _____ 6 ()

7. Digit Span: 7-1-3 _____ (C or exact) 7 ()
2-8-5 _____
8. Three Commands: chair _____ door (wall) _____ box _____ 8 ()
Comments _____
9. Balance on one foot: Trial 1 _____ Trial 2 _____ Trial 3 _____ 9 ()
10. Hop on one foot: yes _____ no _____ 10 ()
11. Familiar Sounds: Identify: a. Clap hands _____ (C or exact) 11-a ()
b. Bell _____ 11-b ()
12. Sentences:
a. Hello _____ 12 ()
b. I like ice cream _____
c. He wants a hamburger and french fries _____
d. The happy dog was eating his food fast _____
13. Size Match: 1 _____ 2 _____ 3 _____ 4 _____ (C or # of square identified) 13 ()
14. Look Alikes: 1. L M R 2. L M R 3. L M R 4. L M R 5. L M R 14 ()
6. L M R 7. L M R 8. L M R 9. L M R Total Correct _____
15. Colors: Name: Red _____ Blue _____ Yellow _____ 15 N ()
Recognize: Red _____ Blue _____ Yellow _____ 15 R ()
16. Remember 3 Colors: BRY _____ (C or I) 16 ()
YRB _____
17. Name objects: rock _____ key _____ 17 N ()
spoon _____ button _____
penny _____ comb _____

Recognize: rock _____ key _____

56

17 R ()

spoon _____ button _____

penny _____ comb _____

Memory objects:

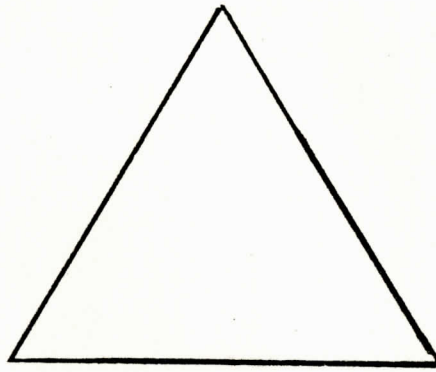
Trial 1 _____

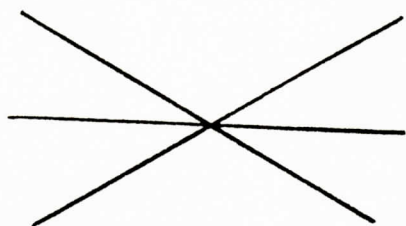
18 ()

Trial 2 _____

Trial 3 _____

- . Copy Triangle 19 ()
- . How things are alike:
- a. meat and potatoes _____ 20 ()
- b. hammer and saw _____
- c. dog and cat _____
- d. car and airplane _____
- . How much and how many: 21 ()
- a. largest elephant _____ (C or I)
- b. dog on ball _____
- c. some monkeys riding _____
- d. all monkeys eating _____
- e. last child _____
- . 5-Digit Series: 22 ()
- 2-7-3-6-8 _____ (C or exact response)
- 4-1-3-5-9 _____
- . Match initial sound: (C or I) 23 ()
- 1. mouse _____
- 2. rabbit _____
- . Remember 5 colors: 24 ()
- R G B Y W _____
- . Define words: 25 ()
- a. apple _____
- b. fish _____
- c. hammer _____
- d. chair _____
- . Copy 3-line cross 26 ()
- . Look Alikes: 10. L M R 11. L M R 12. L M R 13. L M R 27 ()





North Carolina Department of Human Resources

Division of Health Services

North Carolina Psychoeducational Screening Test (PET)
Statewide Prekindergarten Screening ProgramRECORD SHEET INSTRUCTIONS

General Instructions: DHS Form 2181 has been developed for use by trained personnel in the Statewide Prekindergarten Screening Program.

Specific Instructions:

- Telephone: Record home phone number or a number where parent/guardian may be reached.
- Sex: X appropriate gender
- Race: Record race as reported by parent/guardian
- Time: Record time of the beginning and end of the screening.
- DOS: Record date of screening
- DOB: Record child's date of birth
- Age: Record child's age in years, months and days
- Code: The code contains up to 12 digits, thus: 00/00/S-00/00000. The first digit indicates the assigned DEC number. The second digit reflects the last two digits of the fiscal year. The third is an S- designating Screening followed by the number of the screening team according to its county affiliation. County numbers are assigned alphabetically. The fourth digit indicates the child's assigned number.
- Location: Cite of screening, e.g., specific school, church, DEC
- County: Record county.
- Test, Retest: Indicate whether the administration of the instrument is an initial test (=1) or subsequent to initial one, retest (=2).
- UPG: Indicate an unusual pencil grasp
- (): Indicate child's score for corresponding item. If item not administered, put NA.

See examiner instruction sheet for information regarding how to record responses on individual items.

Child's Name _____

DOS: _____
 DOB: _____
 Sex: M _____
 Age: _____ yrs. _____ mos. _____ dys.
 Code: _____
 Location: _____
 County: _____

TOTAL FUNCTION SCORE		SCREENING STATUS	
GROSS MOTOR		Satisfactory 2 Not Yet, Below 2	
VISUAL MEMORY		Advanced 6 Satisfactory 3-5 Not Yet, Below 3	
AUDITORY PERCEPTION		Advanced 6 Satisfactory 4-5 Not Yet, Below 4	
AUDITORY MEMORY		Advanced 8 Satisfactory 4-7 Not Yet, Below 4	
VISUAL MOTOR PERFORMANCE		Advanced 10-11 Satisfactory 4-9 Not Yet, Below 4	
CONCEPT DEVELOPMENT		Advanced 11-13 Satisfactory 6-10 Not Yet, Below 6	
LANGUAGE DEVELOPMENT		Advanced 16-17 Satisfactory 11-15 Not Yet, Below 11	

GROSS MOTOR		GROSS MOTOR	
0 1 2	0 1	Balance on one foot	Hop
9	10		
VISUAL MEMORY		VISUAL MEMORY	
0 1 2	0 1 2	Remember 3 Colors	Remember 5 Colors
16	18 24		
AUDITORY PERCEPTION		AUDITORY PERCEPTION	
0 1 2	0 1 2	Clap Hands 11-a	Match Initial Sound 23
11-a	11-b		
AUDITORY MEMORY		AUDITORY MEMORY	
0 1 2	0 1 2	3 Commands 8	5 Digit Series 22
7	12		
VISUAL MOTOR PERFORMANCE		VISUAL MOTOR PERFORMANCE	
0 1 2	0 1 2	Copy Cross 3-a	Copy Tri-angle 19
12 34	19		
CONCEPT DEVELOPMENT		CONCEPT DEVELOPMENT	
0 1 2	0 1 2	Wheel 4-a, b	How Things Alike 20
4-a, b	21		
LANGUAGE DEVELOPMENT		LANGUAGE DEVELOPMENT	
0 1 2	0 1 2	Colors Name 15N-a, b, c	Define Words 25
1-N	2N-a, b, c		

DHS Form 2306 REV. 11/75
 Prekindergarten Screening

PREKINDERGARTEN ITEMS
 SUPPLEMENTAL ITEMS

North Carolina Department of Human Resources
Division of Health Services

Developmental Profile
Statewide Prekindergarten Screening Program

INSTRUCTIONS FOR USE OF FORM

Fill in heading with appropriate information:

DOS: Date of Screening
DOB: Child's Date of Birth

Code: The code contains up to 12 digits, thus: 00/00/S-00/00000. The first digit indicates the assigned DEC number. The second digit reflects the last two digits of the fiscal year. The third is an S- designating Screening followed by the number of the screening team according to its county affiliation. County numbers are assigned alphabetically. The fourth digit indicates the child's assigned number.

Information on the Developmental Profile is obtained from the Psychoeducational Screening Test. Numerals indicate the degree of appropriateness of a child's response (the higher numeral corresponding to a more appropriate response). Numerals are summed to determine the Total Function Score of a given developmental area. At the far right are criteria indicating the nature of the Total Function Score.

Retention: Hold until further directed by DHS.

APPENDIX B

Parental Permission Form

PARENTAL PERMISSION

Dear Parent,

With the cooperation of the Statewide Prekindergarten Screening Program, I am attempting to gather information for a Masters Thesis. My thesis attempts to compare the two main screening instruments used in the screening program. The N. C. Psychoeducational Screening Test (PET) and the McCarthy Scales of Children's Abilities. Also, a subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) will be used. In addition to aiding me in my efforts, the data collected will enhance the understanding of the overall screening effort.

Although your child ordinarily would not be evaluated further, I would appreciate him or her returning for further testing with the McCarthy and the Auditory Reception of the ITPA Subtest.

The McCarthy includes the following areas:

- _____ Verbal
- _____ Perceptual Performance
- _____ Quantitative
- _____ Memory
- _____ Motor

Please understand that all results will be shared with you, and that no information gathered will be released without your written consent. Upon your request, a copy of the results will be given to you.

I have received a description of each area of the McCarthy. I understand that ordinarily my child would not be evaluated further. I agree to allow _____ (child's name) to participate in this further evaluation with the McCarthy and IPTA Subtest and I am aware that the results of my child's performance on the PET and the McCarthy will be compared.

Signature of Parent or Guardian

(PUT TIME AND DATE OF APPOINTMENT ON PIR)

LIBRARY
Appalachian State University
Boone, North Carolina