DOCUMENT RESUME

ED 342 201 EC 300 960

AUTHOR	Cooper, David H.; Speece, Deborah L.
TITLE	Project SEARCH: A Longitudinal Study of Primary Grade Students At-Risk for School Failure. Final Report.
INSTITUTION	Maryland Univ., College Park. Dept. of Special Education.
SPONS AGENCY	Office of Special Education and Rehabilitative Services (ED), Washington, DC.
PUB DATE	Apr 89
CONTRACT	G008530155
NOTE	189p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Academic Failure; *At Risk Persons; Delivery Systems; Educational Methods; Evaluation Methods; Grade 1; Handicap Identification; High Risk Students; *Instructional Effectiveness; Intervention; Language Handicaps; Learning Disabilities; Longitudinal Studies; Mild Mental Retardation; Prevention; Primary Education; Profiles; *Referral; Risk; *Student Characteristics; Teaching Methods
IDENTIFIERS	Prereferral Intervention

ABSTRACT

This report presents four studies summarizing results of Project SEARCH, which attempted to identify characteristics of children at risk, apply risk analysis procedures, and identify effective educational practices. The first study examined the presence or absence of risk factors in 25 first graders (and 10 controls) referred by their teachers to teacher assistance teams. Resulting profiles were examined to determine degree of risk for referral. The second study examined instructional arrangements in first grade and assessed their effects on academic responses of students (N=67) at risk for failure and controls (N=58). Several specific arrangements were shown to be associated with either accelerated or decelerated rates of at-risk and control students' active, academic, or inappropriate responses. The third study examined the role of dynamic assessment measures in models of individual differences with 193 first grade children. Limited support was found for use of dynamic assessment task prompts in the prediction of academic achievement. The fourth study addressed the ontogeny of school failure in 112 first grade children, 63 of whom were considered at risk. Six clusters were empirically derived with three profiles representing atypical patterns suggestive of learning disabilities, mild mental retardation, and language problems. Appendixes include: a curriculum and methods questionnaire, a 1988 Project Search Update, and a listing of dissemination papers and presentations. (136 references) (DB)

D14

Project SEARCH: A Longitudinal Study of Primary Grade Students At-Risk for School Failure

023 QH 60001

David H. Cooper and Deborah L. Speece
Department of Special Education
University of Maryland
College Park, MD 20742

Final Report submitted to the U.S. Department of Education, Office of Special Education and Rehabilitative Services (Enhancing Instructional Program Options 84-023P, G008530155), April, 1989.

Table of Contents

This report summarizes the results to date of Project SEARCH: A Longitudinal Study of Primary Grade Students At-Risk for School Failure (Enhancing Instructional Program Options 84-023P, G008530155). This report consists of (a) research objectives from the original grant application, (b) the corresponding procedural objectives, (c) accomplishments keyed to each procedural objective and (d) statement of remaining tasks.

Research Objectives

1. Identification of cognitive, behavioral, and demographic characteristics of children at risk.
2. Identification of clusters of at-risk children based on an empirical (cluster) analysis of child characteristics.
3. Risk analysis of cognitive and behavioral characteristics of children in the teacher nominated at-risk group.
4. Identification of clusters of current educational practices used by teachers to promote learning by at-risk children.
5. Identification of educational practices (e.g., curriculum, instructional strategies, materials) associated with reduced risk for special education referral.

Procedural Objectives

The following represented the major research activities for this project:

1. Develop a teacher questionnaire to assess classroom curriculum, instructional methods, and modifications designed for children identified as at risk for academic and behavioral problems.
2. Develop a classroom observation system to capture learning environment variables including teacher-child interaction, time

on task, classroom structure, and task demands.

3. Identify at-risk children (through teacher referrals to the Teacher Assistance Teams - TAT) and classroom control children.
4. Administer individual measures of achievement, intelligence, and information processing to all children.
5. Complete the curriculum and instructional methods questionnaire and behavioral ratings.
6. Observe and characterize the learning environment of all at-risk children.
7. Identify clusters of children based on child characteristics (behavior, achievement, intelligence, information processing and demographics) and identify clusters of children based on characteristics of their learning environments.
8. Perform risk analyses on child characteristic clusters to determine which characteristics are associated with membership in the at-risk group and perform risk analyses on the learning environment clusters to determine characteristics of instruction that serve to maintain at-risk children in the regular classroom.
9. Evaluate progress for each research activity and disseminate project results.

<u>Accomplishments</u>

A summary of the accomplishments with respect to each procedural objective is listed below. Specific results are contained in the four studies that follow this section.

1. <u>Develop teacher questionnaire</u>: The project staff developed the

Curriculum and Methods Questionnaire (CMQ) which can be found in Appendix A. The CMQ has undergone several revisions based on feedback from our participating teachers.

2. Develop classroom observation instrument: Subsequent to the submission of the original grant application, we identified an existing observation system that met the requirements of Project SEARCH. The code for Instructional Structure and Student Academic Response (CISSAR) (Stanley & Greenwood, 1981) was modified, with permission from the authors, for this research. A detailed description of the instrument and specific results pertaining to first grade environments are contained in Study Two.

3. Identify at-risk and control children: Three successive cohorts of first grade at-risk (AR) and control (C) children were identified and recruited during the three project years. Complete data sets were obtained for 104 AR and 89 C children during their first year of schooling. Our identification procedures and the demographic data for the two participating school districts are contained in Study One.

4. Administer individual measures of achievement, intelligence, and information processing to all children: The following measures were used: Diagnostic Achievement Battery (DAB, reading, speaking, listening, mathematics subtests), Cognitive Abilities Test (CogAt, verbal and nonverbal intelligence), Preschool Language Assessment Instrument (PLAI), and Dynamic Assessment Tasks (DAT). Descriptions of each instrument can be found in Study One with specific results in Study Three and Study Four.

5. Complete CMQ and behavioral ratings: As noted above the CMQ

was developed and completed by participating teachers for at-risk children. Teachers also completed the Cooper-Farran Behavioral Rating Scales (CFBRS) on all subjects. The CFBRS is described in Study One. Study Four contains descriptive data for the CFBRS.

6. Observe and characterize the learning environments of all at-risk children: CISSAR data were collected on 204 subjects in first grade. Results are described in Study Two.

7. Identify empirical clusters of children: Study Four contains specific results pertaining to subtypes of AR and C children based on child characteristic measures.

8. Perform risk analyses: Study One describes the methodology of risk analysis and Study Four contains specific results.

9. Evaluate progress and disseminate results: Appendix B contains an example of a Project SEARCH Update that was sent to all participating teachers, principals, and administrators at regular intervals during the project. Appendix C contains a list of publications and paper presentations based on this project. In addition to the efforts described in Appendices B and C we have made several invited presentations concerning this research to the faculty of the College of Education and the Doctoral Seminar, Department of Special Education.

Studies One through Four follow.

Study One: Children at Risk

Abstract

This paper introduces the conceptual framework guiding a three-year investigation of first grade children at risk for referral and placement in special education. Preliminary data for the first of three planned cohorts are presented that illustrated a novel methodological approach to identifying the characteristics of children at risk for referral. The sample (n=35) consisted of 25 first graders who were referred by their teachers to teacher assistance teams and 10 control children, matched for classroom membership and sex. Children's scores on a battery of standardized and experimental tasks were used to form subgroups based on the presence or absence of risk factors. The resulting profiles were examined via a relative risk analysis to determine degree of risk for referral associated with each subgroup. The potential of the proposed methodology to understanding the characteristics and learning environments of at-risk children is discussed.

A Novel Methodology for the Study of Children At Risk for School Failure

The purposes of this article are to present the conceptual design of a three year investigation of first grade children at-risk for school failure and to illustrate the potential of a novel methodology for understanding the contribution of child characteristics and learning environments to the development of a child's status as "at-risk". Two research questions guided our study: (a) what are the characteristics of these children that place them at risk and (b) which instructional practices reduce the risk of failure for some children and thus allow the children to receive their education in regular classroom settings? Both identification of and intervention for these at-risk children present a significant challenge to regular and special education. However, as Reschly (1984) pointed out, there is surprising lack of empirical evidence concerning who gets referred to special education and why; what instructional options, if any, are attempted before referral; and what differences, if any, exist in the functional education needs among different categories of mildly handicapping conditions such as the educable mentally retarded (EMR) and learning disabled (LD). An analysis of research trends in special and regular education provides some insight as to why information is limited on these critical issues.

For the most part, research with mildly handicapped children, particularly the learning disabled, has focused on identifying child characteristics after the children have been classified as handicapped. Thus, a wealth of studies exist that document, for example, information processing deficiencies (Swanson, 1983; Torgesen & Houck, 1980); maladaptive classroom behavior (Feagans & McKinney, 1981; McKinney & Speece, 1983) and problematic communication skills (Bryan & Bryan, 1978; Feagans & Short, 1984). Although these studies have served to broaden our perspective on mildly handicapping conditions, this emphasis on child characteristics assumes that the learning problems are inherent and not a function of classroom processes (Coles, 1978). Also, study of students already identified as handicapped precludes a comparison with children referred but not placed in terms of differences in child characteristics and classroom practices.

While there is an abundance of research documenting group differences, there is a lack of evidence on systematic intervention efforts with at-risk and mildly handicapped children. Pertinent available research tends to be based on single subject designs (e.g., Lloyd, Saltzman, & Kaufman, 1981), thereby limiting generalizability; based on global variables such as school attendance and time in remedial programs, which are too general to determine any significant educational benefits (Helper, 1980); or based on laboratory efforts (e.g., Torgesen, 1977), which limit implications for the natural classroom environment. Complicating these issues is the problem of heterogeneity of children who are at risk or classified as handicapped (Satz & Fletcher, 1980). Given this variability, it is unlikely that a single set of educational practices will work equally well with all children experiencing learning and behavior problems (Finn & Resnick, 1984; Speece, McKinney, & Appelbaum, 1985). The question becomes which practices work best with which children?

In contrast to the traditional concerns of special education research, studies of regular education have focused on classroom practices such as allocation of learning time, classroom management, instructional practices, implementation of curriculum, and characteristics of effective teachers (e.g., Good & Brophy, 1978; Stallings, 1980). While this work provides important instructional and methodological insights for studying children at risk in the classroom, Good (1983) has pointed out that the results are typically in the form of classroom mean scores as opposed to effects for groups of similar children. As in special education research, the linkages among child characteristics, effective instructional practices and regular/special status remain to be identified.

Method

Design

A problem common to both special and regular education research traditions concerns heterogeneity with respect to child characteristics and interventions. At present it is difficult to determine which instructional practices will most likely benefit an individual child. To address this issue, we are implementing two design strategies that are novel to educational research. First, cluster analysis techniques will be used to identify homogeneous subtypes of children in regard to both child characteristics and learning environment variables. Cluster analysis is an empirical, multivariate classification technique that divides a heterogeneous data set into clusters of children who are similar to each other across an array of variables. This technique has been used recently to identify clusters of learning disabled children across a variety of domains (e.g., Satz & Morris, 1981; Speece, 1987). The use of this method in the present study is different from past efforts with respect to both the sample studied (at risk as opposed to handicapped) and the variables selected (learning environment variables in addition to child characteristics).

The second design strategy involves an analysis of the probability of adverse outcome, or "relative risk" associated with child characteristic clusters as well as the learning environment clusters. In the present investigation, we are studying the risk for referral to Teacher Assistance Teams and the risk for special education placement. This approach, used in epidemiological research (Kleinbaum, Kupper, & Morgenstern, 1982), provides a method for determining (a) which clusters of children, based on child characteristics, are associated with a higher degree of risk of being identified by teachers as candidates for learning and behavior problems, and (b) which clusters of interventions are associated with reducing the risk of being classified as handicapped. The usefulness of the relative risk approach to analysis of educational data has been demonstrated in a recent study in which kindergartners exhibiting aberrant work-related behaviors were compared to normally behaving peers (Cooper & Farran, 1988). By means of a relative risk analysis, it was found that the aberrant work-related behaviors (e.g., disorganization and distractibility) placed those children at 12-times greater risk for referral to special education than their normally behaving classmates. In addition to univariate risk analysis, assessment of the risk associated with interacting factors is essential to the present study. In this regard, Cooper (1984) demonstrated that the risk for special education referral associated with aberrant work-related behavior was dramatically increased when assessed in combination with low achievement in reading.

The combination of cluster analysis and relative risk methodologies provides more specific practical outcomes than those afforded by other strategies. By addressing the reality of sample heterogeneity in addition to a straightforward analysis of the risk associated with particular patterns of child characteristics and learning environments, we plan to identify educationally useful findings regarding the match between learners and environments that serve to maintain children in regular education settings.

<u>Measures</u>

To study these issues, we adopted elements of the assessment system proposed by the National Academy of Sciences Panel to address overrepresentation of minority children and males in special education programs (Heller, Holtzman, & Messick, 1982). Heller et al. (1982) suggested that assessment include an evaluation of the child's learning environment as well as the more traditional measurement of child characteristics. In the present study, assessment of the learning environment was approached in two ways. First, teachers completed the Curriculum and Methods Questionnaire (CMQ), an instrument developed by the investigators to operationalize factors in the learning environment identified by Heller et al. (1982) and effective teaching behaviors (e.g., Brophy, 1979). The instrument contains 18 questions regarding reading, mathematics, and oral language curricula, 34 items on teaching methods (e.g., physical arrangements, evaluation of instructional units, social environment for learning), 16 questions on behavior management methods, and 12 items on teaching strategies (i.e., strategies used for acquisition, fluency, and generalization across academic areas). At the end of each section, teachers are asked to indicate how many other students receive the same type of instruction. The validity of this measure rests largely on its content, which was evaluated by practitioners and scholars of curriculum and instruction, with revisions made as necessary.

The second measure of the learning environment was based on a modified version of CISSAR, a classroom observation instrument developed by Stanley and Greenwood (1981). CISSAR is a time sampling method that taps classroom structure and activity as well as teacher, and child behavior. It has adequate reliability and validity (see Greenwood, Schulte, Kohler, Dinwiddie, & Carta, 1986). Interobserver agreement was checked weekly in the present study with a minimum level of 80% required.

Measures of child characteristics included the Diagnostic Achievement Battery (DAB; Newcomer & Curtis, 1984) and the Cognitive Abilities Test (CogAt; Thorndike & Hagen, 1982). The DAB provides measures of achievement in reading, arithmetic, listening, and spoken language. Verbal and nonverbal intelligence quotients are obtained from the CogAt. Both measures have received favorable reviews regarding psychometric adequacy (Brown & Bryant, 1984; Salvia & Ysseldyke, 1985). A third measure of child characteristics was provided by teacher ratings of classroom behaviors on the Cooper-Farran Behavioral Rating Scale (CFBRS, Cooper & Farran, 1984). The CFBRS is a 39-item instrument that provides scores on two factors, interpersonal skills (IPS) and work-related skills (WRS). IPS taps physical and verbal aggressiveness and disruptiveness while the WRS factor measures disorganization, distractibility, and noncompliance. Data supporting adequate reliability of the two factors are reported in Cooper and Farran (1988) and Cooper (1984). Intra-class correlations above .78 were obtained for both factors when analyzed for inter-rater reliability. Content validity was established during scale development and by estimating internal consistency (Cronbach's alpha = .96). Construct validity has been examined by means of factor analytic studies of five data sets, totalling over 1400 subjects.

Two other experimental measures were included in the child characteristics battery. The Preschool Language Assessment Instrument (PLAI; Blank, Rose, & Berlin, 1978) requires the child to respond to four groups of questions that vary in their degree of abstractness. Blank et al. (1978) reported that, across the four groups of questions, split-half reliability ranged between .64 - .86 with test-retest reliability ranging from .73 to .88. Evidence to support content, concurrent and construct validity was also reported (see Blank et al., 1978).

The last measure was adapted from the work of Brown and her colleagues (Bryant, Brown, & Campione, 1983; Ferrara, Brown, & Campione, 1986) on dynamic assessment. The Dynamic Assessment Tasks (DAT) represent a guided teaching experience wherein the child is taught to solve difficult matrix problems that follow a single rule. A sequence of standardized graduated prompts is given until the matrix is solved. For our purposes, the measure of interest was the number of prompts required before the child arrived at the solution. Following Peterson, Homer, and Wonderlich (1982), procedural reliability data were collected to determine the degree to which examiners were adhering to the scripts written for the DAT. Eighteen sessions were recorded on audiotape and seven were randomly selected and rescored by an independent observer. Based on the percentage agreement method, the procedural reliability for all protocols was 100%.

<u>Setting and Participants</u>

<u>School Districts</u>. The sample for this study was drawn from two county school districts adjacent to the Washington, D.C. metropolitan area. County A had a general population of 660,000 with a racial composition of

46% Black, 49% Caucasian, and 5% other ethnicities. The total enrollment for the school system was 103,325 with 62.6% Black, 31% Caucasian, and 6.4% other. The population for County B was 414,074 with a racial distribution of 12.6% Black, 85.4% Caucasian, and 2% other ethnicities. The school system population was 64,552 and reflected the racial composition of the general population with 14.2% Black, 83.8% Caucasian, and 2% other ethnicities.

The at-risk population for this study was all first grade children who were referred to Teacher Assistance Teams (TATs), which operate in both school districts as a "pre-referral" system before special education referral at the building level. The TATs were composed of regular education and sometimes special education teachers whose duties included reviewing the referral with the teacher, providing alternatives for instruction and management, and monitoring progress in the regular classroom. When a child was formally referred to a TAT, a control child of the same sex and from the same classroom was randomly selected from the teacher's list of children who were considered to be progressing normally and who, in the teacher's opinion, would not be referred to either the TAT or gifted and talented programs. During our 3-year investigation, we plan to study 200 at-risk and 200 control children.

Students. Preliminary findings reported below were based on the first of three planned cohorts. In addition to the criterion of referral to the TAT for an academic or behavioral problem, several additional criteria were applied for the selection of at-risk children. Children could not be receiving any special education services, could not be repeating first grade, and, if they were receiving any supplemental instruction outside of the regular classroom, the total amount of time could not exceed 2½ hours per week. The latter criterion was to insure that the learning environment to be studied was primarily the regular education classroom. Thirty-five students on whom we had complete data for the child characteristics were selected for this study; 25 were at-risk (AR) and 10 were controls (C). The racial composition was 37% Black ($\underline{n}$ = 11 AR, 2 C) and 63% Caucasian ($\underline{n}$ = 14 AR, 8 C) with a gender ratio of 34% females ($\underline{n}$ = 9 AR, 3 C) and 66% males ($\underline{n}$ = 16 AR, 7 C).

To illustrate our approach, several child characteristics were selected and analyzed for their importance in defining the at-risk child, the first research question in this investigation. For comparison purposes, results from both regression analysis and relative risk analyses will be presented to highlight potential differences between these methods. Due to small sample size and use of a restricted data set, the results are to be regarded as preliminary rather than as stable indicators of risk factors.

Illustrative Analysis

Definition of Risk Factors

The risk approach identifies factors that are associated with a specified outcome. In the present study, we were interested in determining which factors were associated with a child's referral to a TAT. Relative risk analysis requires use of cut-off scores to divide a sample such that most subjects are said to be free of the risk factor, (e.g., for children scoring below -1 SD, the risk factor is present.) The presence or absence of the risk factor can then be tested for its association with the presence or absence of the outcome of interest.

Our risk factors were derived from selected scores on the battery of educational measures described above. For each measure a cut-off score was determined, and children scoring below the cut-off were differentiated from the rest. Although this procedure results in a loss of variance for the differentiated groups, it is defensible on practical and methodological grounds. From a practical standpoint, it is common educational practice to refer to children's strengths and weaknesses, to describe performance as normal or below average, or to make other similar categorical assessments of children's abilities. These categorizations serve to classify children's characteristics in ways that are potentially useful and familiar to practitioners.

Methodologically, this categorization of subjects is necessary to estimate prevalences of adverse outcome for two distinct groups: subjects who share a risk factor or risk profile, such as poor work-related skills and poor comprehension of discourse, and subjects who do not exhibit that risk factor or profile. The ratio of the prevalence estimates for the two groups is defined as the relative risk, or strength of association between the risk factor, or profiles, and the outcome. The extent to which this ratio deviates from unity is interpreted as the increased risk associated with the risk factor or profile relative to a control group.

For example, Figure 1 represents a hypothetical situation in which a risk factor, poor work-related skills, is evaluated for its association with referral to special education. Assuming a representative sample, the relative risk is equal to the product of cells a and d, divided by the product of cells b and c, or (9 x 80) ÷ (38 x 6) = 3.16. Note that this equation is mathematically equivalent to the ratio of two prevalence estimates (a/a+b) ÷ (c/c+d). Cells c and d represent the baseline condition to which all other prevalences of interest are compared. In this example, the baseline prevalence is 7.5%. A relative risk equal to 3.16 is interpreted as follows: The presence of poor work-related skills results in referral approximately three times as often as when this risk factor is absent.

[INSERT FIGURE 1 ABOUT HERE]

<u>Dichotomization Procedures</u>. The measures used were of two types, standardized and nonstandardized. For the former (including CogAt and DAB) the tabled norms were used to split the sample at a point one standard deviation below the mean. For the latter, (including the CFBRS and DAT) the cut-off scores were based on the sample of control children. Thus, we computed the mean and standard deviation for our sample and used these values to divide the subjects. The last measure, the PLAI, while not a standardized test, did permit derivation of a mastery score, which the technical manual describes as indicative of mastery of discourse at

various levels of cognitive complexity (see Blank et al., 1978). Subjects scoring below this level of mastery were categorized as having this risk factor. Table 1 lists the measures and their respective cut-off scores.

[INSERT TABLE 1 ABOUT HERE]

Risk Analyses. Three types of risk analyses were conducted: univariate, incremental, and risk clusters. Univariate risk provided the relative risk ratio for each variable considered individually while the incremental analysis indicated the risk associated with a specific number of factors, regardless of variable identity. Relative risk of cluster profiles considered the risk associated with specific patterns of factors relative to a cluster of children in which no risk factors were present.

Multiple Regression. By way of comparison, the data were also analyzed by means of stepwise multiple regression, with group (referred, not-referred) regressed on the children's test scores and behavioral ratings. Stepwise regression with backward elimination was used to reduce the full model to the most parsimonious model. Inclusion of the regression analysis serves to illustrate the possible advantages of the relative risk approach over the more familiar analytic strategy. Although an appropriate subject-to-variable ratio was preserved, this analysis, as with the risk analyses, serves only as a heuristic device given the small sample size.

Results

Univariate risk. Table 2 gives the relative risk for referral associated with each of six child characteristics. The statistical significance of each risk estimate is based on examination of the lower bound of the 95% confidence interval. When this bound is greater than 1, the risk estimate is significant. By this criterion, when CFBRS work-related skills or DAB reading achievement scores were below cut-off, the child was at significantly increased risk for referral compared with children who did not demonstrate the risk factor. Table 2 also indicates that the confidence intervals were quite large, owing to the relatively small sample sizes involved in these preliminary analyses. Thus, conclusions regarding the precise magnitude of the risks are unwarranted, although tests of significance are valid.

[INSERT TABLE 2 ABOUT HERE]

Incremental risk. Analysis of the aggregated effect of multiple risk factors may be accomplished initially by examining incremental risk. For the incremental risk analyses, the number of risk factors exhibited by a child was summed, and this sum was used to group children with equivalent numbers of risk factors. Children with zero or one risk factor (out of a possible six) served as the reference or comparison group. Children who exhibited 2 or 3 risk factors had a risk ratio of 17.5; children with 4 or 5 risk factors had a risk ratio of 33.0. These data demonstrated that as the number of risk factors increased, the relative risk increased, without regard to exactly which factors were involved.

Risk Clusters. Due to the small n, a simplified non-empirical clustering approach was adopted. Three child characteristic variables, CogAt verbal intelligence (VIQ), DAB reading achievement (RDG), and CFBRS ratings of work-related skills (WRS) were used to assign children to one of eight possible clusters derived from all possible combinations of risk (1) and not at risk (0) ratings on the three variables (see Figure 2). Cluster 1 was composed of 10 (2 AR and 8 C) children having normal VIQ and normal RDG and positive WRS ratings. These 10 children in Cluster 1 formed the reference or baseline cluster against which all other non-normal clusters were compared. The purpose of this analysis was to estimate the relative risk for referral that was associated with children's membership in each of the remaining seven non-normal clusters (Clusters 2 through 8), where non-normal was defined as having at least one score below the cut-off.

[INSERT FIGURE 2 ABOUT HERE]

Figure 2 illustrates the relative risk for referral associated with children's membership in the clusters. The lower bounds of the 95% confidence intervals (not shown) for Clusters 2, 7 and 8 were greater than 1.0, indicating statistical significance for these risk clusters relative to Cluster 1. Due to the small n's in each cluster, the confidence intervals were large. Therefore, while it is possible to conclude that the relative risks for Clusters 2, 7 and 8 were significant, it is not possible to assess the significance of the differences between and among Clusters 2, 7 and 8 themselves. The complete data set, with a much larger n and correspondingly smaller confidence intervals, will permit these more specific comparisons. This pattern of results suggested that poor work-related skills were associated with referral (Clusters 2, 7 and 8), except when combined with low verbal IQ (Cluster 6). Furthermore, poor reading achievement was not associated with referral (Cluster 3); only when it was accompanied by poor WRS (Cluster 7) was there risk for referral.

Multiple regression. Stepwise regression with backward elimination was used to test a model incorporating the same set of risk factors as the risk analyses on clusters presented above. The full model tested the prediction of referral from the combination of CogAt verbal IQ (VIQ), DAB reading achievement (RDG) and CFBRS work-related skills (WRS). R-square for the full model was .69 (F=23.74, df=3, $p < .0001$) and was reduced only to .68 by the removal of VIQ and RDG (F=71.17, df=1, $p < .0001$). Thus, WRS was the most parsimonious predictor of referral. As with the risk analyses, this result is considered preliminary due to the instability of R-square associated with small sample size.

Discussion

The three-year goal of this investigation is to discover patterns of children's characteristics that may predispose teachers to refer them for consideration by TATs. Additionally, types of learning environments will be identified and used to determine which regular classroom instructional arrangements serve to reduce the risk of special education placement for clusters of children with varying patterns of strengths and weaknesses.

Analysis strategies for these goals were illustrated with a partial data set comprised of the child characteristic measures. For heuristic purposes, "clinical" subgroups of children were formed based on the presence or absence of risk factors. Cluster analysis techniques will be used with an eventual complete data set. Application of relative risk analyses provided preliminary indication that work-related skills, as defined by the Cooper-Farran Behavior Rating Scales, and reading achievement may be important variables differentiating children who were and were not referred to teacher assistance teams by their first grade teachers.

The nature of these results provides a point of comparison with a more familiar analytic strategy, multiple regression. While results must be interpreted with caution due to sample size, the conceptual clarity and practical utility of the cluster analysis/risk analysis approach was illustrated. Both regression and risk analyses converged on the potential importance of work-related skills in defining children likely to be referred to TAT. However, the risk analysis further illustrated the importance of poor reading achievement in conjunction with poor WRS in elevating the risk for referral. In the face of limited intervention resources, this type of result suggests that, while children with poor work-related skills are more likely to be referred (similar to the implications from multiple regression), children who have both risk factors might receive priority. Beta coefficients and significance levels derived from regression are not so easily interpreted.

A potential limitation imposed by the design of this study is the definition of at-risk children. The validity of a sample defined via referral to TAT is difficult to assess. Referrals are the result of teachers' judgment, influenced by their expectations, and other variables that lie at the fringes of objective measurement. These factors are expected to produce variability in the characteristics of children who are referred. In future work, cluster analysis of the child characteristic data will capture this variability. Data regarding referring teachers' professional and educational histories also are collected and will be used to determine if the clusters differ on these variables. In addition, we have restricted our sample to those children referred to school-appointed teams. These teams do not accept casual referrals; rather, all referrals follow a specified format that includes completion of student information forms as well as attendance at a meeting by the referring teacher. Therefore it seems reasonable to assume that teachers are making referrals with an appropriate degree of thought and effort.

Referral and placement in special education, while providing educational benefits to some children, may also be a disservice to others due to the potential for misclassification. On the other hand, at-risk children who demonstrate academic and behavioral problems in the regular setting, may be referred for evaluation but not qualify for special education services. This outcome may also be a disservice if educational modifications are not designed and implemented in the regular classroom. The current project was designed to address both concerns by defining the at-risk learner and learning environments that serve to maintain some children in the general education milieu. The methodology used may provide an initial response to Reschly's (1984) concerns regarding the lack of data on these issues and to teachers who are charged with educating this group of children.

References

Blank, M., Rose, S. A., & Berlin, L. J. (1978). Preschool language assessment instrument: The language of learning in practice. Orlando, FL: Grune & Stratton, Inc.

Brophy, J. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.

Brown, L. & Bryant, B. (1984). Critical reviews of three individually administered achievement tests. RASE, 5(5), 53-60.

Bryan, T. H. & Bryan, J.H. (1978). Social interactions of learning disabled children. Learning Disability Quarterly, 1, 33-38.

Bryant, N. R., Brown, A. L., & Campione, J. C. (1983, April). Preschool children's learning and transfer of matrices problems: Potential for improvement. Paper presented at the Society for Research in Child Development, Detroit, MI.

Coles, G. S. (1978). The learning-disabilities test battery: Empirical and social issues. Harvard Educational Review, 48(3), 313-340.

Cooper, D. H., & Farran, D. C. (1984). Cooper-Farran Behavioral Rating Scales. (Available from D. H. Cooper, Department of Special Education, University of Maryland, College Park, MD, 20742.)

Cooper , D. H. (1984). Identification of critical behaviors in kindergarten. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk factors in kindergarten. Early Childhood Research Quarterly, 3, 1-19.

Feagans, L., & McKinney, J. D. (1981). The pattern of exceptionality across domains in learning disabled children. Journal of Applied Developmental Psychology, 1, 313-328.

Feagans, L. & Short, E.J. (1984). Developmental differences in the comprehension and production of narratives by reading-disabled and normally achieving children. Child Development, 55, 1727-1736.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of indirective reasoning rules: Studies of proximal development. Child Development, 57, 1087-1099.

Finn, J. D., & Resnick, L. B. (1984). Issues in the instruction of mildly mentally retarded children. Educational Researcher, 13(3), 9-11.

Good, T., & Brophy, J. (1978). Looking in classrooms (2nd ed.). New York: Harper & Row.

Good, T. L. (1983). Classroom research: A decade of progress. Educational Psychologist, 18, 127-144.

Greenwood, C.R., Schulte, D., Kohler, F., Dinwiddie, G. & Carta, J. (1986). Assessment and analysis of ecobehavioral interaction in school settings. In R. Prinz (Ed.) Advances in behavioral assessment of children and families (Vol. 2, pp. 69-98). Greenwich, CT: JAI Press.

Heller, K. Holtzman, W., & Messick, S. (1982). Placing children in special education: A strategy for equity. Washington, D.C.: National Academy Press.

Helper, M., M. (1980). Follow-up of children with minimal brain dysfunctions: Outcomes and predictors. In H. E. Rie & E. D. Rie (Eds.), Handbook of minimal brain dysfunctions (pp. 75-114). New York: John Wiley & Sons.

Kleinbaum, D., Kupper, L., & Morgenstern, H. (1982). Epidemiologic research: Principles and quantitative methods. Belmont, CA: Wadsworth.

Lloyd, J., Saltzman, N., & Kauffman, J. M. (1981). Predictable generalization in academic learning as a result of preskills and strategy training. Learning Disability Quarterly, 4, 343-365.

McKinney, J. D., & Speece, D. L. (1983). Classroom behavior and the academic of learning disabled children. Journal of Applied Developmental Psychology, 4, 149-161.

Newcomer, P. L., & Curtis, D. (1984). Diagnostic Achievement Battery. Austin, TX: PRO-ED.

Peterson, L., Homer, A.L., & Wonderlich, S.A. (1982. The integrity of independent variables in behavioral analysis. Journal of Applied Behavioral Analysis, 15, 477-492.

Reschly, D. J. (1984). Beyond IQ test bias: The national academy panel's analysis of minority EMR over-representation. Educational Researcher, 13(3), 15-19.

Salvia, J., & Ysseldyke, J.E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton-Mifflin.

Satz, P., & Fletcher, J. M. (1980). Minimal brain dysfunctions: An appraisal of research concepts and methods. In H. E. Rie & E. D. Rie (Eds.), Handbook of minimal brain dysfunctions (pp. 669-715). New York: John Wiley & Sons.

Satz, P., & Morris, R. (1981). Learning disability subtypes: A review. In F. J. Pirozzolo & M. C. Wittrock (Eds.), Neuropsychological and cognitive processes in reading (pp. 109-141). New York: Academic Press.

Speece, D. L. (1987). Information processing and reading in subtypes of learning disabled children. Learning Disabilities Research, 2, 91-102.

Speece, D. L., McKinney, J. D., & Appelbaum, M. I. (1985). Classification and validation of behavioral subtypes of learning-disabled children. Journal of Educational Psychology, 77, 67-77.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time-on-task. Educational Researcher, 9(10), 11-16.

Stanley, S. O., & Greenwood, C. R. (1981). CISSAR: Code for instructional structure and academic response: Observer's manual. Kansas City, KS: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.

Swanson, H. L. (1983). Effects of rehearsal activity and level of word processing on learning disabled and normal readers' free recall. Journal of General Psychology, 109, 61-72.

Thorndike, R. L. & Hagen, E. P. (1982). Cognitive abilities test, primary battery edition. Chicago: Riverside.

Torgesen, J. K., & Houck, D. G. (1980). Processing deficiencies of learning-disabled children who perform poorly on the Digit Span Test. Journal of Educational Psychology, 72, 141-160.

Torgesen, J. K. (1977). Memorization processes in reading-disabled children. Journal of Educational Psychology, 69, 571-578.

Table 1

Child Characteristic Measures
and the Cut-off Scores for Definition of Risk Factors

Measure	Summary Score	Cut-off score
CogAT[a]	Verbal Standard Age Score	< 85
DAB[b]	Reading quotient	< 85
CFBRS[c]	Work-related Skills (WRS) Interpersonal Skills (IPS)	< 4 < 4
DAT[d]	Numbers of Prompts	> 4
PLAI[e]	Cognitive Level 4	< 2

<u>Note:</u>

[a]Cognitive Abilities Test. Mean = 100, S.D. = 15

[b]Diagnostic Achievement Battery. Mean = 100
S.D. = 15

[c]Cooper-Farran Behavioral Rating Scales.
WRS Mean = 5 S.D. = 1
IPS Mean = 5 S.D. = 1

[d]Dynamic Assessment Tasks.
Number of Prompts = 2.4 S.D. = 1.6
(Fewer prompts indicate better performance.)

[e]Preschool Language Assessment Instrument.
Mastery Score = 2

Table 2

Univariate Relative Risks

Measure	Relative Risk	Confidence Interval[a]
CogAT		
Non-Verbal IQ	3.2	0.4-27.1
CFBRS		
Interpersonal Skills	4.1	1.0-16.4
Work-Related Skills	23.1	5.9-90.2
DAB		
Reading Achievement	4.5	1.3-14.9
DAT		
Prompts to Criterion	1.9	0.2-21.9
PLAI		
Language Reasoning	0.8	0.2-4.2

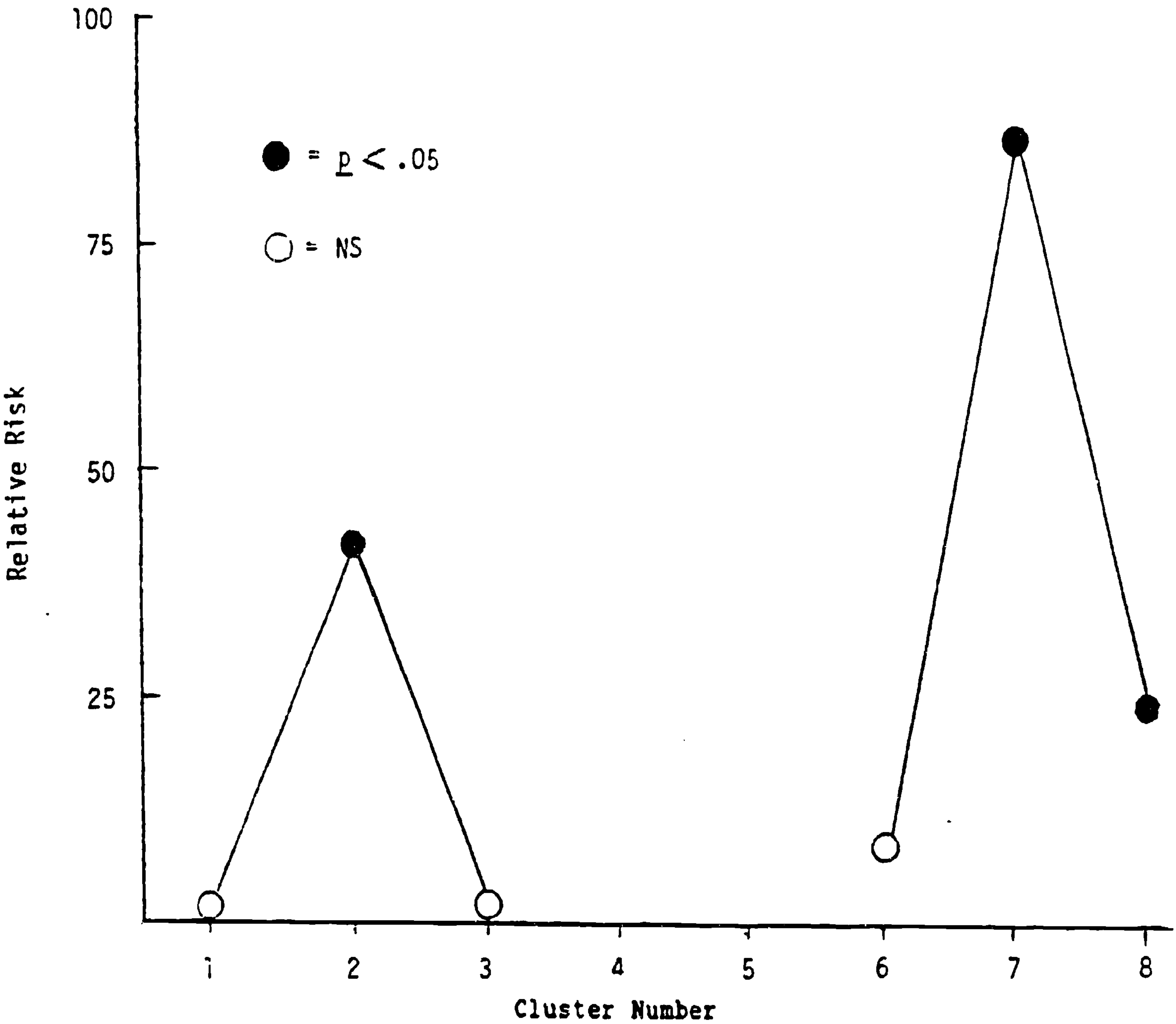
[a]Lower and upper bounds of 95% confidence intervals.

Figure Captions

Figure 1. Example of relative risk analysis.

Figure 2. Relative risk as a function of patterns of risk factors (VIQ= CogAt verbal IQ, RDG= DAB reading achievement, WRS=CFBRS work-related skills).

RISK FACTOR	OUTCOME: Referred	OUTCOME: Not Referred
Present (low WRS)	9 a	38 b
Absent (adequate WRS)	6 c	80 d



Cluster Definition

(0 = Absence, 1 = Presence of Risk Factor)

	1	2	3	4	5	6	7	8
VIQ =	0	0	0	1	1	1	0	1
RDG =	0	0	1	0	1	0	1	1
WRS =	0	1	0	0	0	1	1	1
N =	10	6	3	0	0	1	12	3

Study Two: Instructional Correlates

Abstract

This paper describes instructional arrangements in first grade classrooms and assesses their effects on academic responses of students at risk for failure. A parallel set of analyses of not-at-risk controls provides a basis for comparison. Time-sampled, in-class observations focused on multiple components of instruction (subject-matter, tasks assigned, group size and teachers' behavior) and students' response (active/academic, task management, and inappropriate). The associations of instructional structure with type of student response were considered significant if replicated across two samples. Relative to base-rates (unconditional probabilities) a number of specific arrangements were shown to be associated ($p < .003$) with either accelerated or decelerated rates (conditional probabilities) of at-risk and control students' active, academic or inappropriate responses. Although controls were more academically active than their at-risk peers, control and risk groups appeared to respond to similar arrangements. Implications regarding classroom practices and the achievement-related value of those instructional arrangements on students' academic responding are discussed.

Instructional Correlates of Students' Academic Responses: Comparisons Between At Risk and Control Students

Studies conducted over the last decade have elaborated on Carroll's model of classroom learning (1963) linking time-on-task and students' achievement. The partitioning of academic time into specific arrangements of instruction has demonstrated differential effects on achievement. Brophy (1979) and Rosenshine & Berliner (1978) have reviewed the literature, and found converging evidence to support the effectiveness of instruction characterized as "direct instruction" (Brophy, 1979) or "academic engaged time" (Rosenshine & Berliner, 1978). The most effective instructional arrangements were found to include a sequenced combination of explanation and demonstration, followed by opportunities to practice at a high rate of accuracy, with immediate corrective feedback. In contrast, independent silent reading, writing and listening without opportunities to make responses to which feedback was accessible were negatively associated with achievement. These findings were generated for the most part from studies of children achieving in the normal range, but have been subsequently supported by studies of children at risk for failure (Stallings, 1980; Englert, 1984).

The conceptualization of this study was based on the work of Greenwood and his colleagues, (Greenwood, Delquadri, Stanley, Terry and Hall, 1985) who presented a detailed eco-behavioral study of instructional arrangements' effects on fourth grade students' academic responding. Briefly, Greenwood et al. (1985) concluded that teacher/student discussion decelerated academic responding, while paper/pencil tasks accelerated academic responding. The results to be presented in this paper extend this line of inquiry by examining the classroom ecology of primary grade students at risk.

As Good (1983) pointed out, the time-on-task literature, while showing the linkage with achievement, has failed to account for observed variance in time-on-task. Good further suggested the need for classroom process research that would integrate teacher, student, and curriculum/task variables. The present study addresses these issues directly: an integrated view of classroom instructional ecology is adopted to explain student response variance.

This report presents data on the features of classroom environments that enhance active academic responses for first grade children at risk for failure and not-at-risk controls. Specifically, classroom instructional arrangements (activity, structure, task, and teacher behavior) are described which are associated with increased or decreased academic responding of students.

The probabilities of students' responses, conditional upon each of sixteen variants in the instructional ecology, were compared with unconditional probabilities (base-rates) of those student responses, and the differences tested for significance to identify arrangements that accelerated or decelerated academic and inappropriate responding. We were specifically interested in assessing whether at-risk and control students responded similarly to frequently occurring instructional structures.

Method

Subjects

The present investigation is part of a larger study in which the target population was first grade students at risk for referral to special education, but who were not yet classified as handicapped. The two public school districts that participated were selected because "pre-referral" identification systems were used in each. Elementary schools in both districts used variants of the Teacher Assistance Team (TAT) model to identify students at risk prior to referral to and evaluation by special education teams. The TATs were composed of regular and sometimes special education teachers whose duties included reviewing the TAT referral with the referring teacher, providing alternatives for instruction and/or behavior management, and monitoring students' progress in the regular classroom.

Inclusion criteria. The selection of at-risk first graders for this study was predicated on the following criteria: (a) referral to TAT for academic or behavioral problems, (b) the children were in first grade for the first time, (c) the children's native language were English, and (d) if supplementary instruction, such as Chapter One assistance, was provided outside the regular classroom, the total amount of time did not exceed 2 1/2 hours per week. To reduce the data collection burden on any one teacher, we included only the first two at-risk children nominated per classroom. (Referral of more than two students per classroom was a rare occurrence.) When permission for inclusion was obtained from the parents of an at-risk child, a control child of the same sex and from the same classroom was randomly selected from the teacher's list of children achieving and behaving normally, and who, in the teacher's opinion, would not be referred either to TAT or programs for the gifted and talented. Thus, the initial subject pool contained equal numbers of at-risk and control children. Parent permission return rates for this sample were 66% for at-risk students and 82% for control.

The participating school districts were adjacent to the Washington, D.C., metropolitan area. County A had a general population of 660,000 with a racial composition of 46% Black, 49% Caucasian, and 5% other ethnicities. The total enrollment for the school system was 103,325 with 62.6% Black, 31% Caucasian, and 6.4% other. Twenty-eight schools from County A are represented in the sample. The population for County B was 414,074 with a racial distribution of 12.6% Black, 85.4% Caucasian, and 2% other ethnicities. The school system population was 64,552 and reflected the racial compositi... of the general population with 14.2% Black, 83.8%

Caucasian, and 2% other ethnicities. Eleven schools in County B are represented in the sample.

To strengthen the conclusions of this exploratory study, we took advantage of the availability of two independent samples, each including at-risk (AR) and control (C) students. The derivation sample consisted of all subjects (AR: n = 67, C: n = 58) recruited in years one and two of the project (1985-86, 1986-87), while the replication sample consisted of the third-year's recruits (AR: n = 40, C: n = 39). Results will be reported for analyses that were replicated across these two samples.

Table 1 provides descriptive data for the samples. In comparison to the general school enrollment population data, our sample is disproportionately male (69%). With regard to socio-economic status, the samples appeared to be close to the national median for mother's educational level (AR median = 12 years, C median = 13 years, U.S. median = 12.6 years; Bureau of the Census, 1985).

The child characteristic data in Table 1 for intelligence, achievement and teacher-rated behavioral skills underscore the basic differences between at-risk and control groups while indicating within group comparability of derivation and replication samples.

[INSERT TABLE 1 ABOUT HERE]

Teachers. First grade teachers in the participating schools were included in the study when a referral to TAT from their class resulted in a subject meeting all inclusion criteria. Teacher background data were available on 72 of 80 participating teachers. This information suggested the teachers were well trained. Eighty-two percent had at least 15 hours past the bachelor's degree, and in the appropriate specialty areas: 87% in elementary or early childhood education. The teachers were also, as a group, experienced in teaching first grade: x = 10 years, SD = 8. Class-size was typical for public school first grades: x = 25, SD = 3.7.

Procedures. Children in the risk group were observed on two occasions for a total of thirty minutes. Controls were observed on two occasions totalling twenty minutes. Control children's observations were interspersed among the risk children's, with observers focusing on one child at a time for five minutes, then switching to the other child for five minutes, and so forth. Teachers were observed concurrently, were given no special instructions, and told only that we were interested in observing normal classroom routines for the target students. Observations were conducted between January and June and occurred during morning instructional times, following morning business and prior to lunch.

After being trained to 80% agreement with a standard observation protocol, two observers maintained a minimum of 80% inter-observer agreement on each of 5 aggregated coding categories throughout the study. Thirty-nine observer agreement sessions were conducted (9.9% of the total 385 data collection sessions). Means and standard deviations for percentage agreements for the five aggregated categories were as follows: Activity = 98% (5), Task = 94% (8), Structure = 100% (0), Teacher behavior = 89% (10), and student behavior = 82% (6). These values are quite similar to data reported by the instrument's authors (reported below, Greenwood, Schulte, Kohler, Dinwiddie & Carta, 1986)[1].

Instrumentation. Observers used the Code for Instructional Structure and Student Academic Response (CISSAR, Stanley & Greenwood, 1981). As described by Greenwood et al. (1985) CISSAR provides codes in five areas (see Table 2): Activity (the subject of instruction), Task (the materials or stimuli set by the teacher to occasion responding), Structure (grouping for instruction), Teacher Behavior (relative to the target student), and Student Behavior (specific, active academic responses, prerequisite or enabling responses, or inappropriate responses). Teacher and student behaviors are an instantaneous time-sample, updated each ten seconds. Activity, task and structure are updated each minute.

Reliability and validity of CISSAR were established during instrument development and are well documented, (see Greenwood et al., 1986). Briefly, test-retest (or intra-observer) reliability coefficients were examined for each of 53 coding categories. The mean coefficient was .88, and ranged from .35 to .93 (Greenwood et al., 1986). Inter-observer reliability, examined both at the level of single intervals, and sequences of intervals, is also adequate. Mean percent agreements (interval-level) ranged from 86.3% (students' behaviors) to 99.1% (activities). At the level of sequences, conditional probability data were examined for interobserver agreement by Pearson correlations, and ranged from .76 to .96 across a number of conditional probabilities (Greenwood et al., 1986). Evidence regarding predictive validity of CISSAR has been reported, using criteria of achievement (Greenwood et al., 1986) and differences between Title I and non-Title I groups (Greenwood et al., 1985).

CISSAR codes were adapted slightly (with the authors' permission) for purposes of this study. Modifications to the activity code definitions were minor, and simply aimed to reduce ambiguity arising from assignments given to the target student that differed from the rest of the class. No task codes were changed. The structure code "entire group" was modified to include recitation or other target student responses directed to the whole class. The structure code "individual" was sub-divided by defining two variants on working alone: "individual with teacher" was coded for extended interactions between the target student and teacher, on tasks that may or may not have differed from the rest of the class; "Individual with peer" was coded when the target student worked with one peer, such as in a peer-tutoring arrangement.

The original CISSAR instrument included a "teacher position" code which was not used in the present study primarily to accommodate the major modifications of teacher behavior codes described below. The global code "teaching" was dropped, and replaced by two differentiated codes: *direct* and *indirect* teaching. "Direct teaching" was coded during actual instruction, such as lecturing about factual material, strategies, or providing explanation of concepts or processes. This code captured teacher behavior that conveyed academic content. "Indirect teaching" was coded for behaviors that were intended to be facilitative of academic responses, such as cuing, questioning, monitoring students' responses, reading aloud, or giving directions to elicit student responses. In another adaptation, teacher behaviors directed to students *other* than the target student were specifically coded as either approval of, disapproval of, or teaching other children, and in this way the target students' specific instructional ecology could be more precisely recorded.

Analysis. For the present study, we dichotomized composite codes within each of four CISSAR instructional categories with student responses divided into three types (see Table 2). Twelve activity codes were collapsed into two types: academic (e.g. reading, math, spelling) and non-academic (e.g. business management, transition). Eight task codes were collapsed into two composites based on the demands on the child: active (reader, workbook, worksheet, paper/pencil) and passive (listening to lecture, other media, teacher-student discussion, fetch-put away). The five structure codes were likewise dichotomized based on instructional group-size: focused (small group, one-on-one with teacher) and diffuse (entire group, individual with peer, alone). The nine teacher behavior codes were collapsed with respect to relation with target child: engaged (directly or indirectly teaching target child, approval or disapproval of target child's response) and disengaged (teaching, approving or disapproving other children, no response, other talk). Student Behavior codes were collapsed into three composites: academic (writing, academic game, reading aloud or silently, talking appropriately, answering or asking academic questions), task management (attending, raising hand, looking for materials, moving to new academic stations, playing appropriately) and inappropriate (disrupting, inappropriate talk, task or locale, looking around, self-stimulation). Thus eleven aggregated coding categories were formed.

Modelled on the Greenwood et al. (1985) strategy, values of the activity, task, structure and teacher aggregates were combined in all sixteen possible "Composite Ecological Arrangements" (CEA, defined in Table 3) to determine their relationship to the student response composites (academic, inappropriate). Analyses were conducted to examine behavior within risk and control groups of subjects. The probabilities of students' academic responses, conditional upon each of the CEAs, were examined to determine which CEAs were associated with risk and control subjects' accelerated and decelerated academic responses, relative to the unconditional or "base rate" probabilities. Differences in probabilities (within group, conditional vs. unconditional) were tested via the _z_-statistic proposed by Allison & Liker (1982). A comparable set of analyses examined each CEA's association with students' inappropriate responses. All analyses were conducted using SAS software (1985).

Serial dependency. Following the recommendation of Gardner and Hartmann (1984), the CISSAR data were examined for serial dependency. This preliminary analysis of sequential data is necessitated by the potentially inflating influence of serial dependency on analyses of associations between behavioral streams of two interactants. If the goal is to determine the amount of variance in one interactant (e.g., the student) that is attributable to variance in the other interactant (e.g., the instructional ecology), then it is essential to "partial out" effects within each interactant's behavioral stream. A high degree of serial dependency indicates predictability of an interactant's behavior at time 2 from the same interactant's behavior at time 1. It is this degree of predictability that must be statistically controlled before the effects of another interacting behavior stream can be reliably assessed. Separate analyses were conducted for each of 16 CEAs and student academic responses, for both risk and control groups. These analyses revealed a high degree of lag 1 dependency, that is, a tendency for CEA and student response codes to remain the same from one 10-second interval to the next. Phi correlations (an index of dependency) for CEA data were very high, generally in the .80 to .90 range, and, for student responses, moderately high, in the .50 to

.60 range. Risk and control data were similar in this regard.

After it was determined that serial dependency was a feature of both the instructional (CEAs) as well as response codes, a correction factor (gamma) proposed by Dumas (1986) was applied to the z-statistics before referring to the tabled values for significance testing. Thus, tests of significant differences between conditional and unconditional probabilities controlled for the serial dependency in both instructional and response codes.

Significance level. Due to the exploratory nature of this inquiry, a conventional alpha of .05 was chosen for the overall level of significance for each set of 16 CEAs tested (one set for each sample [risk derivation, risk replication, control derivation and control replication] and for each student response). In order to preserve the probability level at .05 while testing 16 separate z-statistics for each set (one for each of 16 CEA's), Dunn's multiple comparison approach was adopted. Thus, each z, after the gamma correction was applied, was tested at .05/16, or .003.

Replication. The availability of CISSAR data for two independent samples allowed the results of the analyses to be subject to tests of replication. Following the example of Greenwood et al. (1985), results were obtained separately for derivation and replication samples. Results of all analyses will be presented, but conclusions will be based on replicated results.

Results

Descriptive Data

Before presenting the effects of CEAs upon student responses, descriptive data regarding each aggregated category by sample and group are presented. Unconditional probabilities of occurrence for each of 16 CEAs and 3 student responses are given in Table 3 and Table 4 respectively. Results are given separately for derivation and replication samples, and for risk and control groups. (Probabilities have been converted to percentages in Tables 3 and 4. The n given for each column is the total number of intervals coded for the group). Table 3 is described first.

Activity. The subject matter during the observation sessions was predominantly academic in nature (CEA 1-8). Across samples, over 90% of the sequences observed were instructional periods coded as reading, math, language, etc., as opposed to free time, transitions, etc. (Detailed breakdowns are provided in a later section). The vast majority of morning time in these first grades was allocated to academic instruction.

Task. Students were expected to be engaged in tasks requiring an active response, to which correctional feedback could be available (CEA 1-4). Active tasks such as worksheet, workbook and paper-pencil tasks (permanent products) appeared to occur more frequently (approximately 56%) than passive tasks such as listening to lecture, or teacher-student discussion.

Structure. Grouping for instruction was primarily focused (CEA 3, 4, 7, 8), with students divided into reading groups (with or without the teacher present). First-grade teachers in this sample used the morning to meet individually with reading groups, while the other students worked on a common seat-work task, usually involving some type of written response. Approximately 70% of the sequences were coded as "small group" (summed across CEAs 3, 4, 7 and 8).

Teacher. With respect to target students (at risk and control), teachers were disengaged, that is, teaching other students, approximately 60% of the intervals coded (CEAs 2, 4, 6, 8, 10, 12, 14, 16). This reflects the previous finding on structure: teachers instruct small reading groups while the majority of the students at any given time of the morning are not in the reading groups, rather, are working on individual seat-work.

Composite Ecological Arrangements (CEA). CEA 4 and to a lesser extent, CEA 3 were the predominant instructional arrangements, occurring approximately 30% and 13%, respectively, of all intervals. Again, this reflects the reading group strategy alluded to above. Next most frequent was CEA 5, occurring from 7% to 13% of intervals. This arrangement represents a lesson involving the entire class, and generally includes a teacher-focus, with students expected to attend and listen. In a subsequent section, the six most frequently occurring CEAs are further analyzed with respect to specific codes and their effects on student responses.

Student responses. Table 4 gives unconditional probabilities (converted to percentage occurrences) of student responses. The three Composite categories follow Greenwood et al. (1985): academic, task management and inappropriate. For all samples and groups, the task management composite was coded approximately half of all intervals, and consisted largely of the "attending to task" code.

Conditional Probabilities of Academic Responses

Table 5 gives the percentages of student responses coded in one of the academic categories (e.g. reading, writing, answering questions, etc.) conditional upon each CEA. Table 6 presents similar data for students' inappropriate responses. Entries in the tables for each of two samples (derivation, replication) and two groups (risk, control), are conditional probabilities (converted to percentages), and may be compared to the unconditional probability (given as UCP at the head of each column) to determine the accelerating or decelerating effect on academic responding of each CEA. The z statistic (corrected for serial dependency in both CEA and student responses) tests the difference between the conditional and unconditional probability. Significance of z is indicated at the conventional level ($p < .05$) and at a more stringent level to control the alpha within each set of 16 CEAs ($p < .003$). In the following descriptions of results, emphasis is given to findings that were strictly replicated at the $p < .003$

level for derivation and replication samples. However, in a few instances in which effects were consistent across samples, but one of the four contrasts failed to achieve significance, non-significant trends ($p < .10$) are noted by reporting z and p. Replicated results are described in descending order of CEAs' frequency of occurrence.

CEA 4. Academic activity, active task, focused structure and disengaged teacher define CEA 4. Occurring approximately 30% of the time, CEA 4 appears to accelerate academic responding in control, and to a less certain extent, risk samples, (risk group, replication sample: $z = 1.64$, $p < .10$.) Despite the teacher's being disengaged, both risk and control students' active, academic responses (writing, coloring) were greater than unconditional probabilities. Although the control students' absolute level of academic responding is higher than risk students', the conditional probabilities suggest that the accelerating effect on academic responding is equivalent for the two groups. Complicating the picture are the results of analyses of the effect of this CEA on inappropriate responding (see Table 6). Conditional on CEA 4, both risk and control students were significantly more likely to respond inappropriately in comparison to their unconditional probabilities. Analysis of the specific codes that form the composite ecological arrangement suggests the following typical scenario for CEA 4. The target student is working on a paper/pencil or worksheet task while the teacher meets with one of the other reading groups. The task is related to reading; perhaps a copying or coloring task that reinforces a phonetic skill. Typically, the inappropriate behavior during CEA 4 was "looking around".

CEA 3. Academic activity, active task, focused structure and engaged teacher define CEA 3. The second most frequent arrangement for both risk and control groups (13% and 11%, respectively), CEA 3 decelerates academic responding of controls (Table 5) and decelerates inappropriate responding for risk students (Table 6). It appears that both types of students were passively attending to teachers or other students. This CEA did not affect control students' rate of inappropriate responses, nor the risk students' academic rerponses. The scenario for CEA 3 is the reading group; someone was reading aloud (not the target student), while the target student and teacher were listening or following along in their readers. Relative to the number of opportunities for listening, the number of instances of the code "read aloud" were few (risk: 4.8% of CEA 3; control: 3.2% of CEA 3). CEA 3 indicates that the teacher was engaged, with "indirect teaching" the predominant teacher behavior during this arrangement.

CEA 5. Academic activity, passive task, diffuse structure and engaged teacher define CEA 5. CEA 5 occurred at the third-highest frequency (risk: 10.7%; control: 9.5%). Academic responding was significantly decelerated in both groups. Inappropriate responding was not affected. Typical of CEA 5 were discussions among the teacher and the entire class. Language, reading or spelling were the predominant subjects. Teachers were indirectly teaching (i.e., asking questions or listening), while students, risk less so than controls, were attending to task (i.e., listening to and looking at the teacher).

CEA 8. Academic activity, passive task, focused structure and disengaged teacher define CEA 8. Occurring approximately 9% of all intervals coded, CEA 8 had an accelerating effect on risk students' academic responding but a non-replicated effect on controls. Results for inappropriate responding were inconclusive (see Table 6).

CEA 8 appeared to succeed CEA 4 in this fashion: When the target student was not in the teacher's current reading group, and had completed the paper/pencil task (CEA 4), a free-choice time occurred (CEA 8). The subject-matter (activity code) continued to be reading, but the task switched to "other media" -- typically involving an academic game, puzzle or arts and crafts materials. The teacher continued to teach the other reading group, but was likely to be disengaged from them as well, to begin correcting papers or "other talk" (talking about class business, schedules, etc.).

CEA 7. Academic activity, passive task, focused structure and engaged teacher define CEA 7. The probability of occurrence of CEA 7 varied across groups and samples (See Table 3), ranging from 3.6% (control, replication) to 12.1% (risk, derivation). Academic responding was significantly decelerated in both risk and control groups; however, the significance level for the risk/replication sample failed to achieve the strict criterion for replication (z = 2.24, p < .05). While risk students' active, academic responses were suppressed by the teacher's occupying the "academic stage", inappropriate responses were also reduced for the risk but not control students (see Table 6). CEA 7 appeared to capture that portion of the teacher's reading group time when reading *per se* was not the task; rather reading-related vocabulary or specific skill instruction was occurring. Teachers used the board or a discussion format, and thus, were the focus of risk and control students' "attending to task" responses.

CEA 1. Academic activity, active task, diffuse structure and engaged teacher define CEA 1. Although CEA 1 was among the least frequently occurring arrangements (risk, approximately 6%; control approximately 6%), it appeared to successfully accelerate academic responding in both risk and control groups. (Note: the accelerating effect in the risk/derivation sample failed to reach significance [z = 1.6, p < .10], but was consistent with all other samples). No effect on inappropriate responding was observed. Typically, CEA 1 represented a whole-class spelling test, or other dictation activity related either to a specific reading or language skill, handwriting, or paper/pencil math task. This CEA was an especially potent accelerator of risk and control students' writing responses.

Discussion

The classroom ecology and observable responses of first graders were documented in an attempt to identify functional relationships between various instructional arrangements and academic responses. Subject selection and data analyses were designed to provide results that could be compared and contrasted between students at risk for failure and their not-at-risk peers. The dual focus on students' active, academic responses and opportunities provided for such responses was based on consistent findings in the teacher effects literature linking enhancement of achievement to opportunities to respond (Brophy, 1979). Analytic strategies designed to ensure validity of the results included correcting for serial dependency, controlling for inflated alpha within sets of analyses, and replicating results in independent samples. These statistical tests of differences between unconditional and conditional probabilities of students' academic responses revealed interesting patterns regarding differential associations of ecological arrangements with risk and control students' academic and inappropriate behavior in those settings.

Active, academic responding. As defined by Greenwood et al. (1985), the academic response composite comprised those responses "... in which the student was noted to be actively engaged in academic behavior..." (p. 334). In the present study, two CEAs (1 and 4) appeared to be most effective in eliciting these responses in both groups, while CEA 8 accelerated academic responding only in the risk group. The only common component of these three CEAs was an academic activity, especially reading, and to a lesser extent, spelling. Paper and pencil tasks, both in large (CEA 1) and small (CEA 8) group structures, appeared to elicit the desired responses (writing or coloring) in both risk and control groups. These results are in general agreement with Greenwood et al. (1985), who found that Title I and non-Title I students responded academically to paper/pencil tasks.

Active academic responses were accelerated in risk, but not control groups, by academic games or arts and crafts related to reading (CEA 8) while the teacher was meeting with other students in reading group. It is not known whether or not this result is congruent with Greenwood et al., (1985) as only spelling activities were presented. Risk and control students appeared to be indistinguishable during CEA 8 in terms of frequency of active, academic responses, and in this regard, it was among the most productive instructional arrangements.

Although the paper/pencil and game responses accelerated by CEAs 1, 4 and 8 are clearly desirable, it cannot be assumed that enhanced achievement will result for students at risk. Brophy (1979) concluded that achievement is enhanced when students' responses are germane to the immediate learning objectives and susceptible of immediate corrective feedback. With the possible exception of CEA 1 (engaged teacher) the degree to which these conditions were satisfied in the accelerated responses observed here is questionable. Paper/pencil tasks may be germane, but generally are not amenable to immediate, corrective feedback. Academic games often provide feedback (e.g., only the correct letter fits in an alphabet puzzle), but without the teacher's presence, as is the case with CEA 8, the appropriateness of the task and feedback are not ensured. Thus, the arrangements found to accelerate academic responding should not be translated directly to classroom practice without attention to these important links with achievement.

Only one arrangement, CEA 5, decreased academic responding for all children, while CEA 3 and CEA 7 were associated with decreases for at-risk and control students, respectively. The common feature across these CEAs was an engaged teacher with student responses being primarily attention to task, a task management code. This counterintuitive finding of engaged teachers associated with decreased academic responding is explained, in part, by our approach of analyzing concurrent associations (lag zero). Within intervals, the teacher and the student cannot simultaneously occupy the active, academic stage. Therefore, when the teacher or another student is the focus of the group's attention, the target student's choice of responses is generally limited to passive attending at best, or ina' ,priate behavior (typically looking around) at worst. These results parallel Greenwood et al. (1985), who also reported that during spelling instruction, teaching behavior, either to the entire group or to a small group, decelerated students' academic responding. While a lagged sequential analysis may uncover different relationships between engaged teachers and academic student behavior, evidence from CEA 1 suggests the analyses used were sensitive enough to detect these relationships when they occurred. This arrangement consisted of an engaged teacher in a group setting and was associated with increased academic responding for all children. One reason for the differential association with academic responding among these CEAs (1 vs. 5,3,7) may be in how the teacher sequences direct and indirect teaching experiences.

Inappropriate responses. The students in our samples appeared, paradoxically, to be both at their best and at their worst when not in the teacher's reading group. The worksheets to be completed individually (CEA 4) elicited the desired writing or coloring responses, but, not surprisingly, failed to hold students' attention for very long, given that the teacher was involved with other children. As a result, both inappropriate responses, and academic responses were accelerated under this condition. One possible explanation, proposed by Doyle (1983), is the often observed high degree of mismatch between instructional materials and student ability. Brophy (1979) concluded that the optimal level of difficulty for this type of independent work was indicated by better than 90% accuracy in student responses. For low achieving, at-risk students, the optimal worksheet will therefore need to be intrinsically motivating, and carefully selected to provide practice in skills known by the teacher to have been acquired previously. Otherwise, the primary instructional objective of practice, that is fluency-building, will not be achieved, and the probability of inappropriate behavior increases. Also, the sheer amount of time devoted to CEA 4 (approximately 30% of the observational periods) may account for the relationships with student behavior. The tasks may match the child's independent level of functioning but simply lose their appeal. An investment of teaching resources targeted on at-risk students not in reading groups may have significant return. Adult or peer tutors, cooperative learning, individualized computer-assisted instruction or other media are possibilities that have the potential to provide at-risk students the requisite degree of difficulty along with immediate corrective feedback.

Because teachers have responsibility for educating all children in the classroom, student behavior associated with CEA 3 and CEA 7 present a dilemma. Under these conditions, the inappropriate behavior of at-risk children is decreased as is the academic responding of control students. The result appears to be related to increased "attention to task". While this "good news/bad news" scenario may, in fact, be desired by teachers, a positive relation between passive attending and achievement should not be assumed (Trenholm & Rose, 1981).

Teaching. In contrast to Greenwood et al. (1985) in the present study engaged teaching behavior was subdivided into direct and indirect teaching. The overwhelming majority of engaged teaching was indirect, that is, asking questions, listening to students read, etc., as opposed to direct, that is, telling, answering, correcting, or presenting. It is interesting to note that the at-risk students, by increasing attending behavior responded to these instructional arrangements in much the same fashion as control students. If teachers were to adapt their instruction by increasing their overall engaged time with the at-risk students, but in the process decrease at-risk students' opportunities to make active, immediately correctable, academic responses then the effect on achievement is a matter for concern. These students, by virtue of having been referred to teacher assistance teams, have demonstrated less than optimal levels of performance, and therefore, are in need of more, not fewer opportunities for responding as well as increased amounts of direct teaching that includes corrective feedback to their oral, written and/or other (game) responses. At the same time, it is important to note that Good (1983) concluded that active teaching (direct teaching, in our terminology) is the key to increased student achievement.

Rosenshine (1980, cited in Good, 1983) noted the effect of direct teaching: students' academic engaged time rises from 70% during unsupervised seatwork to 84% during a teacher-led discussion. Our approach to analysis of student response, differentiation of active and passive academic engagement (Greenwood et al., 1985), suggests that while students were engaged in Rosenshine's sense, our subjects were not active participants during teacher-centered instruction (CEA 5 and CEA 7). The significance of this finding is emphasized by the findings of Peterson and Swing (1982), who determined from student interviews, that what appeared to observers to be attention to task, was not necessarily isomorphic with cognitive processing of instruction.

Thus we have a contradiction in that we seem to be suggesting the need for both more active teaching as well as more active learning, which by our definition cannot occur simultaneously. The solution may well lie in Doyle's (1983) recommendation regarding the importance of student understanding during direct instruction: "It is essential, therefore, that direct instruction include explicit attention to meaning and not simply focus on engagement as an end in itself" (p. 189). During instruction of low-ability or at-risk students, this may mean frequent, planned interruptions in the teacher's presentation to probe student understanding and to provide corrective feedback to students' incorrect responses (Brophy, 1979). These students in particular may need to be given frequent opportunities to think about the material at hand and to demonstrate their thinking while feedback is accessible (Marx & Walsh, 1988).

With appropriate attention to issues of instructional match, quality of students' attention, and feedback, the results of this study provide several points of departure for research and practice. Given that at-risk and control students responded similarly to a variety of instructional arrangements, it would be important to determine if increased time in the productive arrangements would net improved achievement. Ecological structures related to task management (primarily attending to task) were not explored in this study. However, given that these responses were the most frequently observed, there may be differences between at-risk and control students' use of this time that are related to differential achievement patterns. For example, Anderson, Brubakar, Alleman-Brooks and Duffy (1983) specifically propose that information processing differences between high and low achievers may take the form of metacognitive awareness of task difficulty, triggering a problem-solving response in high but not low achievers. Finally, Good's (1983) suggestion that understanding time-on-task variance requires attention to multiple environmental factors found support in this study. More detailed content analyses (e.g., Eder, 1981) may be profitable in uncovering why some arrangements are more profitable than others.

References

Allison, P.D. & Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. Psychological Bulletin, 91 (2), 393-403.

Anderson, L.M., Brubaker, N.L., Alleman-Brooks, J. & Duffy,G.G. (1983). A qualitative study of seatwork in first-grade classrooms. The Elementary School Journal, 86 (2), 123-140.

Brophy, J. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 773-750.

Bureau of the Census (1985). Statistical abstract of the United States. Washington, DC: U.S. Department of Commerce.

Carroll, J.B. (1963). A model of school learning. Teachers College Record, 64, 723-733.

Doyle, W. (1983). Academic work. Review of Educational Research, 53 (2), 159-199.

Dumas, J.E. (1986). Controlling for autocorrelation in social interaction analysis. Psychological Bulletin, 100 (1), 125-127.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. Sociology of Education, 54, 151-162.

Englert, C.S. (1984). Effective direct instruction in special education settings. RASE, 5(2), 38-47.

Gardner, W. & Hartmann, D.P. (1984). On Markov dependence in the analysis of social interaction. Behavioral Assessment, 6, 229-236.

Good, T.L. (1983). Classroom research: A decade of Progress. Educational Psychologist, 18, 127-144.

Greenwood, C.R., Delquadri, J.C., Stanley, S.O., Terry, B., and Hall, R.V. (1985). Assessment of eco-behavioral interaction in school settings. Behavior Assessment, 7, 331-347.

Greenwood, C.R. Schulte, D., Kohler, F., Dinwiddie, G., & Carta, J. (1986). Assessment and analysis of ecobehavioral interaction in school settings. In R. Prinz (ed.), Advances in behavioral assessment of children and families, (Vol. 2, 69-98) Greenwich, CT: JAI Press.

Marx, R.W. & Walsh, J. (1988). Learning from academic tasks. The Elementary School Journal, 88, (3), 207-219.

Rosenshine, B. & Berliner, D. (1978). Academic engaged time. British Journal of Teacher Education, 4, 3-16.

SAS Institute Inc., (1985). SAS User's Guide: Basics, Version 5 Edition. Cary, NC: SAS Institute Inc.

Skiba, R.J. & Deno, S.L. (1989). Temporal parameters in the sampling of behavior: The accuracy and generalizability of observation. Manuscript submitted for publication.

Stallings, J. (1980). Allocated academic learning time. Educational Researcher, 9(10), 11-16.

Stanley, S.O. & Greenwood, C.R. (1981). CISSAR: Code for instructional structure and student academic response: Observer's Manual. Kansas City, KS: Juniper Gardens Children's Project, University of Kansas.

Trenholm, S., & Rose, T. (1981). The compliant communicator: Teacher perceptions of appropriate classroom behavior. The Western Journal of Speech Communication, 45, 13-26.

Footnote

[1]The brevity of the observation sessions (AR: 30 min, C: 20 min) is at the lower end of the range examined by Skiba and Deno (1989), and therefore, raises questions regarding the representativeness of these relatively brief data sets. The research design of the present study addresses this potential limitation in three ways. First, a relatively large number of subjects were observed, and data were analyzed for <u>groups</u> rather than for individuals. In contrast, the excellent analyses by Skiba and Deno (1989) identified the likelihood of error in estimating <u>individuals'</u> whole-day behavior rates from 30 min samples. Second, two independent samples of objects allowed analyses to be checked for replicability. The resulting large number of statistical tests was then controlled for inflation of alpha by Dunn's procedure. Third, Skiba and Deno (1989) and Greenwood et al. (1986) examined representativeness of brief sessions in comparison to entire day and four-day criteria, respectively. In either instance, the criterion data would necessarily include non-instructional periods, and activities other than those to which we wished to generalize. Our specific interest was in basic skills instruction, which occurs in first-grade classes mainly during the morning hours that were the focus of our observations.

Table 1
Descriptive Data for First Graders

	GROUP				
	At Risk n = 107 Sample		Control n = 97 Sample		Total N = 204
	Derivation (n = 67)	Replication (n = 40)	Derivation (n = 58)	Replication (n = 39)	
Gender					
F n	16	17	13	17	63
%	24%	42%	22%	44%	31%
M n	51	23	45	22	141
%	76%	58%	78%	56%	69%
Race					
B n	32	26	24	19	101
%	48%	65%	41%	49%	50%
W n	34	14	30	19	97
%	51%	35%	52%	49%	48%
O n	1	0	4	1	6
%	1%	0	7%	2%	2%
Mothers Education					
Median years	12	12	13	13	12
Verbal IQ[a]					
M	94	93	106	105	99
SD	11	14	14	14	14
Reading Achievement[b]					
M	82	82	102	104	92
SD	14	13	17	14	18
Math Achievement[b]					
M	84	85	94	95	89
SD	13	10	12	10	13
Work-Related Skills[c]					
M	3.5	3.5	5.4	5.1	4.4
SD	1.1	1.0	0.9	1.2	1.4
Interpersonal Skills[c]					
M	4.8	4.6	5.9	5.8	5.3
SD	1.1	1.1	0.7	0.9	1.1

a Cognitive Abilities Test
b Diagnostic Achievement Battery
c Cooper-Farran Behavioral Rating Scales

Table 2

CISSAR Behavioral Codes and Aggregates[1]

CISSAR Categories	Aggregates	Codes
Activity	Academic	Reading, Mathematics, Spelling, Handwriting, Language, Science, Social Studies, Arts/Crafts
	Non-academic	Free Time, Business Management, Transition, Cannot Tell
Task	Active	Reader, Workbook, Worksheet, Paper/Pencil
	Passive	Listening to Lecture, Other Media, Teacher/Student Discussion, Fetch/Put Away
Structure	Focused	Small Group, Individual with Teacher
	Diffuse	Entire Group, Individual with Peer, Individual Alone
Teacher	Engaged	Direct Teaching, Indirect Teaching, Approval, Disapproval
	Disengaged	Teaching Other Children, Approval of Other Children, Disapproval of Other Children, Other Talk, No Response

Table 2, continued

CISSAR Categories	Aggregates	Codes
Student	Academic	Writing, Reading Aloud, Reading Silent, Asking Question, Answering Question, Academic Talk, Academic Game Play
	Task Management	Attention, Raise Hand, Look for Materials, Move, Play Appropriately
	Inappropriate	Disrupt, Look Around, Inappropriate Locale, Inappropriate Task, Inappropriate Play, Talk not Academic, Self Stimulation

[1] Adapted with authors' permission.

Table 3

Percent Occurrence of Composite Ecological Arrangements

CEA	Activity	Task	Structure	Teacher	Risk Derivation (n=11160)	Risk Replication (n=7110)	Control Derivation (n=6240)	Control Replication (n=4680)
1	Academic	Active	Diffuse	Engaged	7.37	5.32	7.34	4.83
2	Academic	Active	Diffuse	Disengaged	6.04	8.78	5.74	9.02
3	Academic	Active	Focused	Engaged	13.16	12.52	13.09	8.08
4	Academic	Active	Focused	Disengaged	28.87	30.77	33.67	34.62
5	Academic	Passive	Diffuse	Engaged	13.32	6.68	11.06	7.39
6	Academic	Passive	Diffuse	Disengaged	2.48	7.50	2.95	9.40
7	Academic	Passive	Focused	Engaged	12.12	6.56	8.13	3.57
8	Academic	Passive	Focused	Disengaged	5.81	12.93	6.88	12.59
9	Nonacademic	Active	Diffuse	Engaged	0.13	0.07	0.00	0.04
10	Nonacademic	Active	Diffuse	Disengaged	0.56	0.44	0.29	0.21
11	Nonacademic	Active	Focused	Engaged	0.00	0.08	0.00	0.00
12	Nonacademic	Active	Focused	Disengaged	0.86	0.34	0.67	3.08
13	Nonacademic	Passive	Diffuse	Engaged	1.24	1.07	2.04	0.32
14	Nonacademic	Passive	Diffuse	Disengaged	4.46	4.75	4.60	3.14
15	Nonacademic	Passive	Focused	[illegible]ged	0.19	0.30	0.66	0.02
16	Nonacademic	Passive	Focused	Disengaged	2.55	1.48	2.04	3.31

Table 4

Percent Occurrence of Student Responses

	Group			
	At Risk		Control	
	Sample			
Student Response	Derivation	Replication	Derivation	Replication
Academic	23	24	30	28
Task Management	50	54	53	56
Inappropriate	26	22	17	18

Table 5

Conditional Probabilities of Student Academic Response

CEA	Risk Derivation (UCP=23)	Z	p	Risk Replication (UCP=24)	Z	p	Control Derivation (UCP=30)	Z	p	Control Replication (UCP=28)	Z	p
1	26.89	1.62	*	33.60	2.92	***	41.48	3.63	***	43.81	3.51	***
2	24.48	0.46		31.89	3.13	***	48.32	5.05	***	32.23	1.27	
3	19.95	-2.15	**	25.73	0.84		20.32	-4.11	***	19.31	-2.52	**
4	36.34	12.79	***	26.01	1.64	*	42.65	9.34	***	39.38	7.51	***
5	9.82	-8.26	***	6.32	-5.82	***	5.65	-9.39	***	6.36	-5.93	***
6	16.61	-1.79	**	16.14	-2.94	***	32.61	0.51		14.77	-4.32	***
7	11.46	-7.00	***	16.94	-2.24	**	13.21	-5.42	***	4.79	-4.29	***
8	31.17	3.11	***	34.60	5.18	***	41.49	3.28	***	27.33	-0.29	
9	20.00	-0.22		0.00	-0.90		.	.		0.00	-0.68	
10	57.14	4.04	***	0.00	-1.99	**	33.33	0.20		50.00	0.99	
11	.	.		83.33	2.50	**	.	.		.	.	
12	39.58	2.29	**	16.67	-0.51		71.43	3.43	***	30.56	0.38	
13	9.42	-2.82	***	18.42	-0.83		2.36	-4.40	***	0.00	-1.55	*
14	8.84	-4.97	***	7.69	-4.56	***	6.27	-5.48	***	17.69	-1.74	**
15	19.05	-0.35		14.29	-0.77		4.88	-2.35	**	0.00	-0.42	
16	15.09	-2.09	**	17.14	-1.04		27.56	-0.38		22.58	-0.95	

* $p < .10$ ** $p < .05$ *** $p < .003$

Table 6

Conditional Probabilities of Student Inappropriate Response

CEA	Risk Derivation (UCP=26)	Z	p	Risk Replication (UCP=22)	Z	p	Control Derivation (UCP=17)	Z	p	Control Replication (UCP=16)	Z	p
1	30.78	1.94	**	20.37	-0.51		14.19	-1.23		11.06	-1.39	*
2	40.95	5.72	***	19.07	-1.19		24.58	2.64	**	14.69	-0.45	
3	21.58	-2.89	***	14.83	-3.58	***	13.46	-2.05	**	14.29	-0.57	
4	34.51	7.70	***	27.93	4.98	***	24.23	6.82	***	19.01	2.86	***
5	23.13	-1.89	**	32.21	3.46	***	12.90	-2.15	**	8.96	-2.45	**
6	21.30	-1.29	*	29.08	2.73	**	15.22	-0.50		10.91	-2.06	**
7	18.55	-4.41	***	12.76	-3.02	***	11.44	-2.41	**	4.79	-2.67	**
8	21.45	-1.87	**	16.00	-3.02	***	10.96	-2.34	**	22.92	3.34	***
9	33.33	0.44		0.00	- [illegible]5		.	.		0.00	-0.49	
10	14.29	-1.38	*	9.68	-1.05		16.67	-0.04		20.00	0.25	
11	.	.		0.00	-0.95		.	.		.	.	
12	8.33	-2.44	**	12.50	-0.68		16.67	-0.06		32.64	3.57	***
13	23.19	-0.62		14.47	-1.15		3.94	-2.73	**	0.00	-1.15	
14	19.48	-2.26	**	23.37	0.40		10.10	-2.14	**	6 [illegible]0	-1.99	**
15	19.05	-0.56		4.76	-1.41	*	21.95	0.57		0.00	-0.31	
16	7.72	-4.52	***	7.62	-2.24	**	12.60	-0.91		2.58	-3.00	***

* p < .10 ** p < .05 *** p < .003

Study Three: Dynamic Assessment and Individual Differences

Abstract

This study examined the role of dynamic assessment measures in models of individual differences predicting (a) ability to profit from dynamic assessment training and (b) academic achievement. The sample consisted of 193 first grade children representing a wide range of abilities. The first regression model was based on the perspective that dynamic assessment tasks (DAT) represent a type of strategy instruction with the purpose of identifying which of several child characteristic measures contributed to successful posttest performance following training. Verbal intelligence, pretest knowledge, language variables, and number of prompts needed during training accounted for 48% of DAT posttest variance with prompts accounting for a significant amount of variance beyond all other variables in the model. Limited support for DAT prompts was found in the prediction of academic achievement. Secondary analysis of DAT learning profiles (Brown & Ferrara, 1985) revealed that, although indistinguishable by standard achievement measures and DAT prompts, two subgroups of children could be discriminated by DAT posttest, a measure of skill acquisition during training. The findings were interpreted as providing limited but encouraging support for future research with dynamic measures as indices of educationally relevant individual differences.

Dynamic Assessment, Individual Differences, and Achievement

Dynamic assessment represents a training paradigm in which the examiner takes an active role in teaching a purposefully difficult task to an examinee and then measures the degree to which this training resulted in learning, usually controlling for initial knowledge of the task. The theoretical roots of dynamic assessment lie with Vygotsky's conception of cognitive development as a product of social interaction. Dynamic approaches are contrasted with static measures of intelligence and achievement in which the examiner is passive and the child's task is to respond without assistance. Whereas the latter approach provides information on the child's current unassisted performance level, the former provides an indication of the level of performance the child may attain with help from a more knowledgeable teacher. The child's potential is termed the zone of proximal development, the size of which varies according to the amount of instruction the child requires and/or the amount of learning apparent following training. Psychometric intelligence is thought to mark the lower bounds of the zone whereas dynamic measures estimate the upper limit.

The appeal of dynamic assessment for psychologists and educators lies in its potentially enhancing role in the analysis of cognitive functioning and that it may provide a more instructionally relevant indicator of individual differences in learning potential than IQ tests. A low IQ score indicates, at best, that the child will require more instruction to learn a task than will a child with higher intelligence, whereas dynamic assessment may indicate how much instruction is required to obtain a specified result. Theoretically, IQ and learning potential (dynamically assessed) are independent sources of variation in learning (Wertsch & Rogoff, 1984).

While dynamic assessment may be viewed as a potentially valuable measurement approach in its own right, it may also be placed in the larger literature on strategy training. Recent theoretical discussions of communicational dynamics and strategy training instruction (Stone, 1985; 1986; Turnure, 1985; 1987) and summaries of empirical work using techniques to produce strategic learners (Goldman, 1989; Pressley, Johnson & Symons, 1987) contain a common element. Instruction characterized by active participation by both learner and teacher and is crafted in such a way that the learner induces the strategy, may have the desireable effect of improving learning. It is important to note that in this conceptualization of instruction, the learner is not told directly what strategies to perform (Brown & Ferrara, 1985) but rather encouraged to induce a specific strategic approach through the guidance of the teacher. As noted by Goldman (1989), this type of strategy instruction differs from direct instruction and self instruction methods in that the former proceeds from implicit to explicit assistance with the amount of instruction dependent on the child's initial representation of the task.

Stone, a Vygotskian scholar, identifies this type of instruction as "proleptic" and defines prolepsis as "a means of communication in which the initial statement of a message presupposes certain yet unstated components" (Stone, 1986, p. 7). We propose that the graduated prompt method of dynamic assessment developed by Brown and her colleagues (Bryant, 1982; Ferrara, Brown & Campione, 1986) and used in this investigation meets Stone's definition. Stone's (1985, 1986) thesis is that attention to proleptic communication factors may assist our understanding of the mechanisms underlying successful strategy training.

The common ground between proleptic instruction and graduated prompts is that the teacher begins with global hints on how to solve the problem and adds more specific clues if the child is unsuccessful. The sequence of hints initially provides only meager information leaving several unstated components that are systematically added in later prompts. In proposing this type of instruction as a method of understanding strategy training, Stone (1985, 1986) noted there is limited empirical evidence regarding skills needed to profit from strategy instruction, and, on a related point, who may best benefit from any particular strategy training regimen. He proposed oral language skills as a possible factor specifically related to proleptic instruction. Our purpose in viewing dynamic assessment as a type of strategy instruction was to study the contribution of individual difference measures to performance on a dynamic assessment posttest, thereby investigating skills needed to profit from this type of training. The model included a traditional measure of verbal intelligence, pretest knowledge of the target task, oral language skill, and the number of prompts required during training.

A second purpose of this research was to assess the contribution of dynamic measures to predictions of academic achievement. This approach is consistent with the thrust of recent investigations with dynamic measures and represents an extension of this line of inquiry. For example, Bryant (1982) demonstrated that the amount of training required to meet criterion during a dynamic session was not redundant with either prior knowledge of the task or intelligence in preschool children. Ferrara, Brown & Campione (1986) reported reasonable consistency between two types of dynamic measures and further illustrated how the amount of assisted training expands understanding of psychometric intelligence. From a more applied perspective, Delclos, Burns & Kulewicz (1987) found that teachers increased their estimates of handicapped children's learning abilities after viewing the children's performance during a dynamic assessment task.

Linking dynamic measures with achievement is an important question with respect to the validity of the task. However, it was with some ambivalence that we undertook this problem. A strong connection with academic achievement would seem possible only if children were systematically exposed in their classrooms to the type of guided instruction used in the task (e.g., Palinscar & Brown, 1984). A wealth of classroom observational studies suggest that multiple opportunities to solve a problem are generally not provided to individual children especially those with lower academic ability. (e.g., Cooper & Speece, 1989; Eder, 1981; Tobin, 1987). Nonetheless, while the theoretical impetus behind dynamic assessment is compelling, additional evidence of linkages to an educational context is needed. To provide the most rigorous test of the association between dynamic measures and achievement, psychometric intelligence, pretest knowledge of the target task, and language skills were assessed in the present study and entered in a regression model prior to the dynamic measures.

In summary, the nature of dynamic assessment was explored from two perspectives. The first viewed the dynamic assessment sessions as a specific type of strategy instruction leading to questions of who may benefit from this approach. The second perspective was to extend the more prevalent view of dynamic assessment as an individual difference measure by determining its relationship with academic achievement.

An additional element of this research was that first grade children were participants. Previous work has focused on preschool and upper elementary children (Bryant, 1982; Ferrara et al., 1986). Thus the performance of children beginning their school careers is unknown.

Method

<u>Subjects</u>

The children who participated in this investigation were members of a larger study of children's characteristics and learning environments of first grade children who were considered by their teachers to be at risk for school failure (AR). Also included were randomly selected classmates who were achieving at an average rate (Control, C). In total, 193 children, (104 AR, 83 C) had complete child characteristic data used for the present analyses. Although the focus of this study was not AR and C differences, Table 1 reports subject characteristics on all variables (described below) by group and full sample for completeness as secondary analyses used the group distinction to explore relationships between dynamic assessment measures and membership in the at-risk group. Details of subject selection criteria can be found in Speece and Cooper (1988). Briefly, children were considered to be at risk for school failure if they were referred by their teachers for an academic or behavioral reason to a school-based prereferral committee (Teacher Assistance Teams, TAT). These children had not repeated first grade nor had they received special education services. Normally achieving control children matched by gender and classroom were randomly selected from the teachers' lists of such children when an at-risk child meeting all criteria was identified.

[INSERT TABLE 1 ABOUT HERE]

Because group differences were apparent in Table 1, it was necessary to determine the homogeneity of group variances before conducting analyses on the full sample. With the exception of one language variable, (Level 1) all tests were nonsignificant (Cochran's F max test, Kirk, 1968).

Measures

Dynamic Assessment Task (DAT). As previous investigations of dynamic measures focused on children younger or older than our sample, it was necessary to develop an instrument of appropriate difficulty for six year olds. The DAT was an adapted version of the instrument developed by Bryant (1982) for preschool children, the general form of which was used in several subsequent studies (e.g., Ferrara, Brown, & Campione, 1986).

The DAT was composed of 12 3x3 pictorial matrix problems (2 practice, 4 training, 6 posttest items) with practice and training matrices accompanied by eight standard prompts graduated in explicitness to be administered by the examiner until the child solved the puzzle. The posttest was a non-assisted measure of maintenance. All matrices required the understanding of the figural transformational rule used in the first two rows in order to solve for the missing figure in the third row. Each matrix had eight answer pieces from which the child could select to solve the puzzle. Within each set of eight possible solutions, two were designed to be close approximations of the correct answer but one critical component was missing from each (e.g., color, orientation). Examples of a training puzzle and the corresponding prompts can be found in the Appendix. Across the four training items, the amount of new information conveyed by each succeeding prompt was identical. The only variations were specific to the dimensions of each training matrix which were substituted in each prompt sequence as appropriate. Each training matrix varied on two dimension (e.g., color, orientation, shape). The prompt sequence was designed to teach one dimension initially in conjunction with the figural transformation rule (prompts 3,4,5) and then to focus the learner's attention on the second dimension.

Each of the training matrices required understanding of the superimposition rule to find the correct solution. In these matrices, the third picture in each row was the result of superimposing the first two pictures. The decision to teach a single rule (superimposition) was based on our purpose of developing a measure to provide an indication of a child's modificability through instruction in the general sense. Similar to Embretson (1987) the focus was on immediate performance level as opposed to effecting an enduring change in ability.

The matrices were selected from three sources: the Detroit Test of Learning Aptitude-2 (1 matrix, Hammill, 1985), the Matrix Analogies Test-Short Form (6 matrices, Naglieri, 1985) and the Raven Progressive Matrices Sets B and E (5 matrices, Raven, 1960). Pilot testing indicated that the superimposition transformation rule was of sufficient difficulty that most first graders would not be able to solve the puzzle without assistance but could acquire the solution with varying amounts of teaching. This criterion was also met with the study sample as only 2 children solved each of the four training puzzles on the first prompt.

The two practice items (not scored) were designed to acquaint the child with the nature of the task, the puzzle format, and interaction with the examiner to solve the problems. The practice matrices did not follow the superimposition rule and were selected, based on pilot testing, for their relative ease. All four training items and four of six posttest items were superimposition puzzles. The two additional posttest items were similar to the practice items. The order of training and posttest items was randomly determined and then administered in this predetermined sequence. The DAT was administered individually in a single session and took approximately 20 to 25 minutes. As suggested by Peterson, Homer and Wonderlich (1982), procedural reliability data were collected to determine the degree to which examiners were adhering to the scripts written for the DAT (instructions and prompts). Twenty-seven sessions (14%) were recorded on audiotape and scored by an independent rater. Based on the percentage agreement method the procedural reliability fo all protocols was 100%.

Measures derived from DAT were total number of prompts (Prompts) across four training items (range: 4 to 32) and derived posttest score.

During the posttest the child was given a second opportunity to respond if the first choice was incorrect. This procedure was instituted based on our experience with first grade pilot subjects who appeared to respond impulsively to the posttest. A 10s study period was used prior to presenting the eight possible solutions for the same reason. To account for two possible responses and the fact that two of the choices were close approximations to the correct answer, a partial credit scoring system was devised. Full credit (1.25 pts.) was awarded for a correct answer on the first response with a .75 credit for a correct answer on the second attempt. A score of .25 was earned if the two distractor items that were close approximations were selected on either attempt. For example, if child selected a close approximation on the first response and the correct answer on the second, a score of 1.0 would be earned. Thus, possible posttest scores ranged from 0 to 7.5. The internal consistency of the posttest (coefficient alpha) was .60, indicating a modest but acceptable level of consistency for research purposes.

Cognitive Abilities Test (CogAt). The CogAt Primary Battery, Level 1, Form 4 (Thorndike & Hagen, 1982) provided a measure of verbal intelligence (VIQ) and of pretraining skill on matrix type items (Pretest). The latter measure was a subtest of the nonverbal section of the CogAt and consisted of 22 matrix analogy problems. The raw score on the CogAt matrix subtest was used as the pretest measure and correlated .58 with the

DAT posttest. The psychometric properties of the CogAt have been evaluated favorably (Salvia & Ysseldyke, 1985).

Preschool Language Assessment Instrument (PLAI). The PLAI (Blank, Rose & Berlin, 1978) was specifically designed to assess children's ability to process the types of oral language questions found in classrooms, and at four levels of cognitive complexity. The four levels were matching perception (imitation, labeling, "What is this called?"), selective analysis of perception (focus on specific aspects of material and/or integration of components, "Find something that can cut"), reordering perception (restructuring oral or visual information according to demands of the question, "Find the things that are not dolls"), and reasoning about perception (going beyond the immediate oral or visual information by noting logical relationships, "What will happen to the man if he closes the umbrella?").Scores on each of the four levels were used as the language variables in this study. Blank, Berlin, and Rose (1983) asserted that "If children experience difficulty on this test, then it implies that they will experience difficulty in understanding the verbal exchange that is at the heart of the classroom experience" (p. 299). Blank et al. (1978) reported that across the four levels of questions split-half reliability ranged between .64 and .86 with test-retest reliability ranging from .73 to .88. Our assessment of interrater reliability (three individual scorers' agreement with scores on 5 standard protocols) was calculated for each of the 4 levels of cognitive complexity and yielded mean correlations of .97, .99, .99 and .89, respectively. For the regression analyses, the first two levels of discourse understanding (matching perception and analysis of perception) were entered in a single step followed by the second two levels (reordering perception and reasoning about perception).

Diagnostic Achievement Battery (DAB). The DAB (Newcomer & Curtis, 1984) was used for measures of reading and math achievement. The reading score was a composite of alphabet/word recognition and reading comprehension subtests, and math achievement was composed of math reasoning and math calculation subtests. The instrument has been shown to have acceptable psychometric properties (Brown & Bryant, 1984).

Procedures and Analysis

All measures were individually administered by graduate students in education or psychology in three sessions lasting approximately 45 minutes each. Order of testing was randomly determined with the restriction that the CogAt was administered prior to the DAT so that the CogAt matrix subtest could serve as a pretest of the matrix skills taught and assessed on the DAT. The PLAI and DAT were administered in one session. Tests were administered in the schools and prior to the children's lunch periods.

Two sets of multiple regression analyses with forced order of entry were used to address the two primary research questions. First, verbal intelligence, pretest knowledge, language variables, and DAT total prompts were entered in a stepwise manner to predict DAT posttest performance. This analysis was intended to identify sources of individual differences in children's responses to the training. Second, DAT posttest scores were added as the final step in the above model to predict reading and math achievement. This analysis aimed to validate DAT measures as a predictor of achievement. Table 2 presents the correlation matrix for all variables.

[INSERT TABLE 2 ABOUT HERE]

Results

Predicting DAT Posttest

The results of the multiple regression analysis predicting DAT posttest performance are summarized in Table 3. Verbal intelligence and pretest matrix performance accounted for 37% of posttest performance, the language variables accounted for approximately 4%, and DAT total prompts, entered last, accounted for 7%. With the exception of PLAI levels 3 and 4 scores, each step in the analysis accounted for a significant amount of variance. Thus, total prompts contributed variance above and beyond all other variables in the model. Partial correlation coefficients greater than .10 were as follows: Pretest .37, PLAI level 2 .22, and Prompts -.34. The partial correlation for VIQ was .07 indicating that the pretest and number of prompts accounted for the most unique variance in the posttest scores.

[INSERT TABLE 3 ABOUT HERE]

Predicting Achievement

To assess the relationship between DAT measures and academic achievement, DAT posttest scores were added as the last variable in the above model, the results of which are summarized in Table 4. VIQ and Pretest accounted for 25% of the variance in reading achievement with all other variables being nonsignificant. For math achievement, the significant predictors were VIQ (28%), Pretest (3%), and Prompts (2%). Regarding partial correlations with achievement, only VIQ obtained coefficients of any magnitude (.27 and .32 for reading and math, respectively).

[INSERT TABLE 4 ABOUT HERE]

Secondary Analyses

Given the modest but significant contribution of Prompts to math achievement, we were interested in exploring this relationship more specifically. To this end, four groups were created based on quartile prompt scores from the sample (group 1: $\leq$ 8 prompts; group 2: 9-12 prompts; group 3: 13-16 prompts; group 4: $\geq$ 17 prompts). Group 1, receiving 8 or fewer prompts across the four training items, solved the puzzles on the first or second prompt thereby requiring little assistance from the examiner (see prompt sequence in the Appendix). At the other extreme, group 4, requiring a total of 17 or more prompts, needed the more explicit directives contained in prompts 5 through 8.

Separate one-way ANCOVAs were calculated for dependent measures reading and math covarying pretest performance. VIQ could not be used as a covariate due to significant interactions with the group variable (p < .001 and .02 for reading and math, respectively). For both analyses, the main effect of prompt group was significant (reading F (3,188) = 6.76, p < .0002; math F (3,188) = 6.08, p < .0006). Of greater interest were the post hoc analyses. Tukey's HSD test showed prompt group 1 (requiring 8 or fewer prompts) had higher reading achievement (p < .05) than the other three groups, none of which differed from each other (1 > 2 = 3 = 4). Similar analysis for math showed that group 1 math achievement was greater than either group 3 or 4 but did not differ from group 2 (1 = 2, 1 > 3 = 4). Apparently, those children who needed minimal instruction in the DAT session also distinguished themselves by greater reading and mathematics achievement. Interpretation of these data must be tempered by the fact that the effects of verbal intelligence could not be controlled.

Also explored were four learning profiles created on the basis of performance on both DAT training and posttest items (Brown & Ferrara, 1985). A fast/fast profile is defined by performance in which the number of prompts is at or below the sample median (12) with posttest performance above the sample median (4.5). A slow learning profile is defined by performance in the opposite direction (slow/slow). Reflective profiles (slow/fast) are defined by children who require a greater number of prompts but who perform well (i.e., above the median) on the posttest. Context bound learners (fast/slow) require few prompts during training but perform at or below the posttest median. In the present study, 36.3% (28 AR, 42 C) of the sample were considered fast learners and 37.3% (49 AR, 23 C) exhibited profiles associated with the slow learners. In both cases, posttest performance was consistent with training performance. Reflective learners comprised 11.4% (10 AR, 12 C) of the sample, context-bound learners 15.0% (17 AR, 12 C).

The assumption associated with reflective learners is that they may benefit from the type of guided instruction offered during DAT training but that this learning skill may not be realized in the classroom in the absence of such guided instruction. Thus, the reflective learner would appear, on achievement tests, to be unresponsive to instruction. To examine this assumption we compared the reflective and slow profile groups on reading and math achievement controlling for intelligence and pretest scores. There were no achievement differences between these two groups tentatively supporting the view that some individual differences in learning potential are not apparent in achievement scores. Interestingly, the reflective (slow/fast) and context-bound (fast/slow) groups did not differ on any of the measures in the analysis of learning profiles (intelligence, pretest, achievement) even though they differed on both prompts and posttest performance.

Discussion

Posttest scores on the dynamic measure were predicted, as expected, by verbal intelligence and a nonidentical measure of pretest skill. Surprisingly, language skill produced only a modest relationship with the posttest which was due to the PLAI level 2 variable. This latter relationship may be explained by similarities in task demands as the DAT materials required visual superimposition and the PLAI level 2 questions required selective analysis of visually presented materials. It appears that higher levels of discourse processing represented by PLAI levels 3 and 4 were not demanded by the verbal prompts during training or were redundant with the previously entered measure of intelligence (see Table 2).

The addition of number of prompts to the equation produced a modest but significant increase in posttest variance after all other variables were entered. Also of interest is that prompts accounted for approximately 10% of the unique posttest variance whereas verbal intelligence accounted for less than 1%. While these results are limited to a very specific domain, some evidence is accrued in support of proleptic instruction, as embodied by dynamic training, in explaining children's learning competence. The role of prompts in this study is generally supportive of Bryant's (1982) findings in which a younger group of children was studied and different administration procedures were used. Stone's (1986) contention that oral language skills may be important received partial support. Although language skill did not emerge as a general predictor in the present study, the apparent concordance between the language used in training and that used in the PLAI level 2 questions suggests that analysis of language demands in strategy instruction regimens may be worthy of further study.

Verbal intelligence retained its customary role in the prediction of academic achievement while the dynamic measures added little to the equation. Further analysis of children's achievement in relation to prompts revealed that only those children who required on average less than two or three prompts per training item received higher achievement scores than the groups who required more prompts. This finding raises the issue of processes involved in "good" performance in the face of incomplete instruction. Campione and Brown (1984) aptly describe this situation: "Good learners perform thought experiments, seek appropriate analogies and understand some of the principles involved in learning and reasoning from incomplete knowledge ... Instruction may well be incomplete, but they have the skills to 'complete' it for themselves" (p. 286). From this we would hypothesize that our group of fast learners/high achievers are performing "thought experiments" during the initially incomplete DAT instruction just as they are believed to do in classrooms.

While the present study cannot address the skills invoked by high achievers some pertinent evidence was obtained for the slow learners defined by the amount of instruction needed during training. Of the sample, 11.4% would be considered, in Brown and Ferrara's (1985) terms, reflective learners as they required a greater number of prompts but correctly answered more posttest puzzles than the sample median. While the definition of this learning profile is relative to the sample studied, it is noteworthy that our percentage of reflective learners is quite similar to Ferrara's (1982) 15% derived from an older group of elementary school children. The point is that a number of children were able to profit from the type of instruction offered even though they did not demonstrate higher achievement than slow learning children who did not perform well on the posttest. It would be important to test the hypothesis that these reflective learners would benefit from a proleptic manipulation of instruction in an academically relevant domain. Goldman (1989) suggested that this form of instruction may be generally beneficial in producing strategic learners in mathematics as it is designed to be responsive to the child's current understanding and may foster internalization of strategies that are taught.

Given that approximately half of the reflective learners were considered by their teachers as possible candidates for special education, one might speculate on the role "instructional deficiencies" may play in such referrals (Turnure, Buium, & Thurlow, 1976). While referral is a complex and multiply determined phenomenon (Speece & Cooper, 1988) and the reflective at-risk group comprises only 9.6% of the risk sample, it may be that greater attention to the nature of instruction afforded this group could result in heightened school success. The role of dynamic assessment in identifying candidates for instructional modification rather than special education refe al may prove to be its major contribution to educational practice.

References

Blank, M., Berlin, L.J., & Rose, S.A. (1983). Classroom dialogue: Abilities and disabilities of learning disabled children. In J.D. McKinney & L. Feagans (Eds.) Current topics in learning disabilities (pp. 285-307). Norwood, NJ: Ablex.

Blank, M., Rose, S.A., & Berlin, L.J. (1978). Preschool language assessment instrument: The language of learning in practice. Orlando, FL: Grune & Stratton, Inc.

Brown, L. & Bryant, B. (1984). Critical reviews of three individually administered achievement tests. RASE, 5(5), 53-60.

Brown, A.L., & Ferrara, R.A. (1985). Diagnosing zones of proximal development. In J.V. Wertsch (Ed.), Culture, communication, and cognition: Vygotskian perspectives (pp. 273-305). New York: Cambridge University Press.

Bryant, N.R. (1982). Preschool children's learning and transfer of matrices problems: A study of proximal development. Unpublished Master's Thesis, University of Illinois, Urbana-Champaign.

Campione, J.C., & Brown, A.L. (1984). Learning ability and transfer propensity as sources of individual differences in intelligence. In P.H. Brooks, R.D. Sperber, & C. McCauley (Eds.), Learning and cognition in the mentally retarded (pp. 265-294). Baltimore: University Park Press.

Cooper, D.H. & Speece, D.L. (1989). Instructional correlates of student's academic responses: Comparisons between at risk and control students. Manuscript in review.

Delclos, V.R., Burns, M.S., Kulewicz, S.J. (1987). Effects of dynamic assessment on teachers' expectation of handicapped children. American Educational Research Journal, 24, 325-336.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher student interaction. Sociology of Education, 54, 151-162.

Embretson, S.E. (1987). Improving the measurement of spatial aptitude by dynamic testing. Intelligence, 11, 333-358.

Ferrara, R.A. (1983). Children's learning and transfer of inductive reasoning rules: A study of proximal development. Unpublished Master's Thesis, University of Illinois, Urbana-Champaign.

Ferrara, R.A. Brown, A.L., & Campione, J.C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. Child Development, 57, 1087-1099.

Goldman, S.R. (1989). Strategy instruction in mathematics. Learning Disability Quarterly, 12, 43-55.

Hammill, D.D. (1985). Detroit test of learning aptitude - 2. Austin, Tx: Pro-Ed.

Kirk, R.E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Naglieri, J.A. (1985). Matrix analogies test. Columbus: Charles E. Merrill.

Newcomer, P.L., & Curtis, D. (1984). Diagnostic Achievement Battery. Austin, TX: PRO-ED.

Palinscar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. Cognition and Instruction, 1(2), 117-175.

Peterson, L., Homer, A.L., & Wonderlich, S.A. (1982). The integrity of independent variables in behavioral analysis. Journal of Applied Behavioral Analysis, 15, 477-492.

Pressley, M., Johnson, C.J., & Symons, S. (1987). Elaborating to learn and learning to elaborate. Journal of Learning Disabilities, 20, 76-91.

Raven, J.C. (1960). Standard progressive matrices, sets A,B,C,D, and E. Cambridge: Lewis and Co.

Salvia, J., & Ysseldyke, J.E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton-Mifflin.

Speece, D.L., & Cooper, D.H. (1988). Ontogeny of school failure: Classification of first grade children. Manuscript in review.

Stone, C.A. (1985). Vygotsky's developmental model and the concept of proleptic instruction: Some implications for theory and practice in the field of learning disabilities. Research Communications in Psychology, Psychiatry and Behavior, 10, 129-152.

Stone, C.A. (1986). Improving the effectiveness of strategy training for the learning disabled: The role of communicational dynamics. Paper presented at the annual meetings of the International Academy for Research on Learning Disabilities, Evanston, IL.

Thorndike, R.L. & Hagen, E.P. (1982). Cognitive abilities test, primary battery edition. Chicago: Riverside.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. Review of Educational Research, 57, 69-95.

Turnure, J.E. (1985). Communication and cues in the functional cognition of the mentally retarded. In N. Ellis (Ed.), International Review of Research in Mental Retardation, Vol. 13, (pp. 43-77). New York: Academic Press.

Turnure, J.E. (1987). Social influences on cognitive strategies and cognitive development: The role of communication and instruction. Intelligence, 11, 77-87.

Turnure, J., Buium, N., & Thurlow, M. (1976). The effectiveness of interrogatives for promoting verbal elaboration productivity in young children. Child Development, 47, 851-855.

Wertsch, J.V., & Rogoff, B. (1984). Editors' notes. In B. Rogoff & J.V. Wertsch (Eds.), Children's learning in the zone of proximal development (pp. 1-6). San Francisco: Jossey-Bass.

Appendix

Sample Prompt Sequence for Superimposition Rule

When no other directions are noted, point to each picture individually.

1. "Look at each one of the pictures in the puzzle. (Point to each picture, pausing at the end of each row.) Could that help you figure out how to finish the puzzle?"

2. "Let's look at just the top row. (Cover remainder of the puzzle.) These pictures are all different. This one is different from this one, and this one is different from this one. Can you tell me how they are different? (Allow child to answer, correct if necessary). The pictures in the top row follow the Put Together rule. When the first picture is put together with the second picture it makes the third picture."

"Could that help you finish the puzzle?" (Uncover remainder of the puzzle.)

3. "We know that these are different because the first picture has a line on the right side. The second picture has a line across. The third picture has lines on the right side and across. When the first picture is put together with the second picture, it makes the third picture."

(Point to middle row.) "Can you find a line on the right? A line across? A picture with the line on the right put together with a line across? When the first picture is put together with the second picture it makes this picture." (Allow child to point to each picture.)

"Okay. In the top row if we have a line on the right put together with a line across to make this picture and in the middle row we have a line on the right put together with a line across to make this picture, what do we need to finish the puzzle?"

4. "Look at the top row again. (Point.) The pictures in the top row follow the put together rule. The pictures in the middle row follow the put together rule".

"In the bottom row we have a line on the right to put together with a line across".

"What piece do we need to go here?"

5. "Okay, in the top row we have a line on the right side, put together with a line across to make this picture."

"Here in the middle row we have a line on the right put together with a line across to make this picture."

"In the bottom row we have a line on the right side to put together with a line across. We need a picture with a line on the right side put together with a line across to go here. Can you find a picture with a line on the right side and a line across?"

"Any more?" (Allow child to find all 3 answers; help if necessary.) "Which one of these do you need to finish the puzzle?"

6. "These three pieces are different. Can you tell me how they are different? (Pointing to the three possible pieces, verify that the child has

the correct dimension.) Do we need a line across the middle, a line across the bottom or a line across the top to finish the puzzle?"

7. (Point to each picture and to the space for the missing piece.) "In the top row the line across is at the top."

"In the middle row, the line across is in the middle."

"In the bottom row, the line across is at the bottom."

"So which picture with a line on the right side put together with a line across do we need to go here?"

8. "In the top row the line across is at the top."

"In the middle row the line across is in the middle."

"In the bottom row the line across is at the bottom so we need a picture with the line across the bottom to finish the puzzle. Can you find a picture with a line on the right put together with a line at the bottom and finish the puzzle? Good!"

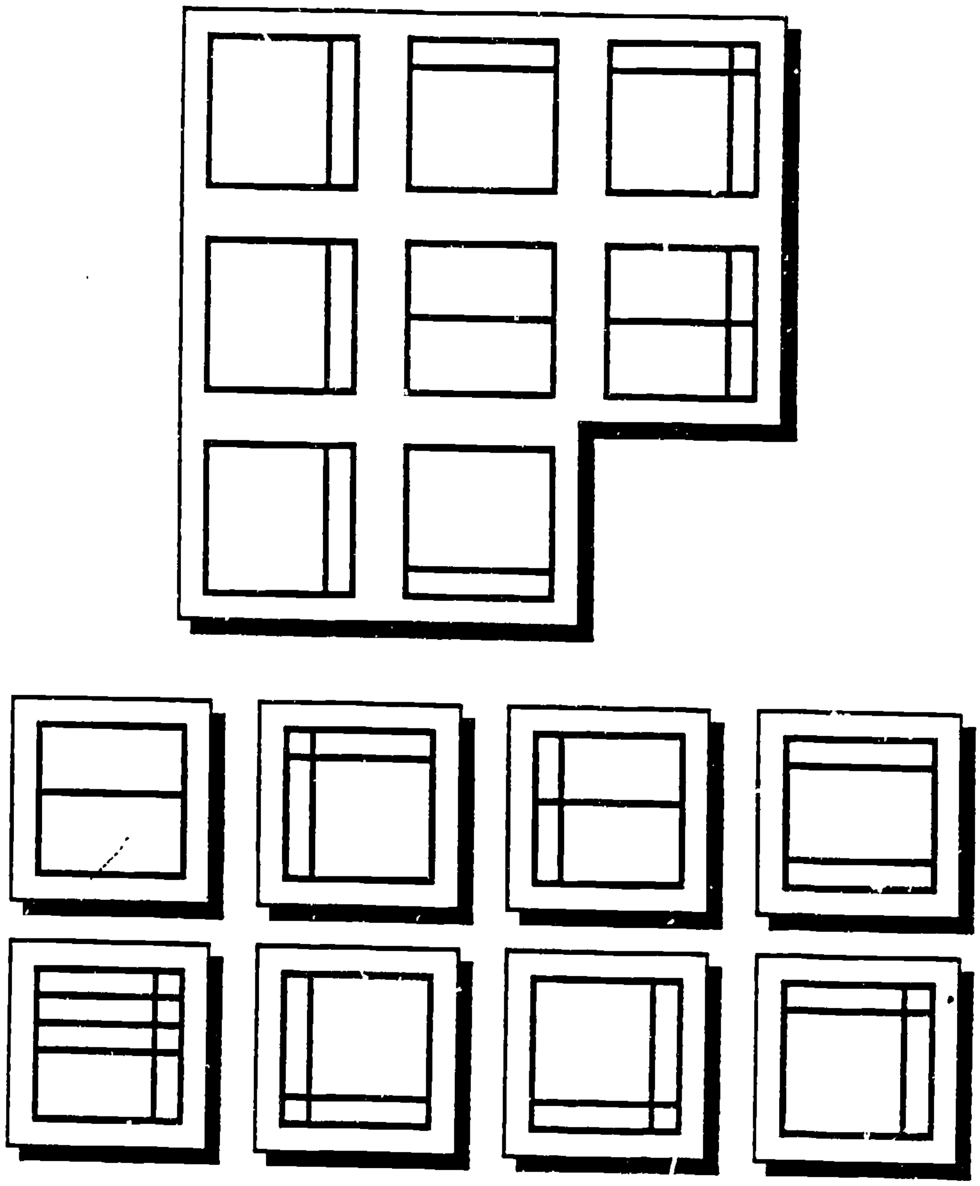


Table 1

Child Characteristic Data

Variable		Group AR	Group C	Sample
n		104	89	193
Gender				
F		30.8%	31.5%	31.1%
M		69.2%	68.5%	68.9%
Race				
Black		53.8%	41.6%	48.2%
White		45.2%	53.9%	49.2%
Other		1.0%	4.5%	2.6%
Mother's Ed.	(Mdn. Yrs)	12.1	13.1	12.3
Verbal IQ[a]	M	93.77	105.48	99.17
	SD	12.23	13.83	14.22
Pretest[a]	M	14.49	16.51	15.42
	SD	3.54	2.86	3.39
Reading[b]	M	81.31	102.53	91.09
	SD	13.97	15.74	18.19
Math[b]	M	84.22	94.51	88.96
	SD	11.79	11.72	12.81
PLAI[c]				
Level 1	M	2.69	2.78	2.73
	SD	.24	.17	.21
Level 2	M	2.44	2.52	2.48
	SD	.28	.31	.30
Level 3	M	2.19	2.36	2.27
	SD	.40	.41	.41
Level 4	M	1.88	2.18	2.02
	SD	.53	.43	.51
DAT[d]				
Prompts	M	13.28	11.00	12.21
	SD	4.92	5.08	5.11
Posttest	M	3.83	4.91	4.33
	SD	1.62	1.65	1.72

[a] Cognitive Abilities Test
[b] Diagnostic Achievement Battery
[c] Preschool Language Assessment Instrument
[d] Dynamic Assessment Tasks

Table 2

Correlation Matrix of Variables Used in Multiple Regression Analysis

Variables	2	3	4	5	6	7	8	9	10
1. VIQ	.52	.19	.41	.35	.55	-.45	.47	.49	.53
2. Pretest		.27	.29	.31	.45	-.44	.58	.36	.44
3. PLAI 1			.26	.29	.25	-.13	.16	.17	.13
4. PLAI 2				.54	.55	-.27	.40	.27	.27
5. PLAI 3					.50	-.28	.27	.30	.18
6. PLAI 4						-.44	.45	.39	.37
7. Prompts							-.55	-.37	-.40
8. Posttest								.37	.44
9. Reading									.64
10. Math									

Table 3

Prediction of DAT Posttest Performance

Variable	R	R^2	R^2 Change	F	df
VIQ	.472	.223	.223	54.70**	1,191
Pretest	.611	.373	.151	45.72**	1,190
PLAI 1,2	.639	.408	.035	5.56*	2,188
PLAI 3,4	.646	.417	.008	1.33	2,186
Prompts	.697	.485	.069	24.69**	1,185

* $p < .005$

** $p < .00001$

Table 4

Prediction of Reading and Math Achievement

Variable	df	Reading R	Reading R^2 Change	Reading F	Math R	Math R^2 Change	Math F
VIQ	1,191	.487	.238	59.53**	.532	.283	75.54**
Pretest	1,190	.504	.017	4.24*	.556	.026	7.17**
PLAI 1,2	2,188	.509	.004	.63	.558	.002	.26
PLAI 3,4	2,186	.525	.016	2.04	.564	.007	.96
Prompts	1,185	.538	.014	3.60	.580	.018	5.05*
Posttest	1,184	.542	.005	1.19	.590	.011	3.15

** $p < .01$

* $p < .05$

Study Four: Classification of First Grade Children

Abstract

The ontogeny of school failure was addressed by examining the characteristics of 112 first grade children, 63 of whom were considered at risk for school failure. Measures of achievement, intelligence, behavior, language, and learning potential were used with cluster analysis techniques to identify six empirically-derived clusters that were replicated across two samples. Three profiles represented variations of normal performance and three represented atypical patterns suggestive of learning disabilities, mild mental retardation, and language problems. These interpretations were validated by cluster differences on the degree of risk for referral, observed classroom behavior, and achievement variables not used in the cluster analysis. The variability of skills evident across the clusters in addition to the composition of the clusters, suggested that teachers cope with a high degree of heterogeneity in their classrooms, that their referrals are not based solely on one dimension of children's psychometric performance, and that certain multivariate profiles are associated with higher risk of referral. By capturing this variability it may be possible to design interventions that ameliorate the risk of failure for some groups of children.

Children's early failure in school is a complex developmental process. Stimulation in the home, cognitive development, social skills, and the quality of their instruction are some of the more obvious factors. Although the reasons for early school failure are complex, the range of educational outcomes for these children is relatively straightforward: continuing on the path of grade promotion, repeating a grade, or receiving services from special or compensatory education programs. Despite the multidimensional nature of school failure, few multivariate data exist to describe, let alone predict, who these children are and how their characteristics may influence placement decisions (Reschly, 1984).

Examination of the research traditions of special and general education provides some insight as to why data are limited on these critical questions. Investigators in special education have focused almost exclusively on within-child characteristics after special education placement occurs, thereby limiting opportunities to describe children before identification takes place (Coles, 1978). Researchers concerned with general education, on the other hand, have identified instructional practices related to academic achievement, but these results are typically reported as classroom averages as opposed to effects for similar groups of children (Good, 1983; Good & Weinstein, 1986). The latter issue is indicative of a phenomenon not well understood in either research or practice, the heterogeneity of children, especially those who are at risk for school failure or classification as handicapped (Satz & Fletcher, 1980).

The current study was designed to describe and assess the effects of this variability among children in their first year of schooling, before special education identification procedures were initiated. We entered into this study of child characteristics mindful of the National Academy of Sciences Panel (Heller, Holtzman, & Messick, 1982) and Messick's (1984) admonition that sole reliance on child characteristics to the exclusion of classroom environment variables ignored an important factor in the ontogeny of school failure. Therefore, this investigation of child characteristics is seen as a necessary first step in a program of research linking developmental factors in the early appearance of school failure with appropriate classification and placement procedures, and ultimately with methods for avoiding over-reliance on special education services in the early elementary grades.

This question, who is at risk for early school failure, is especially critical given recent educational policy initiatives for general education and research findings on the value of special education for the mildly handicapped. Some educators have argued for the divestiture of mildly handicapped children from special education (Reynolds, Wang, & Walberg, 1987; Will, 1986). Even though the empirical base for such a policy is limited and methodologically weak (Fuchs & Fuchs, 1988), it is not clear that special education is the appropriate education setting either. For example, Haynes and Jenkins (1986) found that the instructional processes in special education resource classrooms may not benefit some children especially when reading instruction is the sole province of special education. It seems, however, that an important question has been overlooked. Before focusing on whether general or special education is more effective, it is important to design appropriate classification methods to more clearly define the children who become candidates for placement in different instructional programs. The variability of classroom processes represented by general and special education placements is mirrored by the heterogeneity of children who are served. While this heterogeneity of skills is probably evident across the developmental spectrum, we chose to describe children during their first year of schooling. It is at this time that characteristics that may lead to school failure are first observed systematically by professionals but before formal recognition of failure that leads to classification and differential placement decisions.

To describe the ontogeny of school failure and the first year outcomes associated with this status, both multiple measures and multivariate analysis strategies were selected. First grade children were assessed in several domains including intelligence, achievement, language, behavior and learning potential. Cluster analysis methods were used to identify empirical subtypes of children, and thus describe these children with respect to critical domains of cognitive and social development and independent of classroom membership. Outcome, or validation procedures, included the use of relative risk analysis to determine which subtypes were associated with higher degrees of risk for failure, as indicated by teacher referral. Additionally, achievement variables not included in subtype identification procedures and children's responses to classroom environments were used to validate the empirical subtypes. (For further discussion of the theoretical approach, see Cooper & Speece, 1988.)

These methods were adopted to answer the following questions: (a) can reliable subtypes of children be identified based on child characteristic variables, (b) given reliable subtypes, what is the relative risk of failure associated with each subtype and (c) can these subtypes also be differentiated on the basis of additional achievement variables and observed classroom behavior? The last two questions addressed subtype validity. The process of establishing external validity is an integral component of applied cluster analysis methodology. Because cluster analysis procedures will yield clusters even with random data, validation efforts are necessary to verify that cluster profiles represent meaningful, as opposed to random, groupings of children (Milligan & Cooper, 1987; Skinner, 1981).

Another methodological feature was the clustering of both at-risk and average achieving children in the same analysis to determine the extent to which these two groups shared cluster membership. As noted by Speece (in press), past subtyping investigations typically have used the performance of normally achieving children only as baseline data, leaving open the question of the uniqueness of the obtained subtypes. A developmental perspective would suggest that some at-risk children may share cluster membership with average achievers while others may comprise separate clusters. The methods selected for this investigation allowed examination of this possibility.

Method

Subjects

Children. The sample of children (N = 112) represented two cohorts of first graders. Because our target population was students at risk for referral to special education, but who were not yet classified as handicapped, two public school districts were selected in which "pre-referral" identification systems (described below) were operating. County A had a general population of 660,000 with a racial composition of 46% Black, 49% Caucasian, and 5% other ethnicities. The total enrollment for the school system was 103,325 with 62.6% Black, 31% Caucasian, and 6.4% other. The population for County B was 414,074 with a racial distribution of 12.6% Black, 85.4% Caucasian, and 2% other ethnicities. The school system population was 64,552 and reflected the racial composition of the general population with 14.2% Black, 83.8% Caucasian, and 2% other ethnicities. Elementary schools in both districts used variants of the Teacher Assistance Team (TAT) model prior to evaluation by special education teams. The TATs were composed of regular and sometimes special education teachers whose duties included reviewing the TAT referral with the referring teacher, providing alternatives for instruction and/or behavior management, and monitoring students' progress in the regular classroom.

The selection of at-risk first graders was predicated on the following criteria: (a) referral to TAT for academic or behavioral problems, (b) the child was in first grade for the first time, (c) the child's native language was English, and (d) if supplementary instruction, such as Chapter One assistance, was provided outside the regular classroom, the total amount of time did not exceed 2 1/2 hours per week. To reduce the data collection burden on any one teacher, we included only the first two at-risk children nominated per classroom. In point of fact, referral of more than two students per classroom was a rare occurrence. When permission for inclusion was obtained from the parents of an at-risk child, a control child of the same sex and from the same classroom was randomly selected from the teacher's list of children achieving and behaving normally, and who, in the teacher's opinion, would not be referred either to TAT or programs for the gifted and talented. Thus, the initial sample contained equal numbers of at-risk and control children. Parent permission return rates for this sample were 60.2% for at-risk students and 73.6% for controls[1]. Only children with complete data sets were included in the analyses, thus there were unequal numbers of at-risk and control children.

The last column in Table 1 provides demographic data for the sample of 63 at-risk (AR) and 49 control (C) children. In comparison to the general school population data (referred to earlier), our sample is disproportionately male (76.8%) and, to a lesser extent, Black (42.9%). With regard to socio-economic status, the sample appeared to reflect the national median for mother's educational level; (AR median = 12 years, C median = 13 years, U.S. median = 12.6 years; Bureau of the Census, 1985).

[INSERT TABLE 1 ABOUT HERE]

Teachers. First grade teachers in the participating schools were included in the study when a referral to TAT from their class resulted in a subject meeting all inclusion criteria. In addition to being the source of referrals to TAT, teachers contributed to the children's data set by completing classroom behavior ratings, and by permitting observations of the classroom learning environment and student responses.

Teacher background data were available on 32 of 48 participating teachers.[2] This information suggested the sample was well trained. Sixty-six percent had at least 15 hours past the bachelor's degree, and in the appropriate specialty areas: 87% in elementary or early childhood education. The teachers were also, as a group, experienced in teaching first grade: $\bar{x}$ = 11.4 years, SD = 8.7. Class-size was typical for public school first grades: $\bar{x}$ = 23.8, SD = 2.4.

Measures

Little work has been done in the area of risk factors that are associated with placement of primary grade students in special education. Therefore, selection of constructs and appropriate measures relied on a combination of traditional educational measures (IQ and achievement), putative risk factors that appeared reasonable/logical given classroom demands (classroom discourse skills and learning potential), and the limited evidence available on empirically determined risk factors (e.g., classroom behavior; Cooper & Farran, 1988; McKinney & Speece, 1986). Whether or not our selection of student characteristic variables was correct is largely an empirical question, but there appeared to be partial congruence between our conceptualization and the Project PRIME model of learning competence presented by Kaufman, Agard and Semmel (1985). Both the Project PRIME model and ours included normative academic status (reading and math achievement), classroom academic status (effort and expression), cognitive interactions (participation, discourse), and on-task behavior. Table 2 provides a listing of the measures according to their use as classification or validation measures. Details regarding instrumentation are summarized below.

[INSERT TABLE 2 ABOUT HERE]

Intelligence and achievement. Verbal and non-verbal intelligence scores were obtained from the Cognitive Abilities Test (CogAT; Thorndike

& Hagen, 1982). The Diagnostic Achievement Battery (DAB; Newcomer & Curtis, 1984) provided measures of achievement in reading, arithmetic, listening and spoken language. Both the CogAT and DAB have received favorable reviews regarding psychometric adequacy (Brown & Bryant, 1984; Salvia & Ysseldyke, 1985).

Classroom discourse skills. Several studies (e.g., Blank, Berlin & Rose, 1983) have focused on language of instruction as a key to student achievement. The Preschool Language Assessment Instrument (PLAI; Blank, Rose & Berlin, 1978) was specifically designed to assess children's ability to process the types of oral language questions found in classrooms, across four levels of cognitive complexity from simple perception to abstract reasoning. Blank et al. (1983) asserted that "If children experience difficulty on the test, then it implies that they will experience difficulty in understanding the verbal exchange that is at the heart of the classroom experience" (p. 299). Blank et al. (1978) reported that across the four levels of questions split-half reliability ranged between .64 and .86 with test-retest reliability ranging from .73 to .88. Our assessment of inter-rater reliability (three individual scorers' agreement with scores on 5 standard protocols) was calculated for each of the 4 levels of cognitive complexity and yielded mean correlations of .97, .99, .99 and .89, respectively.

Learning potential. Since Vygotsky advocated the study of children's thought under the "influence of instruction" (cited in Hamilton, 1979), several researchers have evaluated the advantages of dynamic assessment of learning over static measures of intellectual development (see Hamilton, 1979). Recently, Brown and her colleagues (Bryant, Brown, & Campione, 1983; Ferrara, Brown, & Campione, 1986) have extended the earlier work with mentally retarded subjects to studies of non-handicapped students' flexible use of knowledge as indicated by dynamic assessment methods that are correlated with, but not redundant with IQ. The work cited above did not include a learning potential measure appropriate for first-graders, thus we developed the Dynamic Assessment Tasks (DAT), based on that work. The DAT is a guided tutorial experience wherein the child is taught to solve four difficult matrix problems that follow a single rule (superimposition). A sequence of standard, graduated oral prompts was provided until the child solved the matrix puzzle. Two variables were derived from the DAT. First, the total number of prompts given (TLP) was interpreted as the amount of teaching required by the child to arrive at the correct solution. Second, the child's independent performance on six post-test matrices following training provided a measure of the amount of learning that occurred during the session. To control for prior knowledge on this type of task, a residual post-test gain score (RPT) was calculated for each child within each group (AR and C) using the matrix subtest from the CogAT as the predictor variable. Following Peterson, Homer, and

Wonderlich (1982), procedural reliability data were collected to determine the degree to which examiners were adhering to the tutorial scripts written for the DAT. Eighteen sessions were recorded on audiotape and seven were randomly selected and rescored by an independent observer. Based on the percentage agreement method, the procedural reliability for all protocols was 100%.

Classroom behavior. Two approaches to the assessment of classroom behavior were used. The Cooper-Farran Behavioral Rating Scales (CFBRS, Cooper & Farran, 1984), a 37-item teacher rating instrument, provided scores on two child behavior factors: interpersonal skills (INT) and work-related skills (WRK). INT tapped physical and verbal aggressiveness and disruptiveness while the WRK factor measured disorganization, distractibility, and non-compliance. Data supporting adequate reliability of the two factors were reported in Cooper and Farran (1988) and Cooper (1984). Intra-class correlations above .78 were obtained for both factors when analyzed for inter-rater reliability. Content validity was established during scale development and by estimating internal consistency (Cronbach's aipha = .96). Construct validity has been examined by means of factor analytic studies. In five independent data sets, totalling over 1400 subjects, the two factor structure (INT and WRK) has been replicated.

The second measure of behavior was based on a modified version of the Code for Instructional Structure and Student Academic Response

(CISSAR), a classroom observation instrument developed by Stanley and Greenwood (1981). CISSAR is a time sampling method in which classroom structure, task, and activity were coded every 60 seconds and teacher and child behavior were coded every 10 seconds. It has adequate reliability and validity (see Greenwood, Schulte, Kohler, Dinwiddie, & Carta, 1986). In the present study, interobserver agreement was checked weekly for each of two observers with a minimum level of 80% agreement required within each of the 5 aggregate categories.

The observation system categorized student responses into three categories (examples of specific codes listed in parentheses): active academic responses (reading, writing, answering questions); task management (attending to teacher, looking for materials); and inappropriate behaviors (looking around, inappropriate task, disruptive). For the purposes of this study, active academic responses and inappropriate behavior categories were converted to proportions (unconditional probabilities of occurrence) and used as cluster validation variables.

Procedures. Each child participated in three 50-minute test sessions with one of three examiners who were graduate students in education or psychology. The three sessions were devoted to (a) the Diagnostic Achievement Battery (DAB); (b) the Cognitive Abilities Test (CogAT; in some cases two children participated in this session); and (c) the Dynamic Assessment Tasks and the Preschool Language Assessment Instrument (DAT;

PLAI). The order of testing was randomly determined with the restriction that the CogAt always preceded the DAT. This procedure allowed the matrix subtest of the CogAt to be used as a pretest of the matrix problem-solving skills taught during the dynamic phase of the DAT. The tests were administered in private rooms in the schools and outside the regular classroom. All sessions, including classroom observations, occurred before the children's lunch period and were usually administered one week apart between January and May. Each child was the focus of two sessions of classroom observation using the CISSAR system; for at-risk children, 40 minutes (2 20-minute sessions) and control, 20 minutes (2 10-minute sessions). At-risk and control children's observations were interspersed within a session, with the observer focusing on the at-risk child for 5 minutes, then switching to the control child for 5 minutes, and so on. Teachers completed the CFBRS during the spring school term.

Data Analysis

The analysis plan for this study incorporated empirical cluster analysis techniques to identify and replicate subtypes of children. Relative risk analysis, analysis of variance and tests of proportions were used to validate the subtypes. Although cluster analysis techniques are being used more frequently in educational research, they remain widely misunderstood and sometimes misapplied techniques (Milligan & Cooper, 1987). As such, the procedures used in this study to address the more difficult issues of the number of clusters and correct cluster membership are discussed in detail.

Cluster analysis procedures. The overall plan guiding the cluster analysis procedures included identification of candidate solutions on half of the sample, (randomly split with the restriction that each sample contained equally proportionate numbers of AR and C children), selection of solutions that performed best across several guidance functions, and replication of the selected solutions with the other half of the sample to identify one (or more) replicable solutions to carry forward to the validation stage. The variables used for the cluster analysis were the measures of intelligence, reading and math achievement, language, classroom behavior ratings, and learning potential described earlier. External data for validation included the listening and speaking composite scores from the DAB, child behavior variables derived from the classroom observation instrument (CISSAR), and relative risk for referral to TAT.

Hierarchical cluster analysis with correlation as the measure of similarity was used with several algorithms (Ward's minimum variance, complete linkage, average linkage, single linkage) in the initial stage of the cluster analysis with the first sample. To identify candidate solutions (i.e., the number of clusters evident for each algorithm), solutions were evaluated on (a) the pseudo F and t^2 statistics (Milligan & Cooper, 1985; SAS Institute, 1985), (b) the Cubic Clustering Criterion (Milligan & Cooper, 1985; Sarle, 1983; SAS Institute, 1985), (c) visual inspection of cluster separation via a plot of the canonical discriminant functions for each solution (Aldenderfer & Blashfield, 1984), and (d) agreement on subjects' cluster membership across algorithms for a particular number of clusters (Lorr, 1983) assessed via the kappa statistic (Cohen, 1960). Each of these criteria suggested a specific number of clusters in the sample. Agreement across these approaches for each algorithm provided evidence from which to select the appropriate number of clusters.

Other criteria were used to determine cluster membership. Correct cluster membership of subjects is a problem specific to hierarchical techniques since members are not reassigned in later stages of the clustering process (as the number of clusters approaches one). To address this issue, discriminant function analysis ("forecasting") was used to determine if subjects had a higher probability of membership in a cluster other than the original assignment (based on Feild & Schoenfeldt, 1975) and were reassigned as necessary. Membership concordance was also assessed across algorithms with the kappa statistic. A final method compared membership agreement between the hierarchical clusters (with reassignments) and clusters identified through a nonhierarchical technique using the centroids from the hierarchical solutions as seeds (Milligan, 1980; Morris, Blashfield, & Satz, 1986).

In the replication phase, the solutions that were most successful as evaluated against the above criteria were applied to the second sample with the appropriate number of clusters retained. Following Morey, Blashfield, and Skinner (1983), subjects in the second sample were also assigned to the initial sample clusters via the classification function derived from a discriminant function analysis. The degree of replication was determined by assessing membership agreement (via kappa) between the assigned and empirical clusters. These procedures represented the most stringent test for replicability.

The identification of reliable clusters using correlation (shape of profile) as the similarity measure was followed by an additional cluster analysis with distance as the similarity measure. This procedure assessed the effects of scatter and elevation within each cluster and was based on membership of the total sample (Skinner, 1978). The decision to retain or reject the resulting clusters was based on agreement between the pseudo $\underline{F}$ and $\underline{t}^2$ statistics. All analyses were performed with SAS software (SAS Institute Inc., 1985). Further support for the analysis strategies used is provided in a methodological review by Speece (in press).

Validation Procedures. Following the identification of reliable clusters, the third stage of the analysis plan involved assessing cluster differences with procedures and variables external to the data set used for cluster formation. Because cluster formation was based on multiple domains, we also adopted a multivariate perspective toward cluster validation. That is, we anticipated that several different pieces of evidence would be necessary to capture cluster differences and that, taken together, converging evidence would be found to support validation. Relative risk analysis, analysis of variance on achievement variables and differences in probabilities of observed classroom behavior codes were used to assess external validation.

Relative risk analysis, a design strategy adapted from epidemiological research (Kleinbaum, Kupper, & Morganstern, 1982), provided a method for determining which clusters of children were associated with a higher degree of risk of being identified by teachers as candidates for learning and behavior problems (i.e., referral to Teacher Assistance Teams). Within each cluster, AR children were considered "cases" and C children were considered "controls" in epidemiologic terminology. The relative risk ratio compared the prevalences of cases in the clusters of interest to the prevalence of cases in the baseline cluster that exhibited normal performance on the multiple measures. The extent to which this ratio deviated from unity was interpreted as the increased risk for referral associated with the cluster profile, relative to the baseline cluster.

Results

Identification and Replication of Clusters

The initial cluster structures obtained from four algorithms were evaluated against seven criteria to provide evidence regarding the number of clusters and appropriate cluster membership for the split sample. Three solutions were retained for further analysis (evidence for these selections described in parentheses): (a) Ward's algorithm, three clusters (strong concordance between pseudo $\underline{F}$ and $\underline{t}^2$ statistics, well separated clusters in the plot of the canonical discriminant functions, moderate but significant membership agreement across different algorithms, $\underline{K}$ = .40); (b) complete linkage algorithm, four clusters (strong concordance between pseudo $\underline{F}$ and $\underline{t}^2$ statistics, visible separation in the plot of canonical discriminant functions, and moderate to strong membership agreement with a different algorithm, $\underline{K}$ = .60); and (c) average linkage algorithm, six clusters (partial concordance between pseudo $\underline{F}$ and $\underline{t}^2$ statistics, evidence for six clusters partially supported by Ward's algorithm, moderate but significant membership agreement across different algorithms, $\underline{K}$ = .47). Clusters obtained from the single linkage algorithm were not supported by any criterion and the Cubic Clustering Criterion did not provide useful data for evaluating any solution.

Each of these three solutions was then compared with results from a nonhierarchical technique that used the hierarchical centroids of each cluster for seeds (Milligan, 1980). Membership agreement between these techniques was high with $\underline{K}$ = .94, .96, and 1.00 for Ward's three clusters, complete linkage four clusters, and average linkage six clusters, respectively.

Replication of these three solutions with the second sample yielded kappa values of .71, .51, and .32 respectively, which indicated that the Ward's three cluster solution had the highest degree of agreement between the two samples and should be retained for further analysis. The two samples were combined and each of the three Ward's clusters was submitted to another cluster analysis with distance as the similarity metric to determine if the addition of elevation and scatter to the shape data provided evidence of additional clusters. Each of the three clusters split into two clusters, supported by strong concordance between the pseudo $\underline{F}$ and $\underline{t}^2$ statistics, yielding a final six cluster solution.

Cluster Description

Table 1 provides demographic data for each of the six clusters and Figure 1 depicts the profiles across the classification variables. Each pair of clusters (i.e., 1-2, 3-4, 5-6) resulted from the splits of the three cluster solution. Inspection of these profiles indicated that the six cluster solution served to further decrease the heterogeneity within clusters.

The data points in Figure 1 represent the mean $\underline{z}$ score for the cluster members calculated from either the normative data (achievement and intelligence x = 100, SD = 15) or the x and SD for the control group on the nonstandard measures. Thus, zero represents average performance. Note that average performance across the PLAI variables (PL1-PL4) would be represented by scores at zero for each level. Additionally, the DAT variable called Total Prompts (TLP) was reversed such that a negative score was associated with a high number of prompts required by subjects to obtain the solutions to the matrix puzzles.

In addition to the empirical support for the obtained solution, the clusters depicted in Figure 1 presented interpretable profiles from educational and psychological perspectives. Three clusters, 2, 3 and 5, appeared to represent variations of normal performance while clusters 1, 4 and 6 appeared more deviant with respect to profile strengths and weaknesses. Each will be described next.

[INSERT FIGURE 1 HERE]

Cluster 1, composed of 28 AR and 2 C children, was suggestive of a learning disability profile. Children in this cluster were distinguished by the lowest ratings of any cluster on work-related (WRK) and interpersonal skills (INT), a discrepancy between achievement (RDG, MTH) and intelligence measures (VIQ, NIQ) and a gentle decline in discourse understanding associated with the increased complexity of the language task (PL1-PL4). Interestingly, the children in this cluster appeared to benefit from the instruction provided in the Dynamic Assessment Tasks (DAT). A relatively high number of prompts (TLP) resulted in above average residualized post-test gain (RPT). The DAT represented a highly structured, examiner-child interaction, a perhaps infrequent arrangement in elementary school classrooms. The ratio of males to females was approximately 5 to 1 in this cluster while the total sample reflected a 3 to 1 ratio. White children were slightly overrepresented in comparison to the sample demographics (see Table 1).

Cluster 2 consisted of an almost equal number of AR and C children (which was proportional to the sample) with strengths across most variables, notably nonverbal intelligence, the exception being a relative weakness in work-related and interpersonal skills (WRK, INT). This cluster, as with cluster 1, has a 5-1 gender ratio favoring males.

Children in cluster 3, with only 1 AR member, presented a normal profile of means, with all scores average or above. This cluster serves as the baseline or reference group for subsequent analyses of behavior and relative risk for referral.

Cluster 4 was suggestive of a "slow-learner" profile with generally depressed achievement, intelligence, language and work-related skills. This cluster, composed of 12 AR and 5 C children, demonstrated relative strengths in interpersonal skills but did not appear to benefit from instruction on the DAT. The number of females was second highest of the clusters.

Cluster 5, composed of 6 AR and 14 C children, had mean scores within the normal range but with a pattern different from cluster 3, the other normal profile, and suggestive of an achievement-intelligence discrepancy. Curiously, this profile was the only one in which DAT residual posttest gain scores (RPT) declined after instruction. These children may have had difficulties in applying their knowledge or sustaining attention in a rather intense tutorial setting. Overall, however, work-related and interpersonal skills were evaluated favorably by their first grade teachers.

The composition of cluster 6 was unique with regard to (a) gender, 6 females and 5 males; (b) race, all children were Black; and (c) the sharp decline in PLAI performance across increasingly complex discourse levels (PL1 - PL4). Like cluster 5, these children exhibited low achievement in both reading and math in comparison to intelligence. Like cluster 1, these children appeared to profit from the DAT training to the same degree (approximately 1 SD increase from TLP to RPT). However, they required many more prompts during training to solve the puzzles. The overall pattern was suggestive of weaknesses in language processing. This cluster had 11 members, 7 AR and 4 C children.

<u>Validation</u>

<u>Relative Risk</u>. The relative risk ratios for each cluster are depicted in Table 3 (and Figure 1) as values of odds ratios and were derived from comparison of each cluster to cluster 3, the normal profile or baseline group. The statistical significance of each risk estimate is based on examination of the lower bound of the 95% confidence interval. When this bound is greater than 1, the risk estimate is significant. Four of the five estimates were significant. Using cluster 2 as an example (RRR = 27), the

interpretation of the risk ratio is as follows: risk of referral for the profile associated with cluster 2 is 27 times that of the profile of cluster 3, the "normal" cluster. Only cluster 5 failed to significantly differ from cluster 3 in risk for referral.

In addition to establishing risk estimates relative to cluster 3, each of the five clusters may be compared with each other. Cluster 1 (RRR = 252) was significantly different from the other four clusters as was cluster 5 (RRR = 7). Clusters 2, 4 and 6 did not differ from each other (RRR = 27, 43.2, 31.5, respectively).

[INSERT TABLE 3 ABOUT HERE]

Achievement variables. Two achievement composite scores not used in cluster formation, DAB speaking and listening, were used as dependent variables in separate ANOVAs with cluster membership as the independent variable. The descriptive data are presented in Table 3. The speaking composite reflected subtest scores on synonyms and grammatical completion and planned comparisons were made between each pair of clusters (1 vs. 2, 3 vs. 4, 5 vs. 6). As predicted, the ANOVA was significant, $F\ (5, 106) = 9.29$, $p < .001$, as were all contrasts (p's $< .009$ in all cases). The listening composite score reflected performance on story comprehension and word characteristics subtests (e.g., "All mothers are women, True or False"). The same planned comparisons were conducted as for speaking, but it was predicted that clusters 1 and 2 would not differ. We anticipated that while these two clusters differed with regard to reading and math achievement and intelligence, the generally adequate performance on the language variables (PL1 - PL4) and strong performance on the Dynamic Assessment Tasks (RPT) by children in cluster 1 would attenuate these differences. This did not turn out to be the case as significant differences were obtained from the ANOVA, $\underline{F}$ (5, 111) = 8.26, p < .0001, and for all contrasts ($\underline{p}$'s < .009).

Observed classroom behavior. Two composites of classroom behavior codes were used to contrast selected clusters for purposes of external validation. Following Greenwood et al. (1985), composites were formed for active academic responses (academic) and inappropriate behaviors (inappropriate). Proportions of total observed time (unconditional probabilities) were tested for differences among the clusters using a $\underline{z}$-statistic due to Allison and Liker (1982). Because the focus of this analysis was on cluster validation, only specified contrasts were examined. The data are summarized in Table 3.

The first contrast examined the validity of cluster 3, the group exhibiting the normal profile on the child characteristics battery. For this contrast cluster 3 was compared to aggregated data from all other clusters. The probability of academic responding for cluster 3 was .2950 and for all other clusters, .2447. The difference between .2950 and .2447, was statistically significant ($\underline{z}$ = 5.02, $\underline{p}$ < .001). In addition, children in cluster

3 were found to be coded inappropriate significantly less often than other children (probabilities = .153 and .2398, respectively, z = 8.92, p < .001). These findings supported the mean cluster profile which indicated average to above average skills for this group of children.

The distinctiveness of the profile associated with cluster 1 was demonstrated, in part by the substantially higher relative risk for referral reported above. The validity of this high-risk profile was further confirmed by the CISSAR observations. Children in cluster 1 were, relative to others, less likely to make active, academic responses (probabilities = .204 and .272, respectively, z = 8.568, p < .001) and significantly more likely to be inappropriate (probabilities = .2588 and .2135, respectively, z = 5.909, p < .001).

Additional validation analyses focused on the differentiation by CISSAR observation data among the three "normal" clusters (2, 3 and 5), and likewise, among the three "non-normal" clusters (1, 4, and 6). Because these analyses were post hoc, Fleiss' (1981) suggestion to employ stringent significance levels (p < .01) was adopted.

Contrasts among the three normal clusters, 2, 3 and 5, revealed significant differences on academic and inappropriate response probabilities. Specifically, clusters 2 and 3 (combined due to no difference between them) were observed to be more academic than cluster 5 (probabilities = .2966 and [illegible]14, respectively, z = 3.01, p < .01). Cluster 2 was more inappropriate than cluster 3 (probabilities = .2912 and [illegible]536, respectively, z = 10.7, $p < .001$). And cluster 2 was more inappropriate than cluster 5 (probabilities = .2912 and .2199, respectively, z = 5.34, $p < .001$).

Similarly, contrasts among the three non-normal clusters, 1, 4 and 6, revealed with one exception, significant differences on academic and inappropriate responses. Specifically, cluster 1 (suggestive of learning disabilities) was less frequently coded academic than was cluster 6 (language difficulties, probabilities = .2040 and .2698, respectively, z = 5.37, $p < .001$) and more inappropriate than cluster 6 (probabilities = .2588 and .1704, respectively, z = 7.08, $p < .001$). Clusters 4 (slow learners) and 6 did not differ on academic responding (probabilities = .2391 and .2698, respectively, z = 2.14, $p > .01$). However, cluster 4 was significantly more inappropriate than cluster 6 (probabilities = .2292 and .1704, respectively, z = 4.43, $p < .001$), and less inappropriate than cluster 1 (probabilities = .2293 and .2588, respectively, z = 2.60, $p < .01$).

Thus, observed classroom behavior further differentiated the clusters in that 2, 3, and 5 did not respond uniformly to the classroom environment despite mean cluster profiles in the normal range. Differences among the three atypical clusters were also found and were congruent with the interpretations of the profiles and the relative risk analyses.

Discussion

This study is set in an educational context in which the term "risk" has become pervasive quickly. Across the developmental span, segments of the population are said to be at risk for developmental delay, child abuse, school drop-out, and teen pregnancy. Despite its currency, the concept of "risk" is still a newcomer to the educational lexicon and thus lacks a foundation in the theory, practice, and research traditions of education. The present study provides initial data on children's multidimensional risk profiles and quantification of the risk associated with these profiles.

We found that six clusters, reliable across two samples, captured the variability of child characteristic measures and could be summarized in an educationally meaningful way by examining the multivariate cluster profiles. Our interpretation of the profiles is that three clusters represent meaningfully different variations of normal performance (clusters 2, 3 and 5), while three appear to represent atypical patterns of behavior resembling learning disabilities (cluster 1), mild mental retardation (cluster 4), and language disabilities (cluster 6). Distinctions among the three normal clusters and among the three atypical clusters are supported by the behavioral observation data.

The clusters differ with respect to achievement variables and responses to the classroom environment, providing further evidence of cluster validation. Planned comparisons for the speaking and listening composite variables support the interpretation that the split of the original three clusters produced clusters that differed with respect to elevation or severity (i.e., three "normal" profiles and three atypical profiles).

The differences on observed classroom behavior are small yet meaningful and provide an additional dimension for interpretation of the subtypes. For example, clusters 2 and 3 have essentially identical rates of active academic responding but children in cluster 2 exhibit almost twice the amount of inappropriate behavior. It appears that children in cluster 2 are those who quickly catch on to academic tasks, finish their work early, and then respond inappropriately. Given that these children behave inappropriately at the same or higher rate as children in the highest risk group (cluster 1), yet have a lower risk of referral, it may be that superior achievement, intelligence and language serve as protective factors, ameliorating the risk associated with inappropriate behavior (see Cooper & Farran, 1985). Cluster 6, the lowest achieving cluster and the one thought to be associated with language problems, exhibited the second lowest rate of inappropriate behavior and the third highest rate of academic responding. While it may be reasonable to expect that low achievement would be associated with inappropriate behavior, such was not the case with these children. Teachers rated their interpersonal skills as adequate, and the children responded well to the DAT tutorial instruction. Their below average ratings on work-related skills may result from their deficient processing of instructional discourse. Overall, the behavioral observation data substantiate the interpretation of six distinct clusters.

Whether the children in these clusters will be classified as handicapped during their tenure in school is a matter for further study. Placement in special education was a rare occurrence in this sample during first grade (n = 5) probably due to early identification procedures that had identified manifestly handicapped children prior to school entry, and eligibility criteria that were difficult to meet with young children. In addition to a longitudinal perspective to further support the validation data offered here, additional replication efforts are needed to assess the generality of the cluster profiles. While the sample size is adequate for the procedures, we plan to extend the cluster analysis and validation approaches to an independent sample as a second test of the consistency of the classifications obtained in this study.

The validation issues are especially important in regard to the results of the relative risk analysis. In contrast to baseline cluster 3, four of the five other clusters exhibit significantly elevated degree of risk of being identified by teachers as inadequately responding to the regular classroom environment. Perhaps most striking is the extreme risk ratio associated with cluster 1 which is higher than the risk associated with each of the other clusters. While the linkage between risk of referral to Teacher Assistance Teams and risk of classification as handicapped cannot be made with the present data set, it seems reasonable to speculate that children whose profiles are associated with higher degrees of risk prior to entering the special education identification process will likely evidence more risk of being classified as handicapped in the absence of well-defined intervention efforts. Further, it is notable that none of the four clusters with elevated risk share a common profile feature. That is, it appears the multivariate profiles were more salient in defining these children than was any single variable. This finding suggests that the development of school failure should be viewed as a more complex phenomenon than unitary risk factors would suggest.

While this more general interpretation of the findings and the corollary that teachers cope with a high degree of heterogeneity in their classrooms is not surprising, the value of this work lies in giving shape to the amorphous notion of heterogeneity via educationally useful markers. Heterogeneity of students is a truism unless that variance can be captured and described by a compromise between the idiographic and nomothetic extremes. Therefore, beyond statistical validation we plan a process of ecological validation that will include evidence that the clusters may respond differentially to differentiated instruction. Efforts along these lines, the design for which is described elsewhere (Cooper & Speece, 1988), include teacher-reported data on curriculum and methods, as well as time-sampled observations of classroom ecologies. We may find, for example, that students having the profile associated with cluster 2 (high IQ and inappropriate behavior) are at risk for failure, referral and placement only in classrooms that do not afford opportunities for individualized, self-paced learning. Similarly, cluster 6 (language and reading difficulties with adequate work-related and interpersonal skills) may be at risk only in classrooms that feature a great deal of group oriented oral instruction. Having indulged in the foregoing speculation, we wish to re-emphasize the necessity of taking the first step toward instructional relevance: valid classification.

A more specific finding with regard to child variance is that each cluster contained both at-risk and control children and each profile with the exception of cluster 5, was associated with elevated risk of referral compared with cluster 3. This finding provides empirical support for the belief that teacher referral is motivated by more than children's psychometric performance (Gerber & Semmel, 1984; Messick, 1984). That is, control children exhibit patterns and levels of performance similar to the risk children but are not referred by their teachers. Our intent is to follow these clusters of children to identify differences between at-risk learners who will and will not have been placed in special education. Thus the developmental course from school entry, through teacher recognition of at-risk performance to classification and placement can be mapped.

References

Aldenderfer, M.S., & Blashfield, R.K. (1984). Cluster analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-044. Beverly Hills and London: Sage Publications.

Allison, P.D. & Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. Psychological Bulletin, 91, 393-403.

Blank, M., Rose, S. A., & Berlin, L. J. (1978). Preschool language assessment instrument: The language of learning in practice. Orlando, FL: Grune & Stratton, Inc.

Blank, M., Berlin, L.J. & Rose, S.A. (1983). Classroom dialogue: Abilities and disabilities of learning disabled children. In J.D. McKinney & L. Feagans (Eds.) Current topics in learning disabilities (pp. 285-307). Norwood, NJ: Ablex.

Brown, L. & Bryant, B. (1984). Critical reviews of three individually administered achievement tests. Remedial and Special Education, 5(5), 53-60.

Bryant, N. R., Brown, A. L., & Campione, J. C. (1983, April). Preschool children's learning and transfer of matrices problems: Potential for improvement. Paper presented at the Society for Research in Child Development, Detroit, MI.

Bureau of the Census (1985). Statistical abstract of the United States.

Washington, DC: U.S. Department of Commerce.

Coles, G. S. (1978). The learning-disabilities test battery: Empirical and social issues. Harvard Educational Review, 48(3), 313-340.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cooper, D. H., & Farran, D. C. (1984). Cooper-Farran Behavioral Rating Scales. (Available from D. H. Cooper, Department of Special Education, University of Maryland, College Park, MD, 20742.)

Cooper , D. H. (1984). Identification of critical behaviors in kindergarten. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.

Cooper, D.H. & Farran, D.C. (1985). Behavioral risk: An ecological perspective on classifying children as behaviorally maladjusted in kindergarten. In W. Frankenburg, R. Emde, & J. Sullivan (Eds.) Early identification of children at risk: An international perspective (pp. 192-202) New York: Plenum Press.

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk factors in kindergarten. Early Childhood Research Quarterly, 3, 1-19.

Cooper, D.H. & Speece, D.L. (in press). A novel methodology for the study of children at risk for school failure. Journal of Special Education.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. Child Development, 57, 1087-1099.

Feild, H.S. & Schoenfeldt, L.F. (1975). Ward and Hook revisited: A two-part procedure for overcoming a deficiency in the grouping of persons. Educational and Psychological Measurement, 35, 171-173.

Fleiss, J.L. (1981). Statistical methods for rates and proportions. New York: Wiley.

Fuchs, D., & Fuchs, L.S. (1988). An evaluation of the adaptive learning environments model. Exceptional Children, 55, 115-127.

Gerber M.M. & Semmel, M.I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. Educational Psychologist, 19, 137-148.

Good, T. L. (1983). Classroom research: A decade of progress. Educational Psychologist, 18, 127-144.

Good, T.L. & Weinstein, R.S. (1986). Schools make a difference. American Psychologist, 41(10), 1090-1097.

Greenwood, C.R., Delquadri, J.C., Stanley, S.O., Terry, B. & Hall, R.V. (1985). Assessment of eco-behavioral interaction in school settings. Behavioral Assessment, 7, 331-347.

Greenwood, C.R., Schulte, D., Kohler, F., Dinwiddie, G. & Carta, J. (1986). Assessment and analysis of eco-behavioral interaction in school settings. In R. Prinz (Ed.) Advances in behavioral assessment of children and families (Vol. 2, pp. 69-98). Greenwich, CT: JAI Press.

Hamilton, J.L. (1979). Assessment in mental retardation: Toward instructional relevance. In R.B. Kearsley & I.E. Sigel (Eds.) Infants at Risk: Assessment of cognitive functioning (pp. 181-203). Hillsdale, NJ: Lawrence Erlbaum.

Haynes, M.C. & Jenkins, J.R. (1986). Reading instruction in special education resource rooms. American Educational Research Journal, 23, 161-190.

Heller, K. Holtzman, W., & Messick, S. (1982). Placing children in special education: A strategy for equity. Washington, D.C.: National Academy Press.

Kaufman, M., Agard, J.A., & Semmel, M.I. (1985). Mainstreaming: Learners and their environment. Cambridge, MA: Brookline Books.

Kleinbaum, D., Kupper, L., & Morgenstern, H. (1982). Epidemiologic research: Principles and quantitative methods. Belmont, CA: Wadsworth.

Lorr, M. (1983). Cluster analysis for social scientists. San Francisco: Jossey Bass Inc.

McKinney, J.D. & Speece, D.L. (1986). Academic consequences and longitudinal stability of behavioral subtypes of learning disabled children. Journal of Educational Psychology, 78, 365-372.

Messick, S. (1984). Assessment in context: Appraising student performance in relation to instructional quality. Educational Researcher. 13 (3), 3-8.

Milligan, G.W. (1980). An examination of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45, 325-342.

Milligan, G.W. & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50, 159-179.

Milligan, G.W. & Cooper, M.C. (1987). Methodology review: Clustering methods. Applied Psychological Measurement, 11, 329-354.

Morey, L., Blashfield, R. & Skinner, N. (1983). A comparison of cluster analysis techniques within a sequential validation framework. Multivariate Behavioral Research, 18, 309-329.

Morris, R., Blashfield, R. & Satz, P. (1986). Developmental classification of learning disabled children. Journal of Clinical and Experimental Neuropsychology, 8, 371-392.

Newcomer, P. L., & Curtis, D. (1984). Diagnostic Achievement Battery. Austin, TX: PRO-ED.

Peterson, L., Homer, A.L., & Wonderlich, S.A. (1982). The integrity of independent variables in behavioral analysis. Journal of Applied Behavioral Analysis, 15, 477-492.

Reschly, D.J. (1984). Beyond IQ test bias: The National Academy panel's analysis of minority EMR over-representation. Educational Researcher, 13(3), 15-19.

Reynolds, M.C., Wang, MC., & Walberg, H.J. (1987). The necessary restructuring of special and regular education. Exceptional Children, 53, 391-298.

SAS Institute Inc., (1985). SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc.

Salvia, J., & Ysseldyke, J.E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton-Mifflin.

Sarle, W.S. (1983). Cubic clustering criterion. SAS Technical Report A-108. Cary, NC: SAS Institute Inc.

Satz, P., & Fletcher, J. M. (1980). Minimal brain dysfunctions: An appraisal of research concepts and methods. In H. E. Rie & E. D. Rie (Eds.), Handbook of minimal brain dysfunctions (pp. 669-715). New York: John Wiley & Sons.

Skinner, H.A. (1978). Differentiating the contribution of elevation, scatter, and shape in profile similarity. Educational and Psychological Measurement, 38, 297-308.

Skinner, H.A. (1981). Toward the integration of classification theory and methods. Journal of Abnormal Psychology, 20, 68-87.

Speece, D.L. (in press). Methodological issues in cluster analysis: How clusters become real. In H.L. Swanson & B.K. Keogh (Eds.), Learning disabilities: Theoretical and research issues. Hillsdale, NJ: Lawrence Erlbaum.

Stanley, S. O., & Greenwood, C. R. (1981). CISSAR: Code for instructional structure and student academic response: Observer's manual. Kansas

City, KS: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.

Thorndike, R. L. & Hagen, E. P. (1982). Cognitive abilities test, primary battery edition. Chicago: Riverside.

Will, M. (1986). Educating children with learning problems: A shared responsibility. Exceptional Children, 52, 411-415.

Footnotes

[1] The modest return rate of permission letters for AR children (60.2%) suggests the possibility of a sampling bias. The most likely effect of this return-rate bias would be under-representation of low SES children and perhaps a negatively skewed distribution of abilities. To address this question we sought evidence in our data that would be consistent with a restricted range of cognitive abilities, as measure by the CogAT. Results for the AR group indicated normal distribution of verbal intelligence, without negative skewness (skewness = .60) and a mean IQ for AR children ($\bar{x}$ = 94.1, SD = 10.16) one standard deviation below the mean for C children ($\bar{x}$ = 106.1, SD = 13.8). We interpreted these data as not consistent with selection bias.

[2] In order to take advantage of the data available on 32 teachers, yet to account for the missing data on 16 teachers, an estimation procedure was used. For each variable, the data were summarized (means or percentages) for the 32 teachers, along with estimates of the ranges that the mean or percentage would change if the data were not missing, and assuming that the 16 missing cases were within the same range as the 32 non-missing cases. If we assume that the teachers on whom data were not available were all at the lower end of the distributions, the data still reflect an experienced group of educators. The results of the estimates under the

"worst case" assumption are: 43.7% with a bachelors degree plus 15 hours, 58.3% with degrees in elementary or early childhood, and 7.9 years teaching first grade.

Table 1

Demographic Data for the 6 Cluster Solution

		Cluster						
		1	2	3	4	5	6	Total Sample
n (%)		30 (26.8)	15 (13.4)	19 (17.0)	17 (15.2)	20 (17.9)	11 (9.8)	112
Group								
AR	n (%)	28 (93.3)	9 (60.0)	1 (5.0)	12 (70.6)	6 (30.0)	7 (63.6)	63 (56.3)
C	n (%)	2 (6.7)	6 (40.0)	18 (95.0)	5 (29.4)	14 (70.0)	4 (36.4)	49 (43.7)
Gender								
F	n (%)	5 (16.7)	2 (13.3)	3 (15.8)	6 (35.3)	4 (20.0)	6 (54.6)	26 (23.2)
M	n (%)	25 (83.3)	13 (86.7)	16 (84.2)	11 (64.7)	16 (80.0)	5 (45.4)	86 (76.8)
Race								
B	n (%)	10 (33.3)	6 (40.0)	6 (31.6)	7 (41.2)	8 (40.0)	11 (100.0)	48 (42.9)
W	n (%)	20 (66.7)	9 (60.0)	12 (63.2)	9 (52.9)	11 (55.0)	0	61 (54.5)
O	n (%)	0	0	1 (5.2)	1 (5.9)	1 (5.0)	0	3 (2.6)
Mother's Education								
Median		12	14	13	12	12	12	12

Table 2

Summary of Measures

Measures	Cluster Formation	Validation
Diagnostic Achievement Battery (DAB)		
Reading Composite	x	
Math Composite	x	
Speaking Composite		x
Listening Composite		x
Cognitive Abilities Test (CogAT)		
Verbal IQ	x	
Non Verbal IQ	x	
Cooper-Farran Behavioral Rating Scale (CFBRS)		
Work-Related Skills	x	
Interpersonal Skills	x	
Preschool Language Assessment Instrument (PLAI)		
Level 1 (matching perception)	x	
Level 2 (analysis of perception)	x	
Level 3 (reordering perception)	x	
Level 4 (reasoning about perception)	x	
Dynamic Assessment Task (DAT)		
Total Training Prompts	x	
Residual Posttest Gain	x	
CISSAR Observation Instrument		
Active Academic Responses		x
Inappropriate Responses		x

Table 3

Descriptive Data for Cluster Validation Methods

Method		Cluster					
		1	2	3	4	5	6
Relative Risk[a]		252.0	27.0	-	43.2	7.7	31.5
Confidence Interval		(42-1512)	(4.1-177.9)		(6.9-271.3)	(1.0-57.9)	(4.4-226.8)
Speaking Composite	x	83.30	97.80	95.58	80.94	88.45	77.91
	SD	8.76	16.45	8.57	9.24	10.89	8.51
Listening Composite	x	87.20	102.80	101.58	85.00	96.40	84.18
	SD	11.17	15.02	13.80	13.33	8.85	9.50
CISSAR Proportions[b]							
Active Academic		.204	.298	.295	.239	.261	.270
Inappropriate		.259	.291	.154	.229	.220	.170

[a] Odds ratio and 95% Confidence Interval.

Risk is relative to Cluster 3.

[b] Each cluster's proportion for each code is the total number of occurrences of that code divided by the number of intervals observed.

Figure Caption

Figure 1. Mean profile and relative risk ratios (RRR) of clusters. (RDG = reading; MTH = math; VIQ = verbal IQ; NIQ = nonverbal IQ; WRK = rating of work related skills; INT = rating of interpersonal skills; PL1-PL4 = level scores on PLAI; TLP = total prompts on Dynamic Assessment; RPT = residual posttest gain score on Dynamic Assessment).

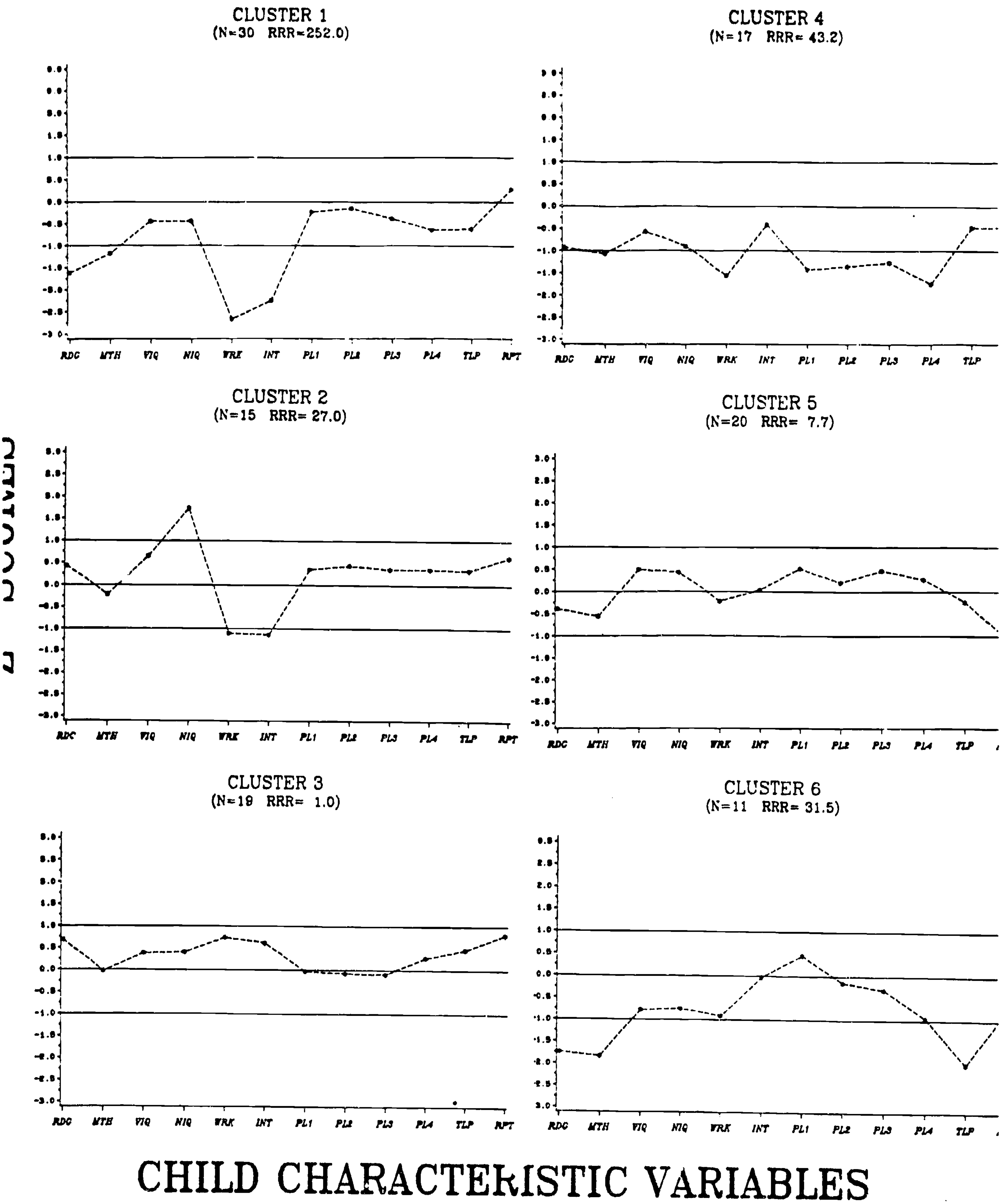


Figure 1. Mean profile and relative risk ratios (RRR) of clusters. (RDG=reading; MTH=mat VIQ=verbal IQ; NIQ=nonverbal IQ; WRK=rating of work related skills; INT=rating of interpersonal skills; PL1-PL4=level scores on PLAI; TLP=total prompts on Dynamic Assessment; RPT=residual posttest gain score on Dynamic Assessment).

Work in Progress

We continue to address the procedural objectives for this project even though the funding period has ended. As explained in our Year 3 Continuation Grant, the number of subjects recruited during Year 1 of the study was less than anticipated. This led to the recruitment of an additional cohort during Year 3 of the study that was not anticipated in the original grant application. To address procedural objectives 7 and 8, pertaining to subtypes of learning environments and risk reduction, second year data on cohort 3 were needed. Collection of these data will be completed by May 1989, and analyses will be completed by August, 1989. An addendum to this Final Report will be completed by December, 1989.

Appendix A

Curriculum and Methods Questionnaire

C.M.Q.
Curriculum and Methods Questionnaire

David H. Cooper & Deborah L. Speece
Project Search
The University of Maryland

DIRECTIONS FOR COMPLETION

THIS QUESTIONNAIRE IS TO BE COMPLETED WITH A FOCUS ON ONE AT-RISK STUDENT. THE STUDENT'S TEACHER SHOULD COMPLETE THE FORM AFTER SUFFICIENT TIME HAS PASSED TO ALLOW AN INSTRUCTIONAL ROUTINE TO HAVE BEEN ESTABLISHED (6-8 WEEKS). THIS ROUTINE, INCLUDING CURRICULUM AND METHODS, IS TO BE DESCRIBED BY MEANS OF ANSWERS TO THE QUESTIONS FOLLOWING.

PLEASE ANSWER THE QUESTIONS BASED ON THE TYPICAL, MOST OFTEN USED METHODS FOR THIS STUDENT. EMPHASIS IS GIVEN TO THE TERM "TYPICAL." IT IS RECOGNIZED THAT VARIATIONS OCCUR IN CURRICULUM AND INSTRUCTIONAL METHODS, HOWEVER, TO DOCUMENT ALL POSSIBLE VARIATIONS EMPLOYED WITH A GIVEN CHILD WOULD BE TOO CUMBERSOME.

PLEASE NOTE

THE CMQ TAKES APPROXIMATELY 45 MINUTES TO COMPLETE. ON THE ADVICE OF TEACHERS WHO HAVE USED THIS INSTRUMENT, WE SUGGEST YOU COMPLETE THE QUESTIONNAIRE OVER TWO OR THREE SESSIONS.

[staff use: ID: ____ ____ ____ ____ card = 1]
(1) (2) (3) (4) (5)

Descriptive Data

PLEASE PRINT

M D Y

TODAY'S DATE ____/____/____
(6 - 11)

STUDENT'S NAME: ____________ ____________ SCHOOL ____________
(first) (last) (12 - 27)

TEACHER ____________ ____________ GRADE ____ CLASS SIZE ____ ____
(28-57) (first) (last) (58) (59 - 60)

1. IS THIS CHILD'S NATIVE LANGUAGE ENGLISH? ____(Y/N) IF NO, DO NOT COMPLETE - RETURN QUESTIONNAIRE TO PROJECT SEARCH.
(61)

2. IS THIS CHILD RECEIVING CHAPTER ONE SERVICES?____(Y/N)
(62)

IF YES, BRIEFLY DESCRIBE (in class or pull out; number of hours per week) ____________

____________ [staff use: ___ ___]
(63 - 64)

3. IS THIS CHILD RECEIVING ANY SUPPLEMENTAL INSTRUCTION? (such as code 13 - speech, reading, resource, etc.) ____(Y/N)
IF YES, PLEASE ANSWER THE FOLLOWING: (65)

___ ___ # MINUTES PER DAY
(66 - 67)

____ # TIMES PER WEEK
(68)

____ TYPE OF SERVICE:
(69)

S = SPEECH (articulation)
L = LANGUAGE (oral or written)
B = BEHAVIOR MANAGEMENT PROGRAM
D = DIAGNOSTIC ASSESSMENT OF STUDENT'S SPECIAL NEEDS
R = READING
M = MATH

____TYPE OF INSTRUCTION:
(70)

S= SUPPLEMENTAL INSTRUCTION; USES SAME METHODS AND MATERIALS AS CLASSROOM
D= SUPPLEMENTAL INSTRUCTION; USES DIFFERENT METHODS AND MATERIALS THAN CLASSROOM

4. COMPARED TO PEERS, HOW WOULD THIS CHILD GENERALLY SCORE ON THE FOLLOWING: (S=STRONG A=AVERAGE W=WEAK)

____READING (71) ____MATH (72) ____FINE MOTOR AND HANDWRITING (73) ____GROSS MOTOR (74)

____ SPEECH (articulation) (75) ____ORAL LANGUAGE (76) ____SOCIAL BEHAVIOR (77)

5. ____ WHICH OF THE FOLLOWING HAS PROVIDED YOU WITH THE BEST INFORMATION IN DETERMINING THIS CHILD'S
(78) STRENGTHS AND WEAKNESSES?

S= STANDARDIZED TEST: ____________
(name)

L= LOCALLY DEVELOPED TEST
C= COMBINATION OF STANDARDIZED TESTS AND OTHER METHODS
T= TEACHER MADE TEST
I= INFORMAL ASSESSMENT
X= OTHER (please specify)____________

J = JUDGEMENT OF OTHER PROFESSIONAL
P = PREVIOUS RECORDS
R = CRITERION REFERENCED TEST
O = TEACHER OBSERVATION

Curriculum

IN THIS SECTION, THE TERM "CURRICULUM" REFERS TO THE PLAN FOR INSTRUCTION, INCLUDING OBJECTIVES RELATING TO KNOWLEDGE AND SKILLS TO BE LEARNED. CURRICULUM IS THE ORGANIZATIONAL GUIDE TO WHAT IS TO BE LEARNED. IF MORE THAN ONE CURRICULUM IS IN USE, PLEASE REFER ONLY TO THE PRIMARY OR MOST FREQUENTLY USED CURRICULUM FOR THIS CHILD.

Reading

[staff use: card = 2]

(5)

IF READING WAS NOT DESIGNATED AS A "WEAKNESS" ON ITEM 4 - PAGE 1, PLEASE PROCEED TO MATHEMATICS, PAGE 3.

1._____INDICATE WHICH AREA LISTED BELOW IS CURRENTLY THE MAJOR CONCERN FOR THIS CHILD
(6)
L= LETTER IDENTIFICATION
S= SOUND-SYMBOL RELATIONSHIPS
A= WORD ATTACK AND IDENTIFICATION
C= WORD/CONCEPT COMPREHENSION
F= ORAL READING FLUENCY
P= PASSAGE COMPREHENSION
O= OTHER (please specify)____________________

2._____INDICATE ANOTHER AREA OF CONCERN IF APPLICABLE TO THIS STUDENT
(7)

3._____PLEASE DESCRIBE THE CURRICULAR EMPHASIS THAT YOU USE FOR WORD IDENTIFICATION TECHNIQUES
(8)
W= WHOLE WORD
P= PHONETIC
C= COMBINATION OF WHOLE-WORD AND PHONETIC
L= LANGUAGE EXPERIENCE
T= CONTEXT BASED
O= OTHER (please specify)____________________

4. FOR THE MAJOR CONCERN INDICATED ABOVE, NAME THE READING CURRICULUM IN USE TO ADDRESS THAT CONCERN.

____

(9 -29) (curriculum)

5. _____PLEASE SPECIFY THE CURRICULUM SOURCE (indicate one of the following)
(30)
C= COMMERCIAL
T= TEACHER DESIGNED
L= LOCALLY DEVELOPED
P= PUBLISHED, NON-COMMERCIAL

6. _____DO YOU KNOW OF EVIDENCE THAT SUPPORTS THE SUCCESS OF THIS CURRICULUM WITH STUDENTS IN GENERAL?
(31)
N= NO KNOWLEDGE OF ANY EVIDENCE
S= STANDARDIZED TEST SCORES FOR LARGE SAMPLE OF STUDENTS
L= PUBLISHER'S LITERATURE, SUCH AS CURRICULUM GUIDE
R= RESEARCH STUDIES IN PROFESSIONAL LITERATURE
E= LOCAL EVALUATION STUDY
P= PERSONAL EXPERIENCE
O= OTHER (please specify)____________________

7.____ ____HOW MANY OTHER STUDENTS IN YOUR CLASS USE THE SAME READING CURRICULUM?
(32-33)

8. DOES THE CURRICULUM INCLUDE THE FOLLOWING? (Y/N)

_____A DESCRIPTION OF ITS THEORETICAL BASIS
(34)

_____OBJECTIVES STATED IN BEHAVIORAL TERMS
(35)

_____STUDENT MATERIALS (books, worksheets, etc.)
(36)

_____SUGGESTIONS FOR SPECIFIC MODIFICATIONS WHEN NEEDED
(37)

_____EVALUATION CRITERIA FOR STUDENT MASTERY
(38)

_____RECORD-KEEPING SYSTEM FOR STUDENT PROGRESS
(39)

9. IN YOUR EXPERIENCE, HAS THIS CURRICULUM BEEN EFFECTIVE WITH OTHER CHILDREN OF THE SAME:

SEX AS THIS CHILD?______(Y/N) (40)   CULTURE AS THIS CHILD?______(Y/N) (41)

Mathematics

IF MATHEMATICS WAS NOT DESIGNATED AS A "WEAKNESS" ON ITEM 4 - PAGE 1, PLEASE PROCEED TO ORAL LANGUAGE, PAGE 4.

10._____INDICATE ONE AREA CURRENTLY OF MAJOR CONCERN FOR THIS STUDENT:
(42)

R= NUMBER RECOGNITION
Q= QUANTITY AND MEASUREMENT
C= COUNTING AND ORDERING
S= SHAPE RECOGNITION
O= OPERATIONS WITH NUMBERS
F= MATH FACTS
P= PLACE VALUES
T= TIME
M= MONEY
N= NONE OF THESE (please specify)____________________

11.______INDICATE ANOTHER AREA OF CONCERN IF APPLICABLE TO THIS STUDENT
(43)

12. FOR THE MAJOR CONCERN INDICATED ABOVE, NAME THE MATHEMATICS CURRICULUM IN USE WITH THIS STUDENT TO ADDRESS THAT CONCERN

____
(44-64) (curriculum)

13.______ PLEASE SPECIFY THE CURRICULUM SOURCE (choose one)
(65)

C= COMMERCIAL
T= TEACHER DESIGNED
L= LOCALLY DEVELOPED
P= PUBLISHED, NON-COMMERCIAL

164

14.______DO YOU KNOW OF EVIDENCE THAT SUPPORTS THE SUCCESS OF THIS CURRICULUM WITH STUDENTS IN GENERAL?
(66)
N= NO KNOWLEDGE OF ANY EVIDENCE
S= STANDARDIZED TEST SCORES FOR LARGE SAMPLE OF STUDENTS
L= PUBLISHER'S LITERATURE, SUCH AS CURRICULUM GUIDE
R= RESEARCH STUDIES IN PROFESSIONAL LITERATURE
E= LOCAL EVALUATION STUDY
P= PERSONAL EXPERIENCE
O= OTHER:(please specify)______

15.__ __HOW MANY OTHER STUDENTS IN YOUR CLASS USE THE SAME MATH CURRICULUM?
(67-68)

16. DOES THIS CURRICULUM INCLUDE: (Y/N)
______ THEORETICAL BASIS
(69)
______ OBJECTIVES STATED IN BEHAVIORAL TERMS
(70)
______ STUDENT MATERIALS (books, worksheets, etc.)
(71)
______ SUGGESTIONS FOR SPECIFIC MODIFICATIONS WHEN NEEDED
(72)
______ EVALUATION CRITERIA FOR STUDENT MASTERY
(73)
______ RECORD-KEEPING SYSTEM FOR STUDENT PROGRESS
(74)

17. IN YOUR EXPERIENCE, HAS THIS CURRICULUM BEEN EFFECTIVE WITH OTHER CHILDREN OF THE SAME:
SEX AS THIS CHILD?_____(Y/N) CULTURE AS THIS CHILD?______(Y/N)
(75) (76)

Oral Language

[staff use: card=3]
(5)

IF ORAL LANGUAGE WAS NOT DESIGNATED AS A "WEAKNESS" IN ITEM 4 - PAGE 1, PROCEED TO METHODS, PAGE 6.

18. WHEN DOES THIS CHILD EXPERIENCE DIFFICULTY IN UNDERSTANDING OR USING ORAL LANGUAGE?(please check all that apply)
______ IN LARGE GROUP INSTRUCTION
(6)
______ IN SMALL GROUP INSTRUCTION
(7)
______ IN INDIVIDUAL INSTRUCTION
(8)
______ IN PEER RELATIONS
(9)
______ OTHER (please specify)______
(10)

(CURRICULUM - ORAL LANGUAGE CONTINUED)

19. DO YOU MODIFY YOUR REGULAR INSTRUCTIONAL STRATEGIES TO ADDRESS THIS CHILD'S ORAL LANGUAGE PROBLEM? (check all that apply)

_____ NO MODIFICATIONS
(11)

_____ SIMPLIFY SENTENCES
(12)

_____ HAVE CHILD REWORD WHAT HAS BEEN HEARD
(13)

_____ INDIVIDUAL INSTRUCTION
(14)

_____ DEMONSTRATE OR MODEL
(15)

_____ PEER ASSISTANCE
(16)

_____ CORRECTIVE FEEDBACK FOR INCORRECT ORAL RESPONSES
(17)

_____ OTHER (please specify)_____
(18)

20. DO YOU USE A SPECIFIC CURRICULUM THAT ADDRESSES THIS CHILD'S PROBLEM IN ORAL LANGUAGE? _____(Y/N)
(19)

IF NO, PLEASE PROCEED TO METHODS SECTION, PAGE 6.

21. PLEASE SPECIFY THE CURRICULUM_____
(20-40) (curriculum)

22. _____ PLEASE SPECIFY THE CURRICULUM SOURCE
(41)

C = COMMERCIAL
T = TEACHER DESIGNED
L = LOCALLY DEVELOPED
P = PUBLISHED, NON-COMMERCIAL

23. _____ DO YOU KNOW OF EVIDENCE THAT SUPPORTS THE EFFECTIVENESS OF THIS CURRICULUM WITH STUDENTS IN GENERAL?
(42)

N = NO KNOWLEDGE OF ANY EVIDENCE
S = STANDARDIZED TEST SCORES FOR LARGE SAMPLE OF STUDENTS
L = PUBLISHER'S LITERATURE, SUCH AS CURRICULUM GUIDE
R = RESEARCH STUDIES IN PROFESSIONAL LITERATURE
E = LOCAL EVALUATION STUDY
P = PERSONAL EXPERIENCE
O = OTHER (please specify)_____

24. ___ ___ HOW MANY OTHER STUDENTS IN YOUR CLASS USE THE SAME ORAL LANGUAGE CURRICULUM?
(43-44)

25. DOES THE CURRICULUM INCLUDE THE FOLLOWING: (Y/N)

_____ A DESCRIPTION OF ITS THEORETICAL BASIS
(45)

_____ OBJECTIVES STATED IN BEHAVIORAL TERMS
(46)

_____ STUDENT MATERIALS (books, worksheets, etc.)
(47)

_____ SUGGESTIONS FOR SPECIFIC MODIFICATIONS WHEN NEEDED
(48)

_____ EVALUATION CRITERIA FOR STUDENT MASTERY
(49)

_____ RECORD - KEEPING SYSTEM FOR STUDENT PROGRESS
(50)

26. IN YOUR EXPERIENCE, HAS THIS CURRICULUM BEEN EFFECTIVE WITH OTHER CHILDREN OF THE SAME:
SEX AS THIS CHILD? _____(Y/N) (51) CULTURE AS THIS CHILD? _____(Y/N) (52)

166

Methods

IN THIS SECTION, INSTRUCTIONAL METHODS AND ARRANGEMENTS FOR EACH SUBJECT AREA (READING, MATH, LANGUAGE) WILL BE IDENTIFIED.

PLEASE INDICATE ROUTINE OR TYPICAL METHODS. WE REALIZE THAT TEACHERS MAY USE SEVERAL APPROACHES, BUT WE WISH THE RESPONSES HERE TO REFLECT METHODS MOST FREQUENTLY USED WITH THIS CHILD. COMPLETE ONLY THOSE COLUMNS THAT APPLY TO THE WEAKNESSES THAT YOU INDICATED UNDER STRENGTHS AND WEAKNESSES ON PAGE 1. (FOR EXAMPLE, IF YOU INDICATED THE CHILD'S WEAKNESSES WERE ORAL LANGUAGE AND READING, - YOU WOULD NOT COMPLETE THE COLUMN FOR MATH). WHERE THE QUESTION DOES NOT APPLY TO A SPECIFIC WEAKNESS, "NA" WILL APPEAR IN THE COLUMN.

QUESTIONS / **ANSWERS**

PHYSICAL ARRANGEMENTS

QUESTIONS	READING	MATH	LANGUAGE
1. WORK SPACE (53-54) D= DESK C= CARREL G= GROUP TABLE T= DESKS TOGETHER O= OTHER(please specify)______	☐	☐	NA
2. ASSIGNED SEATS (Y/N) (55-56)	☐	☐	NA
3. MATERIALS ACCESS (57-58) T= TEACHER CONTROLLED F= FREE ACCESS	☐	☐	NA
4. HOW MANY STUDENTS ARE INSTRUCTED IN THE SAME WAY? (59-62)	☐	☐	NA

[staff use: card = 4] (5)

MOTIVATION FOR PROGRESS

1. REWARDS USED (please check all that apply) (6-26)

	READING	MATH	LANGUAGE
TANGIBLES (stars, stickers etc.)	☐	☐	☐
PRIVILEGES	☐	☐	☐
FOOD	☐	☐	☐
COMPUTER TIME	☐	☐	☐
GRADES	☐	☐	☐
SOCIAL (praise)	☐	☐	☐
NONE	☐	☐	☐

OTHER: (please specify)______
(27-29)

QUESTIONS / **ANSWERS**

MOTIVATION FOR **PROGRESS CONTINUED**

QUESTIONS	READING	MATH	LANGUAGE
2. PROVIDING REWARDS (30-32) E= EVERYTIME A DESIRED RESPONSE OCCURS OR ON A PRE-SET SCHEDULE V= VARIES S= AFTER EVERY WORK SESSION D= DAILY EVALUATION AND REWARD W= WEEKLY EVALUATION AND REWARD N= NO REINFORCERS OR REWARDS GIVEN O= OTHER (please specify)______	☐	☐	☐
3. HOW MANY OTHER STUDENTS RECEIVE THE SAME TYPE OF REWARD SYSTEM? (33-38)	☐	☐	☐

PROGRESS RECORDS

QUESTIONS	READING	MATH	LANGUAGE
1. PROCEDURES FOR RECORDING PROGRESS (39-41) D= DISPLAY(chart etc.) T= TEACHER'S GRADE-BOOK S= STUDENT'S INDIVIDUAL RECORD C= COMPUTER RECORD O= OTHER (please specify)______	☐	☐	☐

167

QUESTIONS	ANSWERS		
PROGRESS RECORDS CONTINUED	READING	MATH	LANGUAGE
2. FREQUENCY OF RECORDING PROGRESS (42-44) D= DAILY W= WEEKLY T= 3-4 TIMES WEEKLY P= GRADING PERIOD			
3. ARE INSTRUCTIONAL DECISIONS BASED ON PROGRESS DATA RECORDS? (Y/N) (45-47)			
4. ARE PROGRESS RECORDS SHARED WITH PARENTS REGULARLY? (Y/N) (48-50)			
5. HOW MANY OTHER STUDENTS' PROGRESS RECORDS ARE KEPT IN THE SAME WAY? (51-56)			

HOMEWORK

QUESTIONS	READING	MATH	LANGUAGE
1. FREQUENCY HOMEWORK IS SENT HOME WITH STUDENT? (# NIGHTS PER WEEK) (57-58)			NA
2. LENGTH? (# OF MINUTES TO COMPLETE EACH ASSIGNMENT) (59-62)			NA
3. METHOD OF EVALUATION MOST FREQUENTLY USED? (63-64) S = SELF CORRECTION T = TEACHER CORRECTION P = PEER CORRECTION H = PARENT CORRECTION O = OTHER (please specify)________________			NA

QUESTIONS	ANSWERS		
HOMEWORK CONTINUED	READING	MATH	LANGUAGE
4. ELAPSED TIME BETWEEN HOMEWORK BROUGHT IN AND RETURNED WITH CORRECTIONS? (65-66) S = SAME DAY O = ONE TO TWO DAYS W = WEEKLY N = NOT RETURNED TO CHILD, FOR TEACHER'S INFORMATION			NA
5. HOW MANY OTHER STUDENTS RECEIVE APPROXIMATELY THE SAME TYPE AND AMOUNT OF HOMEWORK? (67-70)			NA

[staff use: card = 5]
(5)

FINAL EVALUATION OF INSTRUCTIONAL UNITS

QUESTIONS	READING	MATH	LANGUAGE
1. FORMAT (6-7) T = TEST G = GOAL ATTAINMENT CHECK OFF BASED ON PROGRESS DATA I = INFORMAL S = SELF-ASSESSMENT O = OTHER (please specify)________________			NA
2. TIMING OF FINAL EVALUATION (8-9) R = REGULARLY SCHEDULED S = SELF PACED N = AS NEEDED O = OTHER(please specify)________________			NA
3. RESULTS OF EVALUATION TO:			
PARENT? (Y/N) (10-11)			NA
CHILD?(Y/N) (12-13)			NA
CLASS?(Y/N) (14-15)			NA
4. HOW MANY STUDENTS ARE GIVEN UNIT EVALUATIONS IN THE SAME WAY? (16-19)			NA

168

QUESTIONS	ANSWERS		
CORRECTIVE FEEDBACK			
	READING	MATH	LANGUAGE
1. INCORRECT ORAL RESPONSES:			
ARE IMMEDIATELY CORRECTED BY TEACHER? (Y/N) (20-22)			
PROMPTS GIVEN FOR SELF-CORRECTION (Y/N) (23-25)			
DEMONSTRATION OF CORRECT RESPONSES BY PEERS(Y/N) (26-28)			
2. INCORRECT WRITTEN RESPONSES:			
SELF-CORRECTION (answer key) (Y/N) (29-30)			NA
TEACHER CORRECTS AND RETURNS(Y/N) (31-32)			NA
PEER CORRECTION(Y/N) (33-34)			NA
OPPORTUNITY TO RESUBMIT(Y/N) (35-36)			NA
3. HOW MANY STUDENTS RECEIVE CORRECTIVE FEEDBACK THE SAME WAY? (37-42)			
COMMUNICATION OF EXPECTATIONS			
1. EXPLICIT SHARING AND DISCUSSING WITH STUDENTS OF:			
ANNUAL GOALS (Y/N) (43-45)			
WEEKLY GOALS?(Y/N) (46-48)			
DAILY LESSON OBJECTIVES? (Y/N) (49-51)			

QUESTIONS	ANSWERS		
COMMUNICATION OF EXPECTATIONS CONTINUED			
	READING	MATH	LANGUAGE
2. TO HOW MANY STUDENTS ARE EXPECTATIONS COMMUNICATED IN THE SAME WAY? (52-57)			
SOCIAL ENVIRONMENT FOR INSTRUCTIONAL PURPOSES			
1. WHAT IS THE TYPICAL GROUP SIZE FOR THIS CHILD (e.g. "03";"12") UNDER THE FOLLOWING CIRCUMSTANCES:			
ACQUISITION OF NEW MATERIAL (58-63)			
FLUENCY(drill and practice) (64-69)			
GENERALIZATION(new application) (70-75)			
2. ROLE MODELS FOR CORRECT RESPONDING PROVIDED BY:	[staff use: card=6] (5)		
TEACHER? (Y/N) (6-8)			
PEERS? (Y/N) (9-11)			
MEDIA?(film,etc)(Y/N) (12-14)			

QUESTIONS	ANSWERS		
SOCIAL ENVIRNOMENT CONTINUED			
	READING	MATH	LANGUAGE
3. ACCESSIBLE HELPERS FOR THIS CHILD (other than teacher:)			
PEERS?(Y/N) (15-17)	☐	☐	☐
ADULT VOLUNTEER?(Y/N) (18-20)	☐	☐	☐
TEACHER AID?(Y/N) (21-23)	☐	☐	☐
ANSWER KEYS?(Y/N) (24-26)	☐	☐	☐
4. COOPERATIVE LEARNING:			
HOW OFTEN IS THIS CHILD ASSIGNED TO GIVE HELP TO ANOTHER? (27-29)	☐	☐	☐

0 = NEVER
1 = ONCE A WEEK
2 = TWICE A WEEK
3 = MORE THAN TWICE A WEEK

5. COOPERATIVE TASKS:			
HOW OFTEN DOES THIS CHILD WORK WITH OTHERS TO COMPLETE A SINGLE ASSIGNMENT OR TO SOLVE A PROBLEM? (30-32)	☐	☐	☐

0 = NEVER
1 = ONCE A WEEK
2 = TWICE A WEEK
3 = MORE THAN TWICE A WEEK

6. HOW MANY STUDENTS' SOCIAL ENVIRONMENT FOR INSTRUCTION IS THE SAME AS THIS CHILD'S? (33-38)	☐	☐	☐

Behavior Management Methods

BELOW ARE LISTED A NUMBER OF METHODS USED BY TEACHERS TO MANAGE CLASSROOM BEHAVIOR. PLEASE CHECK THE BOX THAT BEST DEFINES THE FREQUENCY WITH WHICH YOU USE EACH METHOD <u>WITH THIS CHILD.</u>

THIS CHILD'S BEHAVIOR IS MANAGED BY:	USED EVERY DAY [4]	USED ALMOST EVERY DAY [3]	USED SOMETIMES (2-3 TIMES/WEEK) [2]	USED RARELY (1-2 TIMES/MONTH) [1]	NEVER USED [0]
1. GUIDED INTROSPECTION, UNDERSTANDING FEELINGS (39)	☐	☐	☐	☐	☐
2. NOTE SENT HOME (40)	☐	☐	☐	☐	☐
3. SELF-MONITORING (41)	☐	☐	☐	☐	☐
4. OVERCORRECTION TECHNIQUES (42)	☐	☐	☐	☐	☐
5. PUNISHMENT: (denial of special activity etc.) (43)	☐	☐	☐	☐	☐
6. PUBLIC WARNING SIGNALS (name on board, etc.) (44)	☐	☐	☐	☐	☐
7. CONTRACTS (45)	☐	☐	☐	☐	☐
8. REMINDERS OF RULES (46)	☐	☐	☐	☐	☐
9. GROUP DISCUSSION AND CONSENSUS ON RULES (47)	☐	☐	☐	☐	☐
10. DISCUSSION OF RATIONALE FOR APPROPRIATE BEHAVIOR (48)	☐	☐	☐	☐	☐
11. SYSTEMATIC DELIVERY OF REWARD(S) (49)	☐	☐	☐	☐	☐
12. EXTINCTION (ignoring) (50)	☐	☐	☐	☐	☐
13. SENT TO PRINCIPAL OR OTHER PROFESSIONAL (51)	☐	☐	☐	☐	☐
14. PREVENTIVE ANTICIPATION: (Suggestions for appropriate behavior <u>before</u> problem arises) (52)	☐	☐	☐	☐	☐
15. TIME OUT (removal of child from situation) (53)	☐	☐	☐	☐	☐
16. OTHER:	☐	☐	☐	☐	☐

(54) (please specify) ____________________

17. ___ ___ INDICATE HOW MANY OTHER STUDENT'S BEHAVIOR IS MANAGED THE SAME WAY.
(55-56)

Teaching Strategies

PLEASE INDICATE YOUR TEACHING STRATEGIES USING THE CODES LISTED BELOW. PLEASE COMPLETE ALL SECTIONS REGARDLESS OF THE CHILD'S STRENGTHS OR WEAKNESSES AS LISTED ON PAGE 1.

IN EACH CASE, THE FIRST STRATEGY CODE YOU LIST SHOULD BE FOR THE STRATEGY USED MOST OFTEN WITH THIS CHILD IN A GIVEN SITUATION.

AFTER THE FIRST, YOU MAY LIST UP TO THREE ADDITIONAL STRATEGIES IF YOU USE MORE THAN ONE. IF NO OTHER STRATEGIES ARE USED, LEAVE BLANKS. TWO EXAMPLES ARE GIVEN BELOW.

Code	Strategy	Code	Strategy
01	MANIPULATIVES(flash-cards, games,etc.)	09	GUIDED DISCOVERY
02	FREE PLAY, EXPLORATION	10	BLACKBOARD, CHART OR OVERHEAD
03	LECTURE/DEMONSTRATION	11	GROUP ORAL RESPONDING (choral)
04	WORKBOOK/WORKSHEET	12	ONE-TO-ONE TUTORING(adult)
05	MEDIA(computer, film, tape, etc).	13	ONE-TO-ONE TUTORING(peer)
06	INDIVIDUAL ORAL RESPONDING TO QUESTIONS	14	SPECIAL PROJECTS
07	TEXT-BOOK READING, SILENTLY OR ORALLY	15	COOPERATIVE GROUP TASK
08	GROUP DISCUSSION	16	HOMEWORK
		17	SUGGESTIONS TO PARENTS

EXAMPLES:

	MOST FREQUENT STRATEGY WITH THIS CHILD	OTHER STRATEGIES USED WITH THIS CHILD
(1) ACQUISITION OF NEW SCIENCE MATERIAL	0 3	0 5 1 6 0 7

EXAMPLE 1 SHOWS THE RESPONSES OF A TEACHER WHO USES LECTURE/DEMONSTRATION (03) MOST OFTEN, FOLLOWED BY MEDIA (05), HOMEWORK (16), AND TEXT BOOK READING (07).

	MOST FREQUENT STRATEGY WITH THIS CHILD	OTHER STRATEGIES USED WITH THIS CHILD
(2) FLUENCY IN READING (drill and practice)	0 6	0 5 __ __ __ __

EXAMPLE 2 SHOWS RESPONSES OF A TEACHER WHO USES INDIVIDUAL ORAL RESPONDING TO QUESTIONS (06) MOST OFTEN, FOLLOWED BY MEDIA (05). NO OTHER STRATEGIES ARE USED BY THIS TEACHER.

(staff use: card=7)
(5)

Reading

	MOST FREQUENT STRATEGY WITH THIS CHILD	OTHER STRATEGIES USED WITH THIS CHILD
1. ACQUISITION(new material)	__ __ (6 - 7)	__ __ __ __ __ __ (8 - 13)
2. FLUENCY(drill and practice)	__ __ (14 - 15)	__ __ __ __ __ __ (16 - 21)
3. MAINTENANCE/GENERALIZATION(review and new application)	__ __ (22 - 23)	__ __ __ __ __ __ (24 - 29)

4. HOW MANY OTHER STUDENTS RECEIVE READING INSTRUCTION IN THE SAME WAY? __ __ (30 - 31)

(TEACHING STRATEGIES CONTINUED)

Math

	MOST FREQUENT STRATEGY WITH THIS CHILD	OTHER STRATEGIES USED WITH THIS CHILD
1. ACQUISITION (new material)	__ __ (32 - 33)	__ __ __ __ __ __ (34 - 39)
2. FLUENCY (drill and practice)	__ __ (40 - 41)	__ __ __ __ __ __ (42 - 47)
3. MAINTENANCE/GENERALIZATION (review and new application)	__ __ (48 - 49)	__ __ __ __ __ __ (50 - 55)

4. HOW MANY OTHER STUDENTS RECEIVE MATH INSTRUCTION IN THE SAME WAY? __ __ (56 - 57)

Language

[staff use: card = 8]
(5)

	MOST FREQUENT STRATEGY WITH THIS CHILD	OTHER STRATEGIES USED WITH THIS CHILD
1. ACQUISITION (new material)	__ __ (6 - 7)	__ __ __ __ __ __ (8 - 13)
2. FLUENCY (drill and practice)	__ __ (14 - 15)	__ __ __ __ __ __ (16 - 21)
3. MAINTENANCE/GENERALIZATION (review and new application)	__ __ (22 - 23)	__ __ __ __ __ __ (24 - 29)

4. HOW MANY OTHER STUDENTS RECEIVE LANGUAGE INSTRUCTION IN THE SAME WAY? __ __ (30 - 31)

THANK YOU FOR PROVIDING THIS INFORMATION Project Search

Appendix B

Project SEARCH Update

PROJECT SEARCH

-UPDATE-

June 1988

David H. Cooper, Ph. D. Deborah L. Speece, Ph.D.
Principal Investigators

Department of Special Education
University of Maryland

As we complete our third year of study, we wish to share our initial results with those teachers, administrators, and other professionals who have expressed an interest in our research. This UPDATE will provide a summary of our activities and findings, based on analysis of two of three cohorts of children. Project SEARCH was designed to answer two questions (a) what are the characteristics of children that place them at risk for school failure? and (b) what types of instructional arrangements serve to reduce the risk of failure for certain children, allowing them to continue their schooling in regular education classrooms? This report provides preliminary answers to these questions.

PARTICIPANTS IN PROJECT SEARCH

The Children and Their Parents

More than 260 children have participated in the data collection effort. Over 70% of the parents who were contacted for permission to include their children in this research responded favorably. This indicated a high rate of interest on the part of parents, and we are pleased with their level of cooperation.

For our study, children who were referred by their classroom teachers to the teacher assistance team (TST or SIT) and who met several additional criteria, were considered to be at-risk for school failure. Control children were randomly selected from teachers' lists of average achieving children.

Our research staff found both kinds of children involved in the project to be extremely cooperative. Several tasks given to them were quite challenging, but testers were consistently impressed with the children's determination and patience. One of our favorite stories was of the first grader who successfully solved a multiplication problem counting by 10's to 550 on his fingers!

Data records were incomplete for some children as well as for some teachers. Consequently, the information included in this report was based on 112 first grade children and 48 teachers. These data were collected during the 1985-86 and 1986-87 school years. Overall, our sample can be described as follows:

- parents of children in the study have slightly more years of education than the national average; 87% are high school graduates;

- forty-three percent of the children in our study are black, 3% other minorities and 54% are white;

- seventy-seven percent of children in our study are boys.

The Teachers

We have been privileged to have had the cooperation of 160 teachers. Some contributed their willingness to participate, but were never called upon because the children referred from their classrooms did not meet Project SEARCH criteria. Some participated in each of three project years and completed a multitude of forms. Most of the teachers were between these two extremes and made invaluable contributions of time and effort in completing the Curriculum and Methods Questionnaire, Cooper-Farran Behavioral Rating Scales, and the Teacher Background Information Form. In addition, the teachers were most gracious in allowing us to conduct live observations of the instruction in their classrooms and letting us take children from the classroom for testing. Teachers' time was also taken by attending one or more meetings with Project SEARCH staff, nominating control children, and sending letters home to parents who had not responded to our direct mailings.

Our sample of teachers were well trained and experienced:

- 65% have at least 15 hours of education beyond the bachelor's degree;

- 87% received degrees in elementary or early childhood education;

- 11.4 years in first grade was the average level of experience.

We continue to be impressed not only with the teachers' dedication in the classroom, but also with their dedication to this research effort. The SIT or TST chair in each building provided invaluable facilitation and communication between the University data collectors and the school. Teachers-in-training at the University of Maryland, and practicing teachers who read the published reports from Project SEARCH will benefit from the high degree of professionalism of school personnel.

THE AT-RISK LEARNER

Who Is At Risk?

From measures of achievement, intelligence, language, classroom behavior ratings, and learning potential, six subtypes or clusters of children were identified. With the exception of behavior ratings completed by teachers (Cooper-Farran Behavioral Rating Scales), all measures were based on the child's performance in individual test settings with our graduate research assistants. The goal was to identify stable patterns of strengths and weaknesses across the child characteristic measures. These strengths and weaknesses were derived from the test information accumulated on both at-risk and control children.

The six clusters of children we identified were both statistically valid and educationally meaningful. The profiles are depicted on the attached graph. To interpret the graphs, remember that the PLAI (Preschool Language Assessment Instrument) represented by variables PL1-PL4 on the graph, indicated the child's ability to deal with increasingly complex language. Also, the last two points on each graph (TLP and RPT) refer to our measure of learning potential (Dynamic Assessment Tasks). TLP indicates how much "teaching" a child required to learn a difficult task and RPT represents how much a child "learned" from this teaching. Overall, any point below -1.0 is interpreted as a weakness, while a point at 0.0 represents average performance.

Reviewing the graphs, one can see that three clusters of children (1,4, and 6) appear to have more weaknesses in performance than do the other three clusters (2,3, and 5). Regarding the latter more "normal" looking clusters there still are differences across the profiles. Children in cluster 3 (only 1 at-risk child in this group) probably represent the type of profile we wish all children could exhibit: achievement commensurate with intelligence, very positive work-related and interpersonal skills, average language skills, and an ability to profit from instruction. Cluster 5 may represent our "second choice" of the type of child we wish to populate our classrooms. They do not appear to achieve as highly as expected but are in the normal range. They receive average classroom behavioral ratings and possess relatively strong language skills. Curiously, the children in this cluster did not appear to benefit from the learning potential task (refer to the relationship between TLP and RPT) and were the only group to exhibit this pattern. Cluster 2 appears to consist of children who may be the most frustrating of the 3 "normal" looking clusters. That is, these children earned the highest intelligence scores with achievement scores being somewhat lower than expected and showed a strong performance in language. However, their behavior on both dimensions of work-related and interpersonal skills received relatively low ratings. This cluster turns out to be an interesting group when we look at the classroom observation data reported in a subsequent section.

The three other clusters that exhibited the poorest performance are discussed next. Cluster 1 children exhibited a pronounced problem on teacher ratings of classroom behavior (WRK, INT) in addition to a discrepancy between achievement (RDG,MTH) and intelligence (VIQ, NIQ). A mild decline in language skills is evident (PL3, PL4), but these children benefited from the structured learning potential task (RPT). We suspect that those children may be exhibiting a pattern associated with learning disabilities. Cluster 4, on the other hand, demonstrated almost uniformly depressed scores across the measures. These children appear to be achieving at levels one would expect especially given their very poor language skills. Although work-related skills (WRK) were not rated positively, interpersonal skills (INT) were regarded as a strength by teachers. Cluster 4 may be the group sometimes referred to as "slow-learners".

Cluster 6 represents a profile we interpret as associated with language difficulties. In addition to low achievement and intelligence, these children exhibited a steep decline on the language measure as the task became more complex (PL1-PL4), and required a great deal of teaching to learn the challenging task (TLP). Interestingly, following this teaching, these children improved as much as children in Cluster 1, although the overall amount of learning was lower (RPT).

How Much Risk?

The clusters presented some interesting food for thought and we anticipate that the profiles are recognizable to our participating teachers. After the clusters were identified, we next wanted to know which ones were most at risk for referral to Teacher Assistance Teams. To determine this, we calculated a statistic called a relative risk ratio (represented on the graph as RRR). The rate of referral for each cluster was compared with the rate of referral for cluster 3, the most "normal" cluster. This procedure results in a number for each of the remaining five clusters. For example, cluster 2 has an RRR equal to 27 which is interpreted as follows: Risk of referral for the profile associated with cluster 2 is 27 times that of the profile of cluster 3. Reviewing this information one can see that relative to cluster 3:

- The profile for cluster 1 represents the highest risk of referral to SIT or TST;

- Clusters 2, 4, and 6 have the next highest risk for referral, but did not differ from each other;

- Cluster 5 with an RRR = 7, represents a significant risk compared to cluster 3, but has significantly less risk than the other clusters.

In examining the profile for cluster 1, it is tempting to conclude that teacher ratings on the CFBRS (Cooper-Farran Behavioral Rating Scales) make the difference. While this may turn out to be the case, it is important to interpret the entire pattern, not one or two variables. For example, cluster 1 also differs from cluster 3 on achievement, intelligence and to some extent, language. It may be that this pattern in conjunction with extremely poor work-related and interpersonal skills influences the decision to refer.

What's Next?

As researchers are fond of saying, these results raise more questions than they answer. With regard to our clusters, two major questions remain: (a) What is the risk of referral to and placement in <u>special education</u> associated with these profiles? and (b) What is the developmental course of children in these clusters, that is, how do they change over time? The answer to risk for special education placement will come from our follow-up phone calls and visits to determine each child's status at the end of each year (promoted, retained, placed in special education). Currently, our status data indicate that very few children are placed in special education at the end of first grade, thus the numbers at this time are too small for analysis. By tracking these children over the course of their elementary school careers, the picture will become more complete.

The developmental question will be pursued over the next two years by collecting the individual child measures for 70 at-risk children who are already identified (no new children will enter the study). We will seek approval from the participating school systems and teachers to obtain second and third year data on the appropriate children. This kind of longitudinal information is quite rare, but will provide the type of evidence needed to see which characteristics of children change over time and which profiles lead to the most positive educational outcomes.

CHILDREN'S RESPONSES TO CLASSROOM INSTRUCTION

Differences Between At-Risk and Control Children

Project SEARCH staff completed 65 hours of in-class observation on both at-risk and control first grade children. From that information we found that at-risk children were more likely to be off-task than control children. When on-task was strictly defined as "making active academic responses," such as reading, writing, answering questions, etc., the difference was quite substantial: control children were on task more often than at-risk children (50% more often). When on-task was defined as also including task-related behaviors such as attending to the teacher, raising hands, looking for materials, etc., the difference between control and risk groups remained, but was less dramatic (14% more often).

Inappropriate behaviors, those that compete with academic responses, were relatively infrequent in both groups. However, the at-risk children were observed to be inappropriate 30% of the time while control children were inappropriate 20% of the time. The two most common inappropriate responses in both groups were "looking around" and "inappropriate task" (working or playing with the wrong materials). Disruptive behaviors were rarely observed.

Increasing Academic Responding

Going beyond simple differences in rates of academic responding, we were able to determine which types of instructional arrangement s increased or decreased students' on-task responding. We characterized instruction in terms of:

- The activity (content or subject matter);
- The task (what the students were expected to do);
- The structure (how the students were grouped);
- The teacher's role (providing students the opportunity to make active academic responses).

Several conclusions are suggested by the analysis of the students' responses to various combinations of activity, task, structure and teacher role.

First, students are profoundly influenced by variations in instruction. In other words, students do not behave in one predictable way, but modify their behavior, apparently in response to their perceptions of what is happening in the classroom.

Second, both at-risk and control students' academic responding is accelerated by the same arrangements of instruction. The most potent instructional arrangements in terms of producing active, academic responses are those composed of (a) academic activities (as opposed to class business, tree play, etc.), (b) active tasks (reading, writing, as opposed to passive tasks, such as listening, discussing) and (c) small group structure (as opposed to entire group or working individually). In addition, teachers play a role in giving students ample time to make active academic responses. Such opportunities are possible while the teacher listens, monitors, or teaches another group of children.

Third, the instructional arrangements that promote academic responding appear to be less frequently available to the at-risk students. We do not yet have a satisfactory explanation for this. It may be that these students' behaviors are just as capable of changing their instructional environment as the environment is of changing their behavior. In other words, the characteristics that place these children at-risk may have an effect on the ways in which their instruction is provided.

Cluster Differences In Classroom Behavior

The clusters of children (described above) were also found to differ on their rates of academic and non-academic responding. Cluster 1, the cluster most at risk for referral, was also the one with: (a) the lowest rate of academic responding and (b) the second highest rate of inappropriate responses among the six clusters.

Cluster 6, the lowest achieving subtype, was also the group that took the longest time preparing to be on task, that is, looking for materials and moving to new work stations. These results, coupled with this group's poor language skills, may account for their deficits in reading and math performance.

Cluster 2, referred to earlier as possibly the most frustrating, is also the group with the greatest potential to achieve. This group is strong on all measures of achievement, intelligence, and language, and according to the observation data, is the group most likely to make active academic responses (29.8% of the time). But, these children also exhibit the highest rate of inappropriate behaviors (29%). Because they are among the highest achievers, we suspect that this cluster includes children who catch on quickly, finish their work before anyone else, and then behave inappropriately.

What's Next?

In the coming months we will analyze observational data on an additional 80 children. This information represents an additional 40 hours of observational information and was collected during the 1987-88 school year. We will look closely at the types of instruction that promote academic responding, and reduce the likelihood of referral to special education.

Also, the data obtained from the Curriculum and Methods Questionnaire (CMQ) will be analyzed together with the classroom observation data. The very specific data provided by teachers on the CMQ, (instructional strategies and behavior management) will be critical to our determination of effective instruction with high-risk students. We recognize that the CMQ placed the most demand on teachers' time, but it is also likely to produce some of the most useful results.

Future Reports

We will continue to provide Project SEARCH UPDATES to keep you informed of our activities and findings. We plan to provide you with at least two reports in the next academic year and anticipate that the following information will be shared during 1988-89:

1. Dynamic Assessment Tasks (learning potential measure)

This experimental measure was designed by our staff and shows promise of providing information not usually available from traditional standardized measures. During the summer (1988) we plan to analyze the strengths and weaknesses of this task in relation to intelligence and achievement. Julie Kibler, one of our graduate research assistants, is preparing a masters thesis on this topic and is taking a lead role in the descriptive phase of this analysis.

2. Risk for Referral to Special Education

As described earlier, the numbers of children placed in special education at the end of first grade were too small to derive a meaningful risk ratio. However, we will be able to analyze the outcomes for children at the end of their second year and subsequent years of schooling. These children entered the study during 1985-86 and 1986-87.

3. Precision of Risk Estimates

This year we identified a new cohort of first graders whose data are currently being collected and will be prepared for analysis this summer. The purpose of including more children was to increase the number of children in each cluster which in turn increases the precision of the risk estimates presented on the graphs. At present, we know that five clusters have a significant degree of risk for referral compared with cluster 3. It is difficult to analyze differences among these risk ratios because of the small samples. The addition of this year's new children will allow these comparisons. These results will be available Spring, 1989.

4. Clusters and Classroom Environments

One of the major questions guiding Project SEARCH is assessing the effect of classroom environments on clusters defined by child characteristics. We want to know which types of environments serve to maintain which children in the regular education classrooms. We anticipate these major findings to be available by Fall of 1989.

FOR FURTHER DETAIL

Over the last two years, we have written several articles and conference papers to share our perspective and findings with the academic community. Below is a list of our work that will provide more detail on the results in this UPDATE. Please call or write us if you would like to receive any of these reports.

Cooper, D.H. & Farran, D.C. (1988) Cooper-Farran Behavioral Rating Scales: Technical Manual. Available from D.H. Cooper, Department of Special Education, University of Maryland, College Park, MD. 20742.

> This report details the technical data supporting the psychometric adequacy of the CFBRS, the instrument completed by Project SEARCH teachers.
>
> Over 1600 children in Maryland, North Carolina and Hawaii have been rated on the CFBRS. It appears that this instrument assesses behaviors critical to success in the primary grades. Professionals who wish to use this instrument are invited to contact Dr. Cooper at the above address.

Cooper, D.H. & Speece, D.L. (in press) A novel methodology for the study of children at-risk for school failure. Journal of Special Education.

> This paper is a revised version of our 1987 presentation to the American Educational Research Association. The details of the research methodology guiding Project SEARCH are presented with illustrations of planned analyses. The relative risk analyses are highlighted. The paper will be published in the Summer, 1988 issue, Vol. 22 (2).

Speece, D.L. (in press) Methodological issues in cluster analysis: How clusters become real. To appear in H.L. Swanson and B.K. Keogh (Eds.) Learning disabilities: Theoretical and Research Issues. Hillsdale, N.J.: Laurence Erlbaum Associates.

> This paper summarizes the issues associated with applying cluster analysis techniques from both statistical and conceptual perspectives. A framework is developed to assist investigators in making the many decisions required for correct application of the techniques. This framework guided the cluster analysis procedures

used in Project SEARCH to obtain the six subtypes described in this UPDATE.

Speece, D.L. & Cooper, D.H. (1988, April) The academic responses to learning environments by children at risk for school failure. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

This report elaborates on several of the results provided in this UPDATE by providing an overview of the statistical analysis and results of the child characteristic and learning environment measures. We are currently revising this manuscript to highlight the subtyping results for presentation at a learning disabilities conference at the Pennsylvania State University at the end of May, 1988.

YOUR TURN

We invite your comments, concerns, insights, and, yes, criticisms of this report specifically, and Project SEARCH generally. Your feedback will guide our future reporting, so, if you find the material stimulating (or incomprehensible) please let us know.

A FINAL WORD

In addition to thanking school personnel in Prince George's and Anne Arundel counties, we would be remiss if we did not acknowledge the efforts of our research team who provided the energy for this project over three years. We wish to thank the following individuals:

Lisa Pericola Case	Clare Kelly
Hillary Schwartz	Julie Kibler
Barbra Sto-ward	Bruce Sharkin
Susan Div-ia	Lynn Rogers
Virginia Noble	Christina LeForge
Diane Hunter	Barbara Lockemeyer
Caren Appel	Barbara O'Malley
Patty Hallinan	Beverly Strong

This research was supported by a grant from the U.S. Department of Education, Office of Special Education and Rehabilitative Services (G008530155).

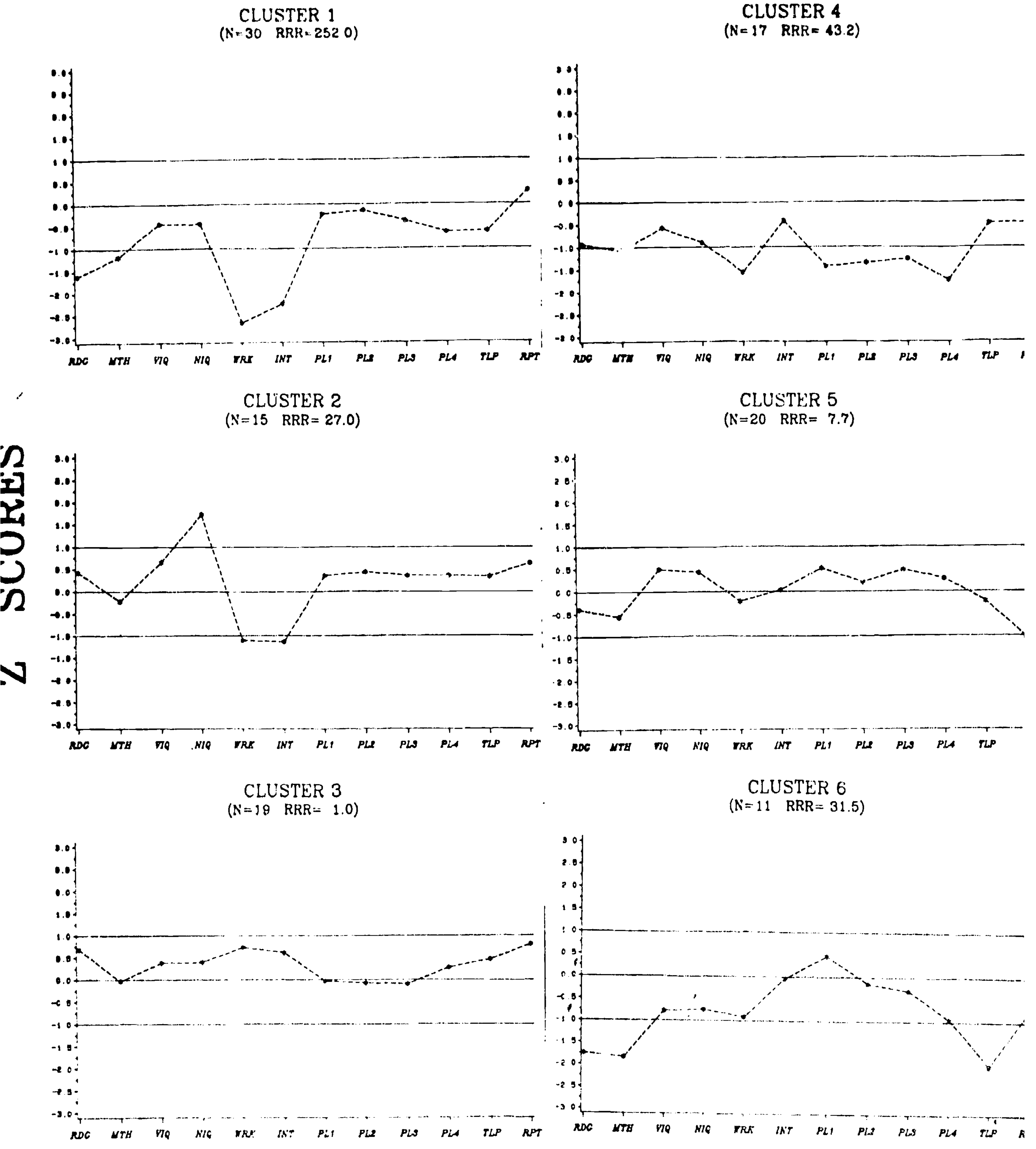


Figure 1. Mean profile and relative risk ratios (RRR) of clusters. (RDG=reading; MTH=mat VIQ=verbal IQ; NIQ=no verbal IQ; WR ating of work related skills; INT=rating of interpersonal skills; PL1-PL4=level scores on PLAI; TLP=total prompts on Dynamic Assessment; RPT=residual posttest gain score on Dynamic Assessment).

Appendix C

Dissemination: Papers and Presentations

Papers

Speece, D. L. (in press). Methodological issues in cluster analysis: How clusters become real. In H. L. Swanson and B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues. Hillsdale, NJ: Lawrence Erlbaum.

Cooper, D. H., & Speece, D. L. (1988). A novel methodology for the study of children at risk for school failure. Journal of Special Education, 22, 186-197.

Cooper, D. H., & Speece, D. L. Instructional correlates of students' academic responses: Comparisons between at-risk and control students. Manuscript in review, 1989.

Speece, D. L., & Cooper, D. H. Ontogeny of school failure: Classification of first grade children. Manuscript in review, 1988.

Speece D. L., Cooper, D. H., & Kibler, J. M. Dynamic assessment, individual differences and academic achievement. Manuscript in review, 1989.

Presentations

Cooper, D. H., & Speece, D. L. (1989, March). Instructional correlates of students' academic responses: Comparisons between at-risk and control students. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Speece, D. L. & Cooper, D. H. (1989, March). Strategy instruction: Who needs it? Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Speece, D. L. & Cooper D. H. (1989, March). Children at risk for school failure: Identification, replication, and validation of empirical subtypes. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Speece, D. L., & Cooper, D. H. (1988, April). The academic responses to learning environments by children at risk for school failure. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Speece, D. L. (1988, Feb.). Methodological issues in cluster analysis: How clusters become real. Paper presented at the International Academy of Research in Learning Disabilities Research Meeting, Los Angeles, CA.

Cooper, D. H., & Speece, D. L. (1987, April). Characteristics of primary grade children at risk for school failure. Paper presented at the Annual Meeting of the American Educational Research Associaiton, Washington, D.C.

Speece, D. L., & Cooper, D. H. (1989, June). What dynamic assessment may add to the learning disabilities equation: Data on children at risk. Paper to be presented at the 1989 Joint Conference on Learning Disabilities, Ann Arbor, MI.

Cooper, D. H., & Speece, D. L. (1989, March). A preview of the mildly handicapped child: What the referring teacher sees. Paper presened at the Gatlinberg Conference on Research in Mental Retardation and Developmental Disabilities, Gatlinburg, TN.

Speece, D. L., & Cooper, D. H. (1988, June). Subtype membership and the risk for referral in the primary grades. Paper presented at the Annual Conference on Research and Theory in Learning Disabilities, The Pennsylvania State University, State College, PA.