DOCUMENT RESUME

ED 101 298 CS 001 588

AUTHOR Satz, Paul; And Others
TITLE Some Predictive Antecedents of Specific Reading Disability: A Two-, Three-, and Four-year Follow-up.
INSTITUTION Florida Univ., Gainesville.
PUB DATE 74
NOTE 54p.; The Hyman Blumberg Symposium on Research in Early Childhood Education, Johns Hopkins University, 1974

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS *Dyslexia; Elementary Education; Elementary School Students; Kindergarten Children; *Longitudinal Studies; *Predictive Ability (Testing); Predictive Validity; Reading Ability; Reading Achievement; Reading Development; Reading Difficulty; *Reading Research; *Theories

ABSTRACT

The purposes of this research project are (1) to test a theory which purports to identify the predictive antecedents (i.e., precursors) of developmental dyslexia (specific reading disability) several years before the disorder is clinically evident, and (2) to evaluate the mechanism which is postulated to underlie and influence later developmental changes in this disorder. The study focused on the predictive accuracy of a developmental-neuropsychological test battery which was given to a total population of white boys during kindergarten (1970) in order to forecast their reading achievement in subsequent years. The predictive accuracy of the test battery was determined by assessing reading in this population at the end of grades one, two, and three. The results revealed that performance on these tests during kindergarten was extremely predictive of the child's reading group membership in later grades, particularly with those children destined to become severely disabled or superior readers. These findings lend substantial support to the validity and utility of an early detection or warning system that could be administered economically before the child begins formal reading and at a time when he is less subject to the effects of academic failure. (RB)

ED101298

U S DEPARTMENT OF HEALTH. EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

The Hyman Blumberg Symposium on Research in Early Childhood Education
Johns Hopkins University 1974

Some Predictive Antecedents of Specific Reading Disability: A Two-, Three- and Four-year Follow-up[1]

Paul Satz, Janette Friel and Fran Rudegeair
University of Florida

PS 007 588

ERIC

BEST COPY AVAILABLE

The present chapter is addressed to a longitudinal review (Years 2 and 3) and follow-up evaluation (Year 4) of the predictive validity of a developmental neuropsychological test battery that was administered to an original population (N = 497) of white boys at the beginning of kindergarten in 1970 in Alachua County, Florida. The children represented virtually all of the white male population (96%) enrolled in the 20 schools (14 urban, 6 rural) in 1970. In addition, the chapter is addressed to a three year cross-validation follow-up on an additional sample of 181 white boys who were tested at the beginning of kindergarten in 1971 with the same battery to forecast reading achievement at the end of Grade 2 (1974).

The purposes of this research project are twofold: (1) to test a theory (Satz and Sparrow, 1970; Satz, Rardin and Ross, 1971; Satz and Van Nostrand, 1973) which purports to identify the predictive antecedents (i.e., precursors) of developmental dyslexia (specific reading disability) several years before the disorder is clinically evident and (2) to evaluate the mechanism which is postulated to underlie and influence later developmental changes in this disorder.

The present chapter is addressed to the first objective (i.e., early detection) because not enough time has elapsed in the longitudinal project (four years) in order to investigate the developmental course of this disorder after the reading disability has been diagnosed and confirmed. However, preliminary data on developmental changes within the first three years (K-G2) are reported in a recent paper (Satz, Friel and Rudegeair, 1974).

The present research springs from the need for an early and valid detection or 'warning system' that could be administered before the child begins formal reading --at a time when his central nervous system may be more plastic and responsive to change --and at a time when he is less subject to the

ERIC

BEST COPY AVAILABLE

shattering effects of repeated academic failure. The need for early detection research in the field of childhood reading disorders has long been recognized but has been marked by a paucity of well-controlled longterm longitudinal studies (Critchley, 1968; Money, 1962; Eisenberg, 1966; Gallagher and Bradley, 1972: Kline, 1962; Ames, 1968; Satz, 1973). The advantages of an early identification and remediation program are highlighted by the recent survey by Keeney and Keeney (1968). It was shown that ". . .when the diagnosis of dyslexia was made in the first two grades of school nearly 82 percent of the students could be brought up to their normal classroom work, while only 46 percent of the dyslexic problems identified in the third grade were remediated and only 10 to 15 percent of those observed in grades five to seven could be helped when the diagnosis of learning problems was made at those grade levels" (Strag, 1972, p. 52).

Experimental studies with humans and infrahumans also suggest that the child may be more responsive to environmental stimulation (e.g., remediation) during that period in which the brain is maturing and when behavior is less differentiated (Caldwell, 1968). Bloom (1964) has shown that variations in the environment have their greatest quantitative effect on a characteristic (e.g., speech) at its most rapid period of change (i.e., ages 2-10) and the least effect at the least rapid period of change (i.e., ages 11-15). Infrahuman studies also suggest that organization can be strongly modified only when active processes of organization are underway and that when facilitated, they progressively inhibit attempts at reorganization (Scott, 1968).

Despite the obvious advantages for an early intervention thrust, they must be based on a valid and efficient detection procedure, especially when the detection measures are applied during preschool -- and <u>before</u> formal reading instruction is commenced. Within such a prevention context, certain types

BEST COPY AVAILABLE

of prediction errors are crucial to the utility of an 'early warning' system--e.g., the false positive and negative signs. To initiate an intervention program for test-classified high risk children in kindergarten may be fruitless if the majority of true high risk children are missed by the tests (i.e., false negatives). Equally serious and perhaps more risky is the case where an intervention program is based on erroneously test-classified positives (high risk) who in three years would have become average to superior readers without remediation (i.e., false positives). These prediction errors can occur with tests despite an apparent demonstration of validity via more simplified descriptive univariate tests of significance (e.g., t-test, Chi-square). In other words, the experimental validation of an early detection battery must incorporate a multivariate design in which multiple measurements are made on the same subjects over time with sufficient temporal separation between the initial test probe (e.g., kindergarten) and the criterion reading assessment in later years (e.g., Grades 1-3). This type of design should utilize a longitudinal framework based preferably on a total population rather than smaller sample of children to offset the potential attrition effects over time and to provide more reliable base rate estimates of reading disability in the designated population. Further, the selection of a more homogeneous population (e.g., white boys) provides an opportunity to obtain a higher at-risk group for later reading disability (i.e., dyslexia) without confounding sex, race or cultural variables (Bentzen, 1963; Satz, 1973). A final methodological requirement, essential to the evaluation of an early detection battery, concerns the use of a separate group of children upon which to cross-validate the predictive validity of the tests administered to the standardization population--and for the same interval of time between the initial probe and the reading criterion assessment.

BEST COPY AVAILABLE

Failure to incorporate the preceding methodological factors into an early detection study would surely limit if not invalidate the results. A review of the current early detection literature reveals an alarming disregard of these problems and a scathing indictment of the area (Gallagher and Bradley, 1972).

An equally serious problem in the early detection literature concerns the lack of a theoretical framework in which to conceptualize the nature of the disorder--namely dyslexia--and its antecedent precursors. Without a testable theory one is bound to be restricted in the selection of a test battery which purports to identify the potentially high risk child. Although a theory represents a framework in which to organize diverse sets of data--often seemingly unrelated--it's ultimate validation must rest on empirical verification over time. In the present context, a theory which postulates the precursors of developmental dyslexia, before the child begins formal reading, can only be evaluated empirically within a longitudinal framework. Further, if support for the theory can be replicated and cross validated in later years, then information is available for application (i.e., early intervention) and for further understanding of this complex disorder of childhood (i.e., etiology and mechanism).

The preceding theoretical rationale reflects some of the basic assumptions that underlie all behavioral research (Lindgren and Byrne, 1971, pp. 18-19):

1. All behavior is caused--that is, it is determined by and is the necessary consequence of antecedent events.
2. The causes of behavior are multiple.
3. The causal factors leading to the variance in the behavior being studied must be identified, if valid prinicples or generalizations are to be developed.
4. The principles or generalizations that govern causal factors or account for variations in behavior are simpler than the original data on which they are based.

BEST COPY AVAILABLE

5. The test of whether prinicples or generalizations are valid is whether they can be used to predict behavior.

Theory. The theory which provides the methodological and conceptual framework for the current longitudinal research has been discussed in previous papers (Satz and Sparrow, 1970; Satz, Rardin and Ross, 1971; Satz and Van Nostrand, 1973). Briefly, it postulates that developmental dyslexia reflects a lag in the maturation of the brain which delays differentially those skills which are in primary ascendancy at different chronological ages. Consequently, those skills which during childhood develop ontogenetically earlier (e.g., visual-perceptual and cross-modal sensory integration) are more likely to be delayed in younger children who are maturationally immature. Conversely, those skills which during childhood have a later or slower rate of development (e.g., language and formal operations) are more likely to be delayed in older children who are maturationally immature.

The theory is compatible with those developmental positions which postulate that the child goes through consecutive stages of thought during development, each of which incorporates the processes of the preceding stage into a more complex and hierarchically integrated form of adaptation (Hunt, 1961; Piaget, 1926; Bruner, 1968). Thus, it is predicted that those children, during preschool, who are delayed developmentally in skills which are in primary ascendancy at this stage, will eventually fail in acquiring reading proficiency. It is predicted, however, that these children will eventually "catch up" on these earlier developing skills but will then lag on those more cognitive-linguistic skills which have a slower and later ontogenetic development (Thurstone, 1955; Bloom, 1964). In other words, the theory predicts that the nature of the disorder will vary as a function of the chrongological age of the child.

ERIC

More specifically, the lag in brain maturation is postulated to delay the acquisition of those developmental skills which have been shown to be crucial to the early phases of reading--namely, learning to differentiate graphic symbols (Gibson, 1968) or the perceptual discrimination of letters (Luria, 1966). Both authors recognize an orderly and developmental sequence in which the early phases of reading are characterized by processes of perceptual discrimination and analysis. In this early phase the child must discriminate the distinctive features of letters (e.g., break vs. close, line vs. curve, rotation and reversal) before he can proceed to later phases which require more complex phonetic and linguistic analysis. Smith (1971) also recognizes the importance of learning the distinctive features of written language in the beginning phases of reading, but cautions that fixation at this level will retard the syntactic process of fluent reading.

The theory, in summary, conceptualizes developmental dyslexia as more than a reading disorder *per se*. That is, the disorder is explained as a delay in those crucial early sensori-perceptual and later conceptual-linguistic skills which are intrinsic to the acquisition of reading and which are triggered by a lag in the maturation of the cerebral cortex. (The lag mechanism, being unobservable at the present time, is treated as a hypothetical construct.) In other words, dyslexia (specific reading disability)is seen as a disorder in central processing, the nature of which varies with the chronological age of the child. This delay in central processing is not meant to imply damage, loss of function or impairment. Such terms are more compatiable with a disease model which often implies a static developmental-acquisition course. With respect to early detection of high risk children, the theory predicts that delays in those developmental skills which are in primary ascendancy during preschool (i.e., kindergarten) are more likely to forecast later problems in reading by Grades 2 and 3.

Current Objectives. The following section is divided into two parts, the first of which reviews the second and third year predictive follow-up of those children (N = 497) who were tested at the beginning of kindergarten in Alachua County (1970) and for whom independent criterion reading measures were obtained at the end of Grades 1 (1972) and Grades 2 (1973). The results of these two follow-up studies are reported elsewhere in detail (Satz and Friel, 1974; Satz, Friel and Goebel, in press).[2] The second part of this section is addressed briefly to two recent unpublished studies: the first subpart concerns the fourth year predictive follow-up of this original population of boys (N = 497) for whom independent reading measures were obtained at the end of Grade 3 (1974); the second subpart concerns a three year cross-validation of the weights derived from the standardization population (1970, N = 497) to an additional sample of 181 white boys who were tested at the beginning of kindergarten (1971) with the same battery to forecast reading achievement at the end of Grade 2 (1974).

REVIEW STUDIES (Part 1).

Second Year Follow-up (1970-1972).[3]

During the spring term of 1972 (Grade 1) independent reading measures were obtained for 95% of the boys who were tested at the beginning of kindergarten in 1970 (N = 473). This figure represents an attrition rate of only 5% during the two year interval. The reading measure, which was completed by the individual classroom teachers (within and outside Alachua County), was adapted from the standard grade report for Grades 1-3 of Alachua County, Florida. On this basis, children were classified into four dichotomous reading groups which can be seen in the column totals of Table 1: Severe (N = 18), Mild (N = 55), Average (N = 339) and Superior (N = 61). These frequencies indicate a 15% incidence of reading disability at the end of Grade 1 (73/473) with an incidence of only 4% in the Severe group (18/473). These incidence figures, which were based on preliminary reading assessment in our population (Grade 1, 1971) were later shown to be premature when assessed at the ends of Grades 2 and 3 (see following sections).

---

insert Table 1 about here

Despite the tentative nature of this reading measure, the developmental tests given in 1970 (n = 19) correctly predicted 84.4% of the children into four discrete reading groups (Satz and Friel, 1974). These results, based on a multiple discriminant function analysis (4 group), can be visualized in Table 1 by comparing the test predictions (+, -) in the rows (1970) against the criterion outcomes (n = 4) in the columns (1972). For valid positives the tests correctly predicted all of the 18 Severe cases (V_p = 100%) and 39 of the 55 Mild cases (V_p = 71%). For valid negatives, the tests correctly predicted 284 of the 339 Average readers (V_n = 84%) and 58 of the 61 Superior readers (V_n = 95%). In other words, the predictive accuracy was largely confined to the extreme reading groups with overlap error largely confined to the Mildly Disabled and Average reading groups. In terms of overall hit-rate, the tests correctly predicted 399 (18 + 39 + 284 + 58) of the 473 Ss or 84.4% of the standardization population.[4]

---

insert Table 2 about here

Predictive ranking of tests. A stepwise regression analysis was then computed to determine the ranking of the predictor variables in terms of their criterion discrimination. Table 2 presents the ranking of the most accurate variables along with their cumulative hit frequencies and factorial loadings. Inspection of this table reveals that the Finger Localization Test ranked highest (76%) followed cumulatively by the Recognition-Discrimination Test (77%), Day of Testing (79%) and the Alphabet Recitation Test (82%). The remaining variables contributed an additional increment of less than three percent to the total hit-rate of 84.4%. This table (Table 2) also shows that three of the most discriminating variables loaded on Factor I, which has previously been defined as a general measure of sensori-perceptual-motor-mnemonic ability (Satz and Friel, 1973).[5] In the original factor analysis, four factors emerged (principal axis solution and orthogonal rotation to varimax criterion): Factor I - tests of sensori-perceptual-motor-mnemonic ability (30.7% common variance): Factor II - kindergarten teacher evaluations (16% common variance): Factor III - tests of conceptual-verbal ability (13.4% common variance) and Factor IV - tests of motor dominance and laterality (7.7% common variance). Factor I is felt to tap those skills which are in primary ascendancy during preschool years (i.e., kindergarten). As such, the results were felt to lend preliminary support for those developmental precursors postulated to underlie and forecast subsequent reading achievement (Satz and Van Nostrand, 1973). When the means and standard deviations were examined for each of the kindergarten measures (1970) it was observed that those children destined to reading problems at the end of Grade 1 (1972) were lagging behind the to-be good readers on most of the developmental tests--but particularly on the Factor I tests. In fact, on the Beery Developmental Test of Visual-Motor Integration (Factor I), the High Risk reading group (Severe and Mild) revealed almost a 12 month lag between their chronological age (65.8 months) and performance age (54.4 months). By contrast, the Low Risk reading group (Average and Superior) virtually matched their chronological and performance age on this test (66.4 months vs. 66.5 months, respectively).

Third Year Follow-up (1970-1973).

Despite the encouraging preliminary findings in the two year follow-up (Satz and Friel, 1974) the interval of an additional year provided an opportunity to obtain more objective and stable measures of reading achievement in this population at the end of Grade 2 in 1973 (Satz, Friel and Goebel, in press). It was also felt that the availability of more objective reading measures by Grade 2 (spring) would provide more valid estimates on the incidence of reading problems in this population (white boys) and a comparison of the differential incidence of reading disability, if any, between Grades 1 and 2. Furthermore, with more objective reading measures, independently assessed, it was felt that a more rigorous evaluation of the predictive validity of the kindergarten tests could be made (K-G2), including their discriminative ranking.

With this objective in mind, two different methods of reading assessment were obtained at the end of Grade 2 (1973). The first method was again based on the Classroom Reading Level which was filled out by the individual teachers both within and outside of Alachua County, Florida. Reading forms were returned for 458 of the original 497 Ss which represents 92% of the population three years later (attrition = 8%). Subjects whose reading was assessed at Levels 0-4 (No Readiness through Primer) were designated as the Severely Disabled readers. Those assigned Level 5 (First Reader) comprised the Mildly Disabled reading group. Subjects reading at Levels 6 and 7 (Second Readers) were designated as Average readers, and those reading above that level comprised the the Superior reading group.

The second criterion reading measure was based on Classroom Reading Level and the IOTA. This combined reading measure afforded the advantage of incorporating both the teacher's assessment of reading, based on nearly a year's interaction with the child, and an independent, individually administered, objective reading test. Both of these measures were obtained on 419 of the original population (85%).[6] After converting the raw scores on each measure to T-scores ($\bar{X}$ = 50, s = 10), the distribution was dichotomized again into four reading groups (<u>Severe</u>, <u>Mild</u>, <u>Average</u>, and <u>Superior</u>).

<u>Criterion I Analysis</u>. The first four-group discriminant function analysis (Program DSCRIM) was computed on Classroom Reading Level in order to compare the predictive validity of the tests (1970) against a similar type of reading criterion at the end of Grade 2 (1973). The results of this analysis can be seen in Table 3 by comparing the test prediction (+,-) in the rows (1970) against the criterion outcomes (n = 4) in the columns (1973). Before discussing the predictive hits, it can be seen that the incidence of reading disability (columns) increased substantially between Grades 1and 2 (1972-1973): <u>Severe</u> (N = 54), <u>Mild</u> (N = 66), <u>Average</u> (N = 270), <u>Superior</u> (N = 68). This criterion distribution reveals a 26% overall incidence of reading disability [(54 + 66)/458)] and an incidence of 12% in the <u>Severe</u> group (54/458). Compared to Grade 2 (1972), this represents a three-fold increase in the frequency of <u>Severe</u> cases (1973).

To further assess the validity of this criterion measure (i.e., Classroom Reading Level), scores on the Gates-McGinitie, which was administered to the <u>total</u> population of children in Grade 2 (Spring) in Alachua County, were averaged for each of the reading groups. Vocabulary recognition scores on this test were compatible with the results of Classroom Reading Level [<u>Severe</u> (-12.3 mos.), <u>Mild</u> (-6.5 mos.), <u>Average</u> (+7.9 mos.) and <u>Superior</u> (+18.8 mos.)]. In other words, the <u>Severe</u> group was well over a full grade level behind in reading (where academic year = 10 mos.). In contrast, the <u>Superior</u> group was almost two years ahead in terms of grade level and month of testing. Moreover, the <u>Mild</u> group was only six months behind grade level whereas the <u>Average</u> reading group was approximately eight months ahead.

---

insert Table 3 about here

Table 3 also reveals the predictive hits. For valid positives, the tests correctly predicted 48 of the 54 Severe cases, (V_p = 89%) and 47 of the 66 Mild cases (V_p = 71%). For valid negatives, the tests correctly predicted 196 of the 270 Average readers (V_n = 73%) and 64 of the 68 Superior readers (V_n = 94%). Once again, the predictive accuracy of the tests was largely confined to the extreme reading groups with overlap error largely confined to the Mildly Disabled and Average reading groups. In terms of overall hit-rate, the tests correctly predicted 355 (48 + 47 + 196 + 64) of the 458 Ss or 78% of the population.

---

insert Table 4 about here

Predictive ranking of tests. A stepwise regression analysis was then computed to determine the ranking of the predictor variables in terms of their criterion discrimination. Table 4 presents the ranking of the most accurate variables along with their cumulative hit frequencies and factorial loadings. Inspection of this table reveals that the Finger Localization Test ranked highest (71%) followed cumulatively by the Alphabet Recitation Test (76%), Recognition-Discrimination Test (77%) and Day-of-Testing (77%). The remaining tests contributed an additional increment of less than one percent to the total hit-rate of 78%. This table (Table 4) again shows the same four variables that ranked highest in the two year follow-up (Grade 1, 1972), three of which, in both studies, loaded on Factor I (Table 2). This finding again strengthens the validity of sensori-perceptual-motor-mnemonic abilities in forecasting subsequent reading achievement levels.

---

insert Table 5 about here

Criterion II Analysis. This analysis was based on the predictive accuracy of the tests (n = 19) given at the beginning of kindergarten (1970) to the combined criterion of Classroom Reading Level plus IOTA Word Recognition at the end of Grade 2 (1973). These results can be seen in Table 5 by comparing the test predictions (+, -) in the rows (1970) against the criterion outcomes (n = 4) in the columns (1973). First, the criterion outcomes in the columns revealed an even higher incidence of reading disability: Severe

(N = 67), Mild (N = 77), Average (N = 214), and Superior (N = 61). Although based on a smaller number of Ss, this criterion assessment revealed a 34% overall incidence of reading disability [(67 + 77)/419] and an incidence of 16% in the Severe group (67/419)--which is four times higher than was observed at the end of Grade 1 (1972).[7]

The validity of this combined reading criterion was further assessed by computing the mean scores of the Gates-McGinitie for each of the four reading groups. Vocabulary recognition scores on the Gates were in essential agreement with the reading groups based on Classroom Reading Level and the IOTA [Severe (-12.1 mos.), Mild (-10.7 mos.), Average (+3.3 mos.) and Superior (+16.3 mos.)]. In other words, the Severe group was over a full grade level behind in reading and the Superior group was over a year and a half ahead in terms of grade level and month of testing. Moreover, the Average reading group was essentially at grade level.

The predictive hits can also be seen in Table 5. For valid positives, the tests correctly predicted 61 of the 67 Severe cases (V_p = 91%) and 51 of the 77 Mild cases (V_p = 66%). For valid negatives, the tests correctly predicted 146 of the 214 Average readers (V_n = 68%) and 59 of the 61 Superior readers(V_n = 97%). Once again, the total hits (61 + 51 + 146 + 59 = 317/419 or 76%) were largely confined to the extreme reading groups.

For purposes of brevity, the stepwise regression analysis also revealed the same discriminative ranking of tests that was found in the second and third year follow-up studies using Classroom Reading Level as the criterion. The Finger Localization Test ranked highest followed cumulatively by the Alphabet Recitation Test, Recognition-Discrimination Test and Day-of-Testing. Three of these four variables each loaded significantly (r $\geq$.37) on Factor I.

Two additional findings emerged from this third year follow-up that merit attention for educators. First, it was found that an abbreviated sample of the 19 tests yielded virtually the same predictive accuracy as the full battery. Selection of this abbreviated battery (seven test and one nontest variables) was based on empirical evaluation of the highest ranking tests in each of the stepwise regression analyses across the years.[8] After separate discriminant function (Program DSCRIM) and stepwise regression analyses were computed against the combined reading criterion (Grade 2, 1973), it was found that eight of the tests given in 1970 detected almost all of those children destined to extremes in the reading distribution at the end of Grade

BEST COPY AVAILABLE

2 (i.e., Severe and Superior groups). Also, the same three Factor I tests again ranked highest (Finger Localization, Alphabet Recitation, and Recognition-Discrimination). This finding, which minimizes both time and cost factors associated with test administration, without sacrificing accuracy, greatly increases its applicability for educators concerned with large scale early screening assessment.

The second finding, of possible interest to educators, concerned the utility of this test battery (abbreviated or standard) to provide a decisional basis upon which to initiate or withhold treatment for an individual child based on his test scores at the beginning of kindergarten. This decisional process, which was prompted by the consistently high predictive accuracy in the extreme reading groups across years, is determined by computing the conditional probability of the differential test signs (+, -). These conditional probability values are based on the inverse probablities of the test signs (valid and false positives and valid and false negatives) and the base rates of reading competency in this population (Meehl and Rosen, 1955; Satz, Fennell and Reilly, 1970). To compute the conditional probabilities for each of the test signs, the (+) and (-) rows in Table 5 were subdivided (on the basis of the original four group DSCRIM analysis) to generate four levels of test decisions [(++), (+), (-), (--)] . Table 6 reveals this 4 X 4 contingency table in which test signs (and decisions) are represented by rows and the outcomes are represented by columns. This table thus allows the educator to determine the probability that a given child is destined to reading disability or competency, in three years, given that his test scores fall in the (++) , (+) , (-) , (--) range. As such, it also provides him with the likelihood that intervention should be instituted or withheld.

---

insert Table 6 about here

The results of this analysis which can be seen in Table 6, indicate that with a severe high risk composite sign (++) , the decision to initiate treatment would have been correct in 82% of the cases; further, this treatment decision would have included 50 of the 67 potential Severe cases (75%) and none of the potential Superior readers in the intervention program. By contrast, the decision to intitiate treatment, given a less severe high risk composite sign (+) , would have been correct only 44% of the time. In other words, extension of treatment services to this group would have yielded only 11 more of the potential Severe cases and 33 of the potential Mild cases, but it would also have involved treatment for 55 children who would not have needed it (false positives). Thus, initiation of treatment for this composite test sign (+) would have resulted in treating more children who didn't need it than did.

This table also shows that the decision to withhold treatment (NT) would have been correct 77% of the time given a mild low risk sign (-) and 96% of the time given a very low risk sign (--). The only risk, given a mild low risk sign (-) is that treatment would have been withheld for a small number of potentially Severe (N = 5) and Mild (N = 23) reading cases. This decision risk, however, would have been virtually eliminated by adopting a more conservative no treatment policy based on very low risk signs (--) .

The latter findings, in summary, illustrate the potential usefulness of a detection procedure which, while brief and economical to administer, can also generate the conditional outcome probabilities (i.e., reading level) for an individual child during kindergarten. Within this framework, educators could base their treatment strategies on a number of factors including the incidence of severe high risk children in the school, available resources, and the risks associated with intervention (false positive errors) vs. non-intervention (false negative errors). If treatment resources were limited and the number of high risk children was large, then surely a conservative decisional strategy should be adopted for treatment [i.e., (++) only]. With respect to the present study, this treatment strategy would have been correct for 82% of the cases; moreover, it would have detected or included 75% of the potential Severe cases in the treatment program and none of the potential Superior readers!

On the other hand, if multiple treatment resources were available, educators might choose to adopt a more liberal intervention policy, given the present utility table i.e., [(++) and (+)] . While this strategy would pick up most of the potential Severe cases, it would also entail considerable efforts and time for many children who would not have needed such help (i.e., false positives). But then again, these latter children, through intervention, may have eventually become superior readers.

It should be apparent that the final decision to initiate treatment or not should be made by the educator who has to consider a multiplicity of factors. The utility table presented in this section merely simplifies this decisional task by generating the likelihood probabilities and risks for either decision (T or NT) in each individual case.

These outcome probabilities, while generated by the test signs and base rate incidence of reading disability in the population, are of little use to educators if the predictive validity of the tests cannot be replicated across years within Ss or cross-validated across years on additional Ss. The following section is addressed to this problem.

NEW STUDIES (Part II)

Fourth Year Follow-up (1970-1974).

In view of the high predictive accuracy of the developmental tests in the second and third year follow-up studies (Grade 1, 1972 and Grade 2, 1973, respectively), the question was raised as to whether the accuracy would attenuate as the test-criterion interval increased to four years. After all, an interval of this length would seemingly increase the number of uncontrolled environmental, growth and treatment factors which could produce changes in criterion group membership and thus lower the predictive accuracy of the tests given in 1970. In fact, by the end of Grade 3 (Year 4, 1974), it was learned that the vast majority of the Severe cases were receiving remedial help in the schools. With this knowledge it also became necessary to determine whether the incidence of reading problems decreased between the third and fourth follow-up years and if not, to determine whether the same Ss again fell in the same criterion groups.

---

insert Table 7 about here

Classroom Reading Level was again used to assign Ss into the four discrete reading criterion groups at the end of Grade 3 (1974). These reading forms, which were independently filled out by the individual classroom teachers, were completed for 459 of the original 497 Ss (1970). This figure represents almost 93% of the original population now residing in Florida, the United States and abroad.[9] The obtained reading distribution can be seen in the column totals of Table 7: Severe (N = 55), Mild (N = 93), Average (N = 234) and Superior (N = 77). This distribution reveals a 32% overall incidence of reading disability [(55 + 93)/459] and an incidence of 12% in the Severe group (55/459). These frequencies can be more easily visualized in Figure 1 by comparing the longitudinal changes in the incidence of reading disability (based on Classroom Reading Level) from the end of kindergarten (1971) to the end of Grade 3 (1974). Inspection of this figure reveals an increasing incidence of overall reading disability from Grades 1-3 but with a leveling off in the Severe reading group between the ends of Grades 2 and 3 (1973-1974). In other words, the increase in overall reading disability in the last follow-up year (Grades 2-3) was due to an additional increase (N = 27) in the number of Mild reading cases. It should be kept in mind that this increasing incidence of reading problems across years has occurred despite the equally increasing intervention of the schools to provide remediation for these disabled readers (Severe and Mild).

insert figure 1 about here

Of particular interest are the changes in reading disability between the ends of Grades 2 and 3 (1973-1974). These changes can be visualized by comparing the column totals in Tables 3 and 7. The totals in the Severe group remained essentially unchanged (Grade 2, N = 54; Grade 2, N = 55). However, there were S changes within groups. Of the 54 Ss in the Severe group (Grade 2), 38 remained in Grade 3. Of the 16 Ss who changed from the Severe group, 10 were reassigned to the Mild group in Grade 3 and six to the Average reading group. Despite these changes in the Severe group between Grades 2 and 3, there were an additional 17 Ss who were reassigned to the Severe group in Grade 3 from the remaining groups in Grade 2 (1973). Further, 13 of these 17 Ss came from the Mild group and four from the Average reading group.

Figure 1 also shows that there was a six percent increase in the incidence of overall reading disability between the ends of Grade 2 and 3 (1973-1974). This increase was due to the reassignment of 27 additional cases to the Mild group (Grade 3), 10 of whom came from the Severe group (Grade 2) and 17 of whom came from the Average reading group (Grade 2).

The next question was whether this reassignment of Ss into different criterion groups (Grades 2-3), plus the uncontrolled treatment intervention in the schools, would lower the predictive accuracy of the tests given in 1970 (n = 19). To answer this question, a four-group discriminant function analysis (Program DSCRIM) was again computed on the test scores (1970) against the criterion reading membership (1974). These results can be seen in Table 7 by comparing the test predictions (+,-) in the rows (1970) against the criterion outcomes in the columns (1974). For valid positives, the tests correctly predicted 50 of the 55 Severe cases (V_p = 91%) and 64 of the 93 Mild cases (V_p = 69%). For valid negatives, the tests correctly predicted 163 of the 234 Average readers (V_n = 70%) and 72 of the 77 Superior readers (V_n = 94%). Once again, the predictive accuracy of the tests was high, particularly in the more important extreme groups, despite an interval of almost four years from test probe to criterion determination. In terms of overall hit-rate, the tests correctly predicted 349 (50 + 64 + 163 + 72) of the 459 Ss or 76% of the population.[10] This hit-rate is almost identical to the results of the third year follow-up (Grade 2, 1973). Of particular importance is the fact that the tests continued to show greater accuracy for those children destined to extremes of the reading distribution in later years. It is these children that educators must identify, hopefully during primary grades, in order to institute more effective programs for future growth and development. In this context, the potentially gifted child should be given the same consideration as the high risk child. To stifle creativity or achievement in a gifted child by improper placement or lack of stimulation is almost as negligent as the failure to identify and help the immature child at a critical time in his delayed development.

Predictive ranking of tests. A stepwise regression analysis was also computed to determine the ranking of the predictor variables in terms of their criterion discrimination. Table 8 presents the ranking of the most accurate variables along with their cumulative hit frequencies and factorial loadings. Inspection of this table reveals that the Finger Localization Test ranked highest (70%) followed cumulatively by the Alphabet Recitation Test (74%), Recognition-Discrimination Test (77%) and Dichotic-Listening Test (Total Recall) (77%). The remaining tests contributed an additional increment of less than two percent to the total hit-rate of 78.2%.[11] This table (Table 8) again shows that three of the same tests ranked highest in predictive discrimination across each of the follow-up years. Further, that each of these tests have loaded on Factor I. In the present study, all of the four top ranking tests loaded on Factor I which again buttresses the significance of this factor in forecasting subsequent reading achievement.

Insert Table 8 about here

Cross-Validation: Third Year Follow-up (1971-1974)

The following section summarizes the results of the final and major evaluation of the predictive test battery given at the beginning of kindergarten in 1970. Despite the high predictive accuracy which was demonstrated across each of the follow-up years (Grades 1-3), for both the standard and abbreviated test batteries, the results must be interpreted as only tentative until replicated on an independent group of children with the same time interval between test probe and criterion determination. This methodological requirement must be met notwithstanding the high predictive validity and internal consistency (predictive ranking) of the tests across follow-up years.

In the present study, it was decided to cross-validate the test weights (lambdas) derived from the third year follow-up of the standardization population (Grade 2) against a new sample of children who were tested at the beginning of kindergarten in 1971 and whose criterion reading scores were obtained three years later at the end of Grade 2 (1974).

This cross-validation sample included 181 Ss who represented the total population of white kindergarten boys enrolled in five of the largest elementary schools in Alachua County (based on the 1970 census). They were individually tested, under identical conditions to the standardization population in 1970, in a mobile testing laboratory with four assessment modules. The only difference in this new sample was the fact that the children were all selected from urban schools.

This additional sample was follwed for three years until they reached the end of Grade 2 (1974) at which time reading assessments were made. For comparison purposes to the third year follow-up study (standardization group), two types of reading assessment were again employed: Criterion I (Classroom Reading Level) and Criterion II (Classroom Reading Level plus IOTA Word Recognition).

Criterion I analysis. This analysis was based on the test scores of the cross-validation group during kindergarten (1971), multiplied by the individual lambda weights derived from the standardization discriminant function analysis (Grade 2), against the criterion of Classroom Reading Level assessed at the end of Grade 2 for the cross-validation group (1974). Reading measures were obtained on 175 of the original 181 Ss or 96% of the sample. The resulting distribution of reading groups can be seen in the column totals of Table 9: Severe (N = 18), Mild (N = 25), Average (N = 92) and Superior (N = 40). This distribution reveals a 25% overall incidence of reading disability [(18 + 25/175] and an incidence of 10.3% in the Severe group (18/175). These percentages are almost identical to the figures obtained with the standardization population at the end of Grade 2 (1973) [Reference Table 3 and Figure 1].

---

insert Table 9 about here

The predictive hits of the abbreviated test battery (n = 8) can be seen in Table 9. For valid positives, the tests correctly predicted 16 of the 18 Severe cases (V_p = 89%) but only 9 of the 25 Mild cases (V_p = 36%). However, for valid negatives, the tests correctly predicted 64 of the 92 Average readers (V_n = 70%) and 37 of the 40 Superior readers (V_n = 93%). These detection outcomes are virtually the same as with the standardization population for the second, third and fourth follow-up years (Grades 1-3). That is, the tests were again more predictive of the extreme groups in the reading distribution (i.e., Severe and Superior). In terms of overall hit-rate, the tests correctly predicted 126 (16 + 9 + 64 + 37) of the 175 Ss or 72% of the sample. This predictive accuracy, based on weights derived from a different population and with a test-criterion interval of nearly three years, lends convincing support for the intrinsic validity of the tests. In fact, the overall hit-rate was even higher when the full test battery was employed (H_T = 76%); however, the full battery failed to detect as many severe high risk children. In other words, its false negative rate was higher than its false positive rate.

---

insert Table 10 about here

Predictive ranking of tests. A stepwise regression analysis was computed on the full battery to see if the discriminative ranking of the tests remained essentially the same for this cross-validation sample of children. The results can be seen in Table 10. The Finger Localization Test ranked highest (66.9%) followed

ERIC

cumulatively by the Embedded Figures Test (68.2%), the WISC Similarities Subtest (70.9%) and the Finger Tapping Test (Total) (71.5%). The remaining tests contributed an additional increment of less than six percent to the total hit-rate of 77.5% (using uncross-validated weights). It is interesting to note that the Finger Localization Test again ranked highest in terms of criterion discrimination; however, the following rankings varied from the standardization studies in terms of tests and factorial loadings. Although the two top ranking tests were again associated with Factor I, the next two were associated with Factor III (Similarities) and Factor IV (Finger Tapping Test). This change in discriminative ranking must reflect subtle differences in the composition of the cross-validation group as compared to the standardization population. One difference already mentioned was the selection bias in favor of urban schools. A second difference might relate to the expected discrepancy between a sample (N = 181) and a population (N = 497). Regardless of the reason for this difference, it does suggest that the differential discriminability of the tests will vary, in part, as a function of the population or sample selected. However, if the battery consists of at least eight tests, the overall hit-rate should not be affected.

----
insert Table 11 about here

Criterion II analysis. This analysis was based on the test scores of the cross-validation group during kindergarten (1971), multiplied by the individual lambda weights derived from the standardization discriminant function analysis (Grade 2), against the combined criterion of Classroom Reading Level and IOTA Word Recognition assessed at the end of Grade 2 for the cross-validation group (1974). The IOTA was individually administered to those children who still resided in Florida and Georgia by the end of Grade 2. For this reason, criterion information was available on only 151 of the original 181 Ss (83%). Inspection of the column totals in Table 11 reveals a 40% overall incidence of reading disability in this sample [(25 + 36)/151] and a 16% incidence in the Severe group (25/151). The higher overall incidence in this sample may simply reflect sampling bias due to the smaller N. However, the percentage of Severe cases (16%) was identical to the standardization group (Grade 2) for the combined reading criterion (Table 5).

The predictive hits against this additional criterion can also seen in Table 11. For valid positives, the tests correctly predicted 19 of the 25 Severe cases (V_p = 76%) and 19 of the 36 Mild cases (V_p = 53%). For valid negatives, the tests correctly predicted 44 of the 60 Average readers (V_n =73%) and 29 of the 30 Superior readers (V_n = 97%). In other words, the tests correctly predicted 111 (19 + 19 + 44 + 29) of the 151 Ss or 74% of the cross-validation sample. Again, the accuracy was largely confined to the extreme groups (Severe and Superior). This hit-rate of 74%, based on the lambda weights from a different standardization group and with a test-criterion interval of nearly three years, provides additional support for the predictive validity of these kindergarten measures. In fact, the shrinkage in this cross-validation analysis was only two percent when compared to the hit-rate for the third year standardization follow-up, using this combined criterion (Table 5).

For purposes of brevity, the stepwise regression analysis (full battery) revealed essentially the same discriminative ranking of tests for this combined reading criterion as for Classroom Reading Level (Table 10). The Finger Localization Test (Factor I) again ranked highest followed cumulatively by the Embedded Figures Test (Factor I), WISC Similarities Subtest (Factor III) and Socio-economic Level (Factor II). The only change in this ranking was the substitution of Socio-economic Level for the Finger Tapping Test. However, the discriminative ranking of tests changed somewhat between the standardization group and the cross-validation group analyses. This change may well reflect sampling differences between the two groups. Nevertheless, the Finger Localization Test (Factor I) retained its primary ranking across follow-up years and groups lending further support to the validity of this factor as a precursor to later reading disability.

CONCLUDING REMARKS

The preceding discussion has focused on the predictive accuracy of a developmental-neuropsychological test battery which was given to a total population of white boys during kindergarten (1970) in order to forecast their reading achievement in subsequent years. This was accomplished by assessing reading in this population independently at the end of Grades 1, 2 and 3 (1972-1974). The results, based on systematic evaluation at yearly follow-up periods, revealed that performance on these tests during kindergarten was extremely predictive of the child's reading group membership in later grades--particularly those children destined to become Severely Disabled or Superior readers. Also, it was shown that this accuracy was largely accounted for by a small number of tests which consistently ranked highest in predictive discriminability. These latter tests represented measures of sensori-perceptual-motor-mnemonic skills which were postulated to be in rapid development during preschool years and crucial to the early phases of reading. It was also shown that these predictive tests could be applied in educational settings as detection procedures for early intervention decisions. A utility matrix was presented to illustrate this concept of statistical decision theory. Finally, it was shown that the test weights derived from the standardization population (K-G2) held up under cross-validation for a separate group of children who were followed for three years (K-G2).

These findings, in summary, lend substantial support to the validity and utility of an early detection or 'warning system' that could be administered economically before the child begins formal reading--at a time when his central nervous system may be more plastic and responsive to change--and at a time when he is less subject to the shattering effects of repeated academic failure. One of the real tradegies in our educational system has been the tendency to institute the intervention program only after the reading disability has become clinically manifest, at which time secondary behavioral and psychological problems are often confounded with the underlying handicap (de Hirsch and Jansky, 1966). It may be that these secondary problems serve to exacerbate if not reinforce the vicious cycle of frustration and failure for the child.

This problem of delayed intervention is probably based on multiple factors including limited diagnostic and treatment resources and, sometimes, the naively optimistic view that the slow starter will eventually "catch up" if left alone. The incidence changes across years in the present study should provide a sobering contrast to this view. The incidence of overall reading disability has continued to increase yearly in our longitudinal population despite the increasing intervention efforts of the schools during this same time period (reference Figure 1). Furthermore, although the incidence of severe reading disability seems to have plateaued between Grades 2 and 3, it has never decreased--again, despite the remedial programs for these children in the schools.

The high predictive accuracy of the test battery (standard and abbreviated) has two interesting implications concerning the nature of this disorder--namely, dyslexia. First, the fact that these kindergarten tests assessed a variety of developmental skills which were shown to forecast subsequent reading competency suggests that the disorder is the consequence of antecedent events which are present in the child as early as kindergarten--and most probably before. Second, the fact that these antecedent factors comprised nonreading skills suggests that dyslexia is more than a reading disorder *per se*. This latter finding is most compatible with those theories that conceptualize dyslexia as a disorder in central processing (Chalfont and Scheffelin, 1969).

One instructive point to be drawn from these comments is that faulty methods of instruction in the schools cannot be singled out as the primary cause of reading failure today (Furth, 1972). The results of the present study indicate that there are precursors which are reflected in the behavior of the child--at least as early as kindergarten--which forecast reading competency *before* formal reading instruction or remediation is commenced. This statement, however, is not meant to exclude an interaction between the child's readiness and the type of instruction involved. Certainly, faulty instruction will serve to intensify the problems of a high risk child. This point has never been disputed.

But what are these precursors or antecedents that forecast future reading competency? In the present context, these precursors seemed to reflect a general developmental readiness in perception, cognition, language and memory. There was no one particularly striking function or skill area that stood out. For those children destined to become severely disabled readers in four years, their performance during kindergarten was depressed or delayed on almost all of the developmental tests--especially on the most discriminative predictors (i.e., Finger Localization, Recognition-Discrimination, Alphabet Recitation). In this respect, the precursors suggest some type of developmental unreadiness or immaturity. For example, during kindergarten assessment, this group of severely high risk children could only recite parts of the alphabet, their visual-motor integration on the Beery was approximately 10 months behind their chronological age. Similar delays in somatosensory integration, perceptual discrimination, auditory discrimination and verbal-cognitive processing also characterized the performance of these severe high risk children.

By contrast, those children destined to become superior readers in three or four years were advanced on almost all of the developmental tests given in kindergarten. This comparison can be quickly visualized by inspection of Table 12 which presents the means for each of the tests in the Abbreviated Battery for each of the reading groups at the end of Grade 2. The potential superior readers were advanced in terms of visual-motor, perceptual, somatosensory and language development--and they were intellectually much brighter (IQ = 113 vs. 92). In fact, those children destined to superior reading levels in four years were advanced developmentally whether compared within Ss (to their chronological age) or between Ss (to other reading groups). In fact, Table 12 shows that as one moved from the Severe group to the Mild, Average and Superior reading groups, that overall performance increased in linear additive increments. There was no special pattern (non linear) that characterized any of the groups except an increasing developmental integration in perception, cognition, language and memory as one proceeded from the high to the low risk groups.

This latter finding should dispel any premature attempts to explain specific reading disability as a primary lag or defect in one particular modality or function (e.g., auditory sequencing or memory, etc.). The problem is unfortunately not that simple, particularly when viewed within a longitudinal framework. Yet, considerable controversy continues to exist concerning the nature of the disorder--namely, dyslexia and it's etiology. All too often these terms are used interchangeably or are poorly defined. Similar problems could occur in the present study if one were to treat the developmental precursors as causal antecedent events. At best, the precursors represent phenotypic behavioral expressions of events intrinsic to the child which are probably mediated by more basic brain mechanisms. Hence, the elusive and complex nature of the problem.

At present, the concept of a developmental readiness seems to represent the most parsimonious and least pernicious way to describe these behavioral differences (precursors) that characterize children destined in later years to extremes in the reading distribution. This concept, which has long been proposed by others (de Hirsch and Jansky, 1966; Critchley, 1968; Money, 1962) is certainly compatible with the theory proposed in this paper (Satz and Van Nostrand, 1973). However, the present results do not permit a test of the mechanism which is postulated to underlie this disorder--namely, a lag in the maturation of the cerebral cortex (Satz and Van Nostrand, 1973). This formulation must still be treated as a hypothetical construct, the mechanism of which is neither observable or directly measurable at the present time. However, the formulation has some heuristic value in that it postulates that this lag mechanism retards the acquisition of those developmental skills which are in primary ascendancy during preschool and which are crucial to the early phases of reading. This statement, which is conceptualized in developmental-behavioral terms, is subject to more direct test and evaluation. The theory is also explicit with respect to this concept of a developmental lag. First, it states that the precursors of this disorder can best be explained in developmental terms, without recourse to terms such as damage, impairment or dysfunction. Second, that this lag in brain maturation will delay the acquisition of the different stages of thought that unfold during the developmental process. This latter position is critical to the second objective in our longitudinal project--namely, developmental changes in this disorder over time. However, not enough time has elapsed in the project to determine the nature of these developmental changes, if any, from ages 5 - 11. The theory nevertheless states that this lag in brain maturation (genetic or otherwise) will delay the acquisition of the developmental stages in thought for the dyslexic child. In this context, the disabled reader is expected to be constantly lagging behind his normal age control. In younger years, the pattern will involve those skills which are in primary ascendancy at that age (e.g., perceptual, cross-modal integration).

In later years the pattern will change as the dyslexic child eventually "catches up" on the earlier developing skills--now only to be lagging on those crucial later-developing cognitive language skills which unfold from the preceding stages of concrete perceptual operations. If the dyslexic child fails to "catch up" on these more cognitive-language skills which develop in later childhood--when the brain is reaching full maturation--then more permanent delays in language and reading skills are predicted (Satz and Van Nostrand, 1973).

Evaluation of this part of the theory will not be completed until the children reach 11 years of age in 1976 (Grade 5). The test of a developmental lag hypothesis, however, will be possible because of the repeat administration of the tests at three year intervals since kindergarten (K, G2, and G5). Preliminary evaluation between grades K and G2 (Satz, Friel and Rudegeair, 1974) provided some initial support for this hypothesis of delayed acquisition. Only time will determine whether these preliminary findings are replicable. If so, they should provide additional knowledge concerning the precursors <u>and</u> developmental course of this disorder over time. This should bring us one step closer to the major questions of etiology and prognosis. These questions can and must be answered if we are to prevent or reduce the large number of human casualties that needlessly result when academic problems begin so early in life.

BEST COPY AVAILABLE

Footnotes:

1 Research supported in part by funds from The National Institute of Mental Health, Behavioral Sciences Research Branch (MH 19415).

2 Nineteen variables derived from the kindergarten battery were utilized in these studies. A detailed description of each variable is reported elsewhere (Satz and Friel, 1973). They consist of (1) Day of Testing (2) Age, (3) Handedness, (4) Finger Tapping Total, (5) Peabody Picture Vocabulary Test, (6) Recognition-Discrimination, (7) Embedded Figures, (8) Verbal Fluency, (9) Beery Developmental Test of Visual-Motor Integration, (10) WISC Similarities Subtest, (11) Alphabet Recitation, (12) Right-Left Discrimination, (13) Finger Localization, (14) Auditory-Discrimination, (15) Dichotic Listening, Right Channel Recall, (16) Dichotic Listening, Left Channel Recall, (17) Auditory-Visual Integration, (18) Behavioral Checklist and (19) Socio-economic Status.

3 A preliminary one year follow-up to the end of kindergarten is reported in a separate study (Satz and Friel, 1973).

4 This hit-rate was based on a standard computer generated cutting line (Satz, 1966); however, the hit-rate increased to 91% when an adjusted cutting line was employed (reference Satz and Friel, 1974).

5 Day of Testing, which also loaded on Factor I, was excluded as a factorial variable because of its low correlation coefficient with this factor ($r \leq .37$).

6 Fewer Ss were included in this combined criterion than the Classroom Reading Level group since the IOTA was administered to only those Ss who still resided in Florida or Georgia at the end of Grade 2, whereas the teacher questionnaires were distributed throughout the USA and abroad.

7 A 34% overall incidence of reading disability, using this more refined composite reading criterion, is approximately what would be expected for a male population. If the national incidence is approximately 20% with a sex ratio of 4:1 in favor of boys, then the incidence should increase to 32% in a male population.

8 The abbreviated battery comprised the following: (1) Peabody Picture Vocabulary Test, (2) Recognition-Discrimination Test, (3) Beery Developmental Test of Visual-Motor Integration, (4) Alphabet Recitation Test, (5) Finger Localization Test, (6) Auditory-Discrimination Test, (7) Dichotic Listening Test (Total R and L Recall) and (8) Socio-economic status.

9 This figure (N = 459) is almost identical to the available population in the third year follow-up (N = 458).

10 Similar discrimination between groups was observed for the abbreviated battery (N = 8) in this four year follow-up.

11 The slight discrepancy in total hit-rate between DSCRIM and BMD07M is believed to reflect computational differences between the two programs.

ERIC

BEST COPY AVAILABLE

References

Ames, L. B. Learning disabilities: The developmental point of view. In H. R. Myklebust (Ed.), Progress in learning disabilities: Volume one. New York and London: Grune and Stratton, 1968. Pp. 39-74.

Bentzen, F. Sex ratios in learning and behavior disorders. American Journal of Orthopsychiatry, 1963, 23, 92-98.

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

Bruner, J. S. The course of cognitive growth. In N. S. Endler, L. R. Boulter, and H. Osser (Eds.), Contemporary issues in developmental psychology. New York: Holt, Rinehart and Winston, Inc., 1968. Pp. 476-494.

Caldwell, B. M. The usefulness of the critical period hypothesis in the study of filiative behavior. In N. S. Endler, L. R. Boulter, and H. Osser (Eds.), Contemporary issues in developmental psychology. New York: Holt, Rinehart and Winston, Inc., 1968. Pp. 213-223.

Chalfont, J. C. and Scheffelin, M. A. Central Processing Dysfunctions in Children: A review of research. NINDS Monograph No. 9, U.S. Department of Health, Education and Welfare, 1969.

Critchley, M. Dysgraphia and other anomalies of written speech. The Pediatric Clinics of North America, 1968, 15, 639-650.

de Hirsch, K., Jansky, J. and Langford, W. S. Predicting reading failure. New York: Harper and Row, 1966.

Eisenberg, L. The epidemiology of reading retardation and a program for preventive intervention. In J. Money (Ed.), The disabled reader: Education of the dyslexic child. Baltimore: Johns Hopkins Press, 1966. Pp. 3-20.

Furth, H. G. Symposium on Current Operation, Remediation and Evaluation: Dallas,

BEST COPY AVAILABLE

Texas. Meeting report. Journal of Learning Disabilities, 1972, 5, 650-652.

Gallagher, J. J. and Bradley, R. H. Early identification of developmental difficulties. In I. J. Gordon (Ed.), Early childhood education. Chicago: University of Chicago Press, 1972. Pp. 291-303.

Gibson, E. J. Learning to read. In N. S. Endler, L. R. Boulter and H. Osser (Eds.), Contemporary issues in developmental psychology. New York: Holt, Rinehart and Winston, Inc., 1968. Pp. 291-303.

Hunt, J. McV. Intelligence and experience. New York: Ronald Press, 1961.

Keeney, A. H. and Keeney, V. T. Dyslexia: diagnosis and treatment of reading disorders. St. Louis: The C. V. Mosby Co., 1968.

Kline, C. L. The adolescents with learning problems: How long must they wait? Journal of Learning Disabilities, 1972, 5, Pp. 262-271.

Lindgren, H. C. and Bryne, D. Psychology: An introduction to a behavorial Science. New York: John Wiley and Sons, Inc., 1971.

Luria, A. R. Higher cortical functions in man. New York: Basic Books, Inc., 1966.

Meehl, P. E. and Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin, 1955, 52, 194-216.

Money, J. (Ed.). Reading disability: Progress and research needs in dyslexia. Baltimore: John Hopkins Press, 1962.

Piaget, J. Judgment and reasoning in the child. New York: Harcourt and Brace, 1926.

Satz, P. A block rotation test: The application of multivariate and decision theory analysis for the prediction of organic brain disorder. Psychological Monographs, 1966, 80, (21, whole no. 629).

Satz, P. Left-handedness and early brain insult: An explanation. Neuropsychologia, 1973, 11, 115-117.

Satz, P., Fennell, E. and Reilly, C. The predictive validity of six neuro-diagnostic tests: A decision theory analysis. Journal of Consulting and Clinical Psychology, 1970, 34, 375-381.

Satz, P. and Friel, J. Some predictive antecedents of specific learning disability: A preliminary one-year follow-up. In P. Satz and J. Ross (Eds.), The disabled learner: Early detection and intervention. Rotterdam, The Netherlands: Rotterdam University Press, 1973. Pp. 79-98.

Satz, P. and Friel, J. Some predictive antecedents of specific reading disability: A preliminary two-year follow-up. Journal of Learning Disabilities, 1974, 7, 437-444.

Satz, P., Friel, J. and Goebel, R. Some predictive antecedents of specific reading disability: A three-year follow-up. Bulletin of the Orton Society, in press, 1974.

Satz, P., Friel, J. and Rudegeair, F. Differential changes in the acquisition of developmental skills in children who later become dyslexic: A three year follow-up. In D. Stein, J. Rosen and N. Butters (Eds.), Plasticity and recovery of function in the central nervous system, New York: Academic Press, 1974. Pp. 175-202.

Satz, P., Rardin, D. and Ross, J. An evaluation of a theory of specific developmental dyslexia. Child Development, 1971, 42, 2009-2021.

Satz, P. and Sparrow, S. Specific developmental dyslexia: A theoretical formulation. In D. J. Bakker and P. Satz (Eds.), Specific reading disability: Advances in theory and method. Rotterdam, The Netherlands: Rotterdam University Press, 1970. Pp. 17-39.

Satz, P. and Van Nostrand, G. K. Developmental dyslexia: An evaluation of a theory. In P. Satz and J. Ross (Eds.), The disabled learner: Early detection and intervention. Rotterdam, The Netherlands: Rotterdam University Press, 1973. Pp. 121-148.

Scott, J. P. Critical periods in behavioral development. In N. S. Endler, L. R. Boulter and H. Osser (Eds.), Contemporary issues in developmental psychology. New York: Holt, Rinehart and Winston, Inc., 1968. Pp. 213-223.

Smith, F. Understanding Reading: A psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart and Winston, 1971. Pp. 15-27.

Strag, G. A. Comparative behavioral ratings of parents with severe mentally retarded, special learning disability, and normal children. Journal of Learning Disabilities, 1972, 5, 52-56.

Thurstone, L. L. The differential growth of mental abilities. Chapel Hill, North Carolina: University of North Carolina Psychometric Laboratory, # 14, 1955.

Table 1

Predictive Classification of Ss into Criterion Reading Groups (1972)[a] based on Discriminant Function Composite Scores (1970)

Composite Test Scores		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	18	39	55	3
	%	(100)	(71)	(16)	(5)
−	N	0	16	284	58
	%	(0)	(29)	(84)	(95)
	T	18	55	339	61

[a] Criterion = Classroom Reading Level (End of Grade 1) (N_T = 473).

Table 2

Discriminative Ranking and Cumulative Classification of Tests by Factor Loadings based on Discriminant Function Composite Score[a]

Ranked Variables	Factor	Cumulative % Correct
1. Finger Localization	I	76.1
2. Recognition-Discrimination	I	77.2
3. Day of Testing		79.1
4. Alphabet Recitation	I	81.6
5. Residual Tests	I-IV	84.4

[a] Criterion = Classroom Reading Level (end of Grade 1).

Table 3

Predictive Classification of Children into Criterion Reading Groups (1973)[a] based on Discriminant Function Composite Scores (1970)

Composite Test Scores[b]		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	48	47	74	4
	%	(89)	(71)	(27)	(6)
–	N	6	19	196	64
	%	(11)	(29)	(73)	(94)
	T	54	66	270	68

[a] Criterion = Classroom Reading Level (N_T = 458).

[b] Program = DSCRIM, Multiple Discriminant Analysis (4 group predictions collapsed), prior probabilities set equal.

ERIC

Table 4

Discriminative Ranking and Cumulative Classification of Tests by Factor Loading based on Stepwise Discriminant Function Composite Scores[a,b]

Ranked Variables	Factor	Cumulative % Correct
1. Finger Localization	I	70.5
2. Alphabet Recitation	I	75.5
3. Recognition-Discrimination	I	77.3
4. Day of Testing		76.9
5. Residual Tests	I-IV	77.7

[a] Criterion = Classroom Reading Level

[b] Program = BMD07M, Stepwise Discriminant Analysis (4 group), prior probabilities set equal.

ERIC

Table 5

Predictive Classification of Children into Criterion Reading Groups (1973)[a] based on Discriminant Function Composite Scores (1970)

Composite Test Scores[b]		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	61	51	68	2
	%	(91)	(66)	(32)	(3)
–	N	6	26	146	59
	%	(9)	(34)	(68)	(97)
	T	67	77	214	61

[a] Criterion = Classroom Reading Level & IOTA Word Recognition (N = 419).

[b] Program = DSCRIM, Multiple Discriminant Analysis (4 group predictions collapsed), prior probabilities set equal.

Table 6

Probability of Decision Risk (Treatment/No Treatment) associated with Differential Composite Test Score Predictions to Third Year Reading Level[a]

Composite Test Scores[b]	Decision	Criterion Groups: Severe	Mild	Average	Superior	Ratio Correct	p
+ +	T	50	18	15	0	68/83	.82
+	T	11	33	53	2	44/99	.44
-	NT	5	23	84	11	95/123	.77
- -	NT	1	3	62	48	110/114	.96
	Total	67	77	214	61		

[a] Criterion = Classroom Reading Level & IOTA Word Recognition (N_T = 419).

[b] Program = DSCRIM, Multiple Discriminant Analysis (4 group), prior probabilities set equal.

Table 7

Predictive Classification of Children into Criterion Reading Groups (1974)[a] based on Discriminant Function Composite Scores (1970)

Composite Test Scores[b]		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	50	64	71	5
	%	(91)	(69)	(30)	(06)
-	N	5	29	163	72
	%	(09)	(31)	(70)	(94)
	T	55	93	234	77

[a] Criterion = Classroom Reading Level (N = 459).

[b] Program = DSCRIM, Multiple Discriminant Analysis (4 group predictions collapsed), prior probabilities set equal.

ERIC

Table 8

Discriminative Ranking and Cumulative Classification of Tests by Factor Loadings based on Stepwise Discriminant Function Composite Score[a,b]

Ranked Variables	Factor	Cumulative % Correct
1. Finger Localization	I	70.4
2. Alphabet Recitation	I	74.1
3. Recognition-Discrimination	I	76.5
4. Dichotic Listening (Total)	I	77.0
5. Residual Tests	I-IV	78.2

[a] Criterion = Classroom Reading Level

[b] Program = BMD07M, Stepwise Discriminant Analysis (4 group), prior probabilities set equal.

ERIC

Table 9

Predictive Classification of Cross-Validation Sample into Criterion Reading Groups (Grade 2, 1974)[a] based on Discriminant Function Weights of Standardization Population (1970-1973)

Composite Test Scores[b]		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	16	9	28	3
	%	(89)	(36)	(30)	(7)
-	N	2	16	64	37
	%	(11)	(64)	(70)	(93)
	T	18	25	92	40

[a] Criterion = Classroom Reading Level (N = 175)

[b] Abbreviated Test Battery

Table 10

Discriminative Ranking and Cumulative Classification of Tests by Factor Loadings based on Discriminant Function Composite Score [a,b]

Ranked Variables	Factor	Cumulative % Correct
1. Finger Localization	I	66.9
2. Embedded Figures	I	68.2
3. Similarities	III	70.9
4. Finger Tapping (Total)	IV	71.5
5. Residual Tests	I-IV	77.5

[a] Criterion = Classroom Reading Level (N_T = 175).

[b] Program = BMD07M, Stepwise Discriminant Analysis (4 group), prior probabilities set equal.

ERIC

Table 11

Predictive Classification of Cross-Validation Sample into Criterion Reading Groups (Grade 2, 1974)[a] based on Discriminant Function Weights of Standardization Population (1970-1973)

Composite Test Scores[b]		Criterion Groups			
		Severe	Mild	Average	Superior
+	N	19	19	16	1
	%	(76)	(53)	(27)	(3)
-	N	6	17	44	29
	%	(24)	(47)	(73)	(97)
	T	25	36	60	30

a Criterion = Classroom Reading Level and IOTA Word Recognition (N = 151).

b Abbreviated Test Battery

BEST COPY AVAILABLE

Table 12

Means and Standard Diviations of Tests (Abbreviated) Battery, Grade K) for Reading Groups (Combined Criterion, Grade 2)

	Criterion Groups			
	Severe N=67 Age[a]=65.9	Mild N=77 Age=65.6	Average N=214 Age=66.4	Superior N=61 Age=67.0
1. Finger Localization	26.2 (7.1)	31.1 (8.0)	35.9 (7.6)	39.5 (5.2)
2. Alphabet Recitation	14.0 (8.8)	17.5 (8.5)	22.2 (6.8)	25.3 (6.9)
3. Recognition-Discrimination	6.9 (2.9)	8.9 (3.0)	9.9 (2.6)	11.6 (2.0)
4. Peabody	92.0 (16.5)	100.2 (12.3)	106.4 (14.9)	113.1 (12.1)
5. Beery[b]	56.3 (7.5)	60.7 (8.1)	66.5 (9.3)	71.4 (10.8)
6. Auditory-Discrimination	1.1 (.5)	1.1 (.5)	1.3 (.4)	1.5 (.3)
7. Dichotic Listening (Total)	57.6 (13.7)	63.0 (14.5)	68.1 (14.2)	76.4 (13.9)
8. Socio-economic Status	1.7 (.4)	1.8 (.4)	1.9 (.3)	2.0 (0.0)

[a] Age in Months

[b] Score in Months

ERIC

BEST COPY AVAILABLE

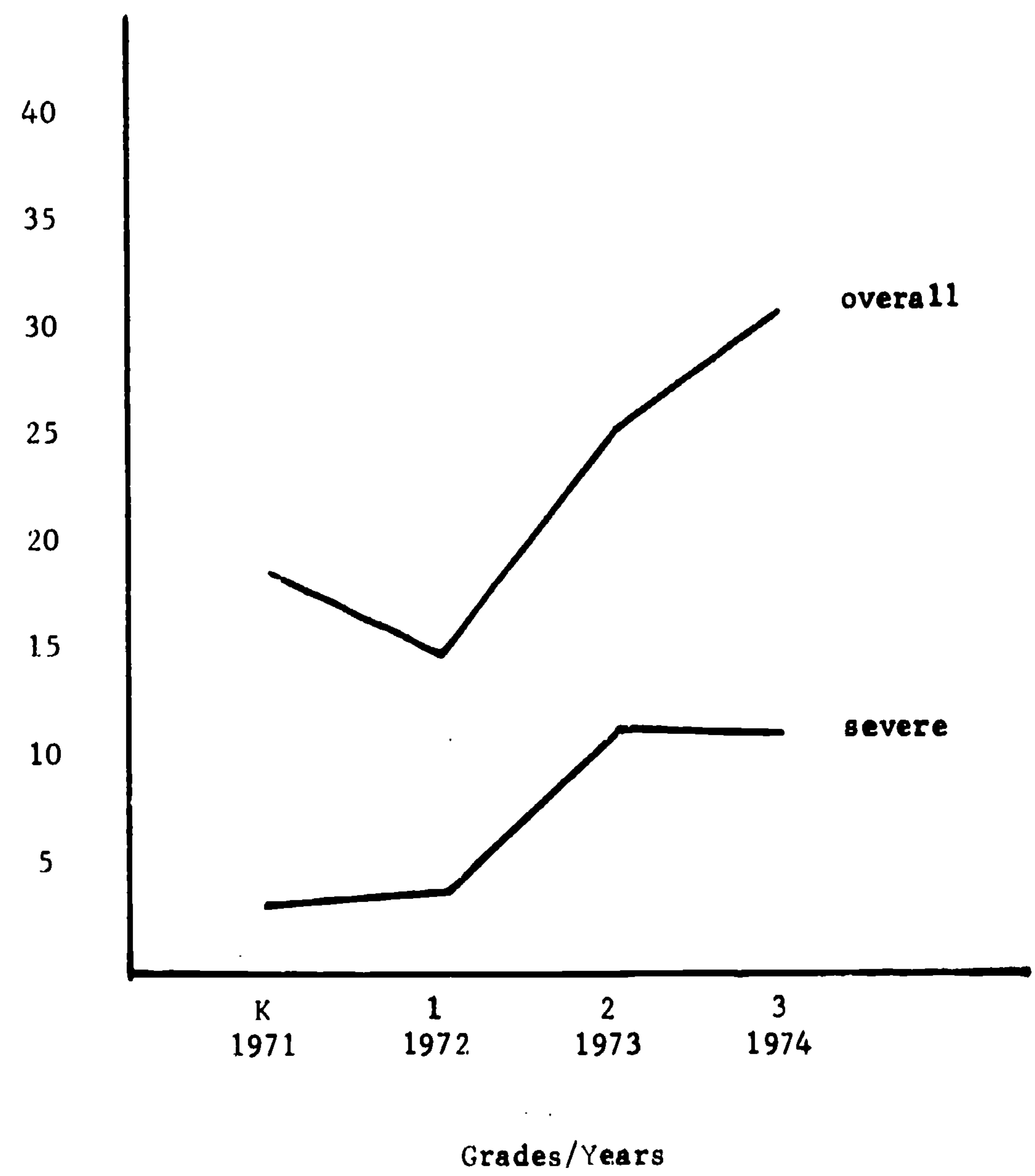

Figure 1. Follow-up changes in Incidence of Overall Reading Disability (Severe and Mild) and Severe Reading Disability in population from end of Grades K - 3 (1971 - 1974).