DOCUMENT RESUME

ED 301 007 EC 211 754

AUTHOR Ysseldyke, James E.; And Others
TITLE Examination of Categorical Practice in Special Education: Is It Supported by Research?
INSTITUTION Minnesota Univ., Minneapolis. Dept. of Educational Psychology.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
PUB DATE 6 Apr 88
NOTE 40p.; Paper presented at the Annual Conference of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; *Educational Environment; Elementary Education; *Emotional Disturbances; Instructional Effectiveness; *Learning Disabilities; Mainstreaming; *Mild Disabilities; *Mild Mental Retardation; Special Education; Student Reaction; *Teaching Methods
IDENTIFIERS *Categorical Special Education

ABSTRACT

This study examined the extent to which categorical differences exist in the instructional environment for students with mild handicaps. Subjects were 30 learning-disabled, 32 emotionally/behaviorally disturbed, 30 educable mentally retarded, and 30 nonhandicapped students in grades 2-4. Twenty-four special education teachers and 54 regular education teachers also participated. The quantitative nature of instruction was measured by gathering data on instructional ecology and student responses. The qualitative nature of instruction was measured using an integrative rating system based on observations, teacher interviews, and student interviews. Few differences were found in how the handicapped groups of students were instructed. More often, differences found were between handicapped and nonhandicapped students, or between regular education and special education settings. There were no differences over the entire school day in the tasks handicapped students were assigned or in the nature of their responses. Within special education settings, educable mentally retarded students' instruction was rated lower than other groups on instructional presentation, checking for student understanding, and task relevance. Considerable variability was found within categories on both quantitative and qualitative aspects of instruction, in both regular and special education. (JDD)

Examination of Categorical Practice in Special Education: Is It Supported by Research?

James E. Ysseldyke, Martha L. Thurlow, and Sandra L. Christenson

Department of Educational Psychology

University of Minnesota

Presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 6, 1988.

The research reported in this presentation was supported by a grant from the U.S. Department of Education, Office of Special Education and Rehabilitative Services (OSERS). Points of view or opinions do not necessarily represent official position of OSERS.

IAP JA5 (QP)

Abstract

The extent to which categorical differences exist in the quantitative and qualitative nature of the instructional environment was examined. Observations were conducted in both regular and special education settings using the quantitative Code for Instructional Structure and Student Academic Response and the qualitative Instructional Environment Scale. Subjects were 30 LD, 32 EBD, 30 MR, and 30 nonhandicapped students in grades 2-4. Few differences were found among groups of students with handicaps. More often, differences found were between handicapped and nonhandicapped students, or between regular education and special education settings. Implications for interventions and effective instruction are discussed.

Examination of Categorical Practice in Special Education: Is It Supported by Research?

Few, if any, educators would disagree with the statement that there are handicapped students in schools. However, which students are mildly handicapped and whether these students should be subcategorized as learning disabled, emotionally/behaviorally disturbed, or educable mentally retarded is a controversial issue.

The provision of services to special education students through Public Law 94-142 reflects two underlying assumptions. First, it is assumed that special education instruction is necessary to optimize instruction. Thus, current practices of referral and placement proceed in the belief that certain students have instructional needs that can be better met through special programming than through standard educational practices provided in mainstream classes. Second, it is assumed that the delivery of special education services should be based on a categorical model. Thus, specific types of learning disorders are diagnosed so that different instructional events and procedures can be implemented to optimize school programs, which will be tailored to match the needs of the students' identified disabilities. The extent to which these assumptions are valid is quickly becoming a major topic of controversy in special education. In large part, the controversy emerged from the questioning of current categorical practices on the grounds that they do not accurately differentiate students and are not treatment specific (cf. Reynolds, Wang, & Walberg, 1987; Stainback & Stainback, 1984; Will, 1986).

The amount of time spent in identifying students as handicapped and declaring students eligible for specific special education

placements is high. An important issue for practitioners is whether the time spent in specific categorical determination for handicapped students leads to changes in the students' instructional environments.

The extent to which categorical practice in special education is supported by research findings has not yet received adequate attention. As part of a continuing research program, we have examined whether categorical differences exist in the quantitative and qualitative nature of instruction for students with mild handicaps, specifically those labeled as learning disabled (LD), emotionally/ behaviorally distrubed (EBD), and educable mentally retarded (EMR). In addition, we have been developing an evolving literature review focusing on the comparison of student characteristics or instructional practices for students experiencing learning and/or behavioral problems.

To date, our literature review has included 33 studies, published in the past three years, that examined categorical differences in one or more of the following areas: (a) academic performance, (b) behavioral characteristics, (c) medical problems, (d) social-emotional characteristics, (e) teacher and instructional characteristics, and (f) post-school outcomes. Of these studies, however, 12 made "categorical" comparisons between one group with handicaps and a second group without handicaps (but often called low achieving). When these studies were eliminated, we had 21 to examine.

Our review of these 21 studies has produced equivocal findings, although the majority of studies seem to provide some support for

categorical differences. However, it is important to examine the nature of the differences being identified. Table 1 is a summary of the results of a quick review of the 21 studies. In this review, we looked at the extent to which there was support of categorical differences (i.e., a significant difference was reported), questionable support (a possible difference was reported, but not reported as significant), or no support (no significant differences were reported). The most frequent areas where differences were found include behavioral characteristics and teacher/instructional characteristics.

Insert Table 1 about here

In comparisons of behavioral characteristics, we found that four of five studies reported differences. For example, behaviorally disordered students are found to have greater adjustment problems than learning disabled students who, in turn, have greater adjustment problems than educable mentally retarded students (Cullinan & Epstein, 1985). Learning disabled students receive higher adaptive behavior scores than mentally retarded students (Harrison, 1987). Teachers rate emotionally/behaviorally disordered students as having greater hyperactivity than learning disabled and educable mentally retarded students (Epstein, Cullinan, & Gadow, 1986). And, teachers estimate that behaviorally disordered students are more victimized by violent crime than are learning disabled students (Lang & Kahn, 1986). Thus,

in the area of behavioral characteristics, the studies finding categorical differences generally focus on ratings provided by adults in the students' school environments. Such ratings, of course, are highly susceptible to preconceived notions about categorical differences.

In the area of teacher/instructional characteristics, we found that five of eight studies reported categorical differences. Some studies finding differences among categories have focused on program placements, such as the amount of time in mainstream settings (Travis, Thomas & Fuller, 1985) or the nature of the special education setting in which students are placed (Singer, Butler, Palfrey, & Walker, 1986). Teachers' expectations for student achievement also have been found to differ, with higher expectations expressed for students with learning disabilities in comparison to expectations for students with mental retardation (Rolison & Medway, 1985). And, teachers' views of their inservice needs are different for teachers of behaviorally disordered students compared to teachers of learning disabled and mentally retarded students in the areas of assessing pupil behavior, curriculum design and use, and professional information. Thus, four of the five studies finding categorical differences are based on either placement data or teacher perceptions. The other study we found in this area has identified differences in the use of instructional time, including greater proportions of time for EMR students in teacher-directed questioning instruction and smaller proportions of time in directed reading, compared to LD students

(Sindelar, Smith, Harriman, Hale, & Wilson, 1986). The study can be contrasted to other studies that have found no differences when they looked at the variety of instructional strategies used, the effective structuring of student time, effective questioning, provision of feedback, effective management techniques, and active student involvement (Algozzine, Morsink, & Algozzine, 1986), or differences in student achievement (Marston, 1987) and in student responses to instruction (O'Sullivan, Marston, & Magnusson, 1987), as a function of the match of teacher licensure with student label. In large part, then, studies in the area of teacher/instructional characteristics have concentrated on placement differences rather than differences within placements.

The purpose of this paper is to report our own findings on the extent to which categorical differences exist in the instructional environment (specifically, the quantitative and qualitative nature of instruction) for students with mild handicaps who have been labeled as learning disabled, emotionally/behaviorally disturbed, and mentally retarded. The quantitative nature of instruction was measured using a 10-second interval observation system that gathered data on instructional ecology (e.g. instructional tasks), and student responses (academic responding, academic engaged time, management, inappropriate behavior). The qualitative nature of instruction was measured using an integrative rating system based on observations, teacher interviews, and student interviews. In this system, global integrative judgments about an individual students' instructional

environment were made along the dimensions of instructional planning, instructional presentation, checking for student understanding, task relevance, practice, and feedback. Two questions were of primary interest:

1. To what extent are there differences among categories of mildly handicapped students in the quantitative nature of instruction?
2. To what extent are there differences among categories of mildly handicapped students in the qualitative nature of instruction?

Since most students were observed in both regular education and special education settings, we also were interested in examining differences in the quantitative and qualitative nature of instruction as a function of setting. Data also were collected on a comparison group of nonhandicapped "average" students in the regular education setting.

Method

Subjects

Subjects were 92 school-identified handicapped students (30 LD, 32 EBD, 30 EMR) and 30 nonhandicapped (NH) students in grades 2-4 from urban (57%, 50 handicapped, 20 nonhandicapped) and suburban (43%; 42 handicapped, 10 nonhandicapped) school districts. Nonhandicapped students were considered average; they received no extra services, such as Chapter I or High Potential.

Students' classifications (LD, EBD, EMR) were verified by special education teachers. LD students had been identified by a two-year

deficit criterion or an ability-achievement discrepancy criterion. EBD and EMR students were identified using behavioral descriptors. EBD students had chronic task incompletion problems, acting out, behavior difficulties, or social interaction difficulties. These characteristics were severe enough to impede academic performance. EMR students were functionally academically retarded in the four basic skill areas, and most received their basic skill instruction within special education settings. In instances where students carried two labels (i.e., LD and EBD), students were selected on the basis of primary classification.

The four groups of students were distributed evenly across grades 2-4 and were similar in sex and racial characteristics. They ranged in age from 91 to 146 months, with the handicapped students being slightly older; the average age for LD students was 113 months (range = 91-136). The average age for EBD students was 115 months (range = 97-137); the average age for EMR students was 119 months (range = 99-146); and the average age for nonhandicapped students was 109 months (range = 91-128).

Participating teachers included 24 special education teachers and 54 regular education teachers; teacher information was available for 51 regular and 24 special education teachers. The mean number years teaching experience was 16.6 years (range = 1-31), Most teachers were female (n = 66; 88%); nine teachers (12%) were male. The majority of teachers held bachelor degrees plus additional credits (n = 32; 42.7%) or master's degrees (n = 24; 32.0%). Nine teachers (12%) held a bachelor's degree only, while 10 teachers (13.3%) held a master's degree plus additional credits.

Teachers and students were volunteer participants in the study. Students were randomly selected within category with two restrictions: (a) parent permission for student participation had been obtained, and (b) regular education teachers would have no more than two students and special education teachers would have no more than four students involved in the study unless willing to do so.

Observation Systems

Two observation systems were employed. The CISSAR (Code for Instructional Structure and Student Academic Response) observation system developed by Greenwood, Delquadri, and Hall (1978) was used to collect information on the quantitative nature of instructional time. This system focuses on the observation of one target student (rather than sampling behaviors of several students). Two event areas coded in the system were of interest: instructional task (8 codes) and student response (7 academic, 6 inappropriate, and 5 management). Table 2 is a summary of the specific events recorded within each area.

Insert Table 2 about here

A momentary 10-second time sampling technique was used to direct the recording of events. An auditory electronic timer attached to a clipboard was used to signal the 10-second intervals. The timer was equipped with an earplug so that only the observer could hear the signal (a short beep sound).

The Instructional Environment Scale (IES) (Ysseldyke, Christenson, McVicar, Bakewell, & Thurlow, 1986) was used to obtain information

about the qualitative nature of instruction for an individual student. After completing classroom observations and interviews with teacher and student, the observer rated 40 items about principles of effective instruction. These were rated on a 4-point Likert-type scale indicating the degree to which the statement was characteristic of the target student's instruction, from "very much like the student's instruction" (4) and "like the student's instruction" (3), to "not much like the student's instruction" (2), and "not at all like the student's instruction" (1). Ratings of NA (not applicable) for 5 items and ? (can't tell) also were possible in clearly specified situations (Ysseldyke et al., 1986). The teacher interview included 20 questions about those areas more difficult to understand through observation only (e.g., instructional planning decisions); the interview required, on the average, 30 minutes to complete (range = 20 to 45 min). The student interview included eight questions about the student's perceptions of the tasks assigned and also provided data on the student's success rate for the assignments during the second day of observations. Four questions were accompanied by cartoon-like pictures to aid the students' understanding of the required 3 or 4-point ratings.

Observers

Four CISSAR and six IES obervers collected data. All were females, selected from a pool of 100 male and female applicants who had responded to an ad in a local newspaper. All but two observers had a child or children in elementary or secondary schools. Observers did not work in schools in which their children were enrolled.

Procedures

Observer training CISSAR training used the Observer and Trainer's Manual developed by Stanley and Greenwood (1980). Two weeks of half-day training sessions were required to cover the material presented in the manual. This was followed by two to three days of additional practice observing within actual classrooms.

IES training required observers to read training manual materials (Ysseldyke et al., 1986), learn definitions for each item, and integrate multiple sources of information to form one rating for each item. Discussion of rating considerations and practice rating items through the use of written examples, videotapes, and classroom practice was used extensively. A major focus of training was to learn to describe rather than evaluate or judge instruction.

Data collection. Each student was observed for one entire school day by one CISSAR observer. Observations were not conducted during breaks, such as those for lunch, recess and bathroom. Also, observers did not code during physical education, music, or special assembly programs since the observation system did not apply to these situations. Typically, observers did not code continuously for a period of more than two hours because of natural breaks within the school day. Observers did follow target students when they left their homerooms to go to other classrooms for other subjects, or when they went to the resource teacher for special instruction. Coding was conducted in these other classrooms in the same manner as in homerooms.

IES ratings for each student were based on two observations in a setting plus interviews with teachers and student. A total of 184 IES ratings was completed: 92 in regular education classrooms and 92 in special education classes. LD and EBD students were observed in both regular and special education settings. Regular education students were observed in their regular class and EMR students were observed only in their special education class. LD, EBD, and 10 EMR students were observed in resource rooms. The remaining EMR students (N = 20) were observed in one of five self-contained classrooms. The average length of individual regular mainstream class observations was 48 minutes (range = 18-118 min). The average length of individual special education class observations was 44 minutes (range = 10-90 min). Most observations in regular education settings were during reading or math.

Generally, CISSAR observations were completed first, followed by IES observations within a one-week period. The preferred data collection sequence for IES ratings was completion of two classroom observations, student interview, and teacher interview. IES items were rated only after all were completed. The student interview was conducted as close to completion of the observation as possible to reduce student forgetting about the observed lesson. Variations from this sequence were due to classroom scheduling difficulties and teacher preferences. In most cases, the teacher and student interviews were conducted on the day of the second IES observation. In some cases, to meet teacher schedules, teacher interviews were conducted before school on the day after the second observation.

The student's name was revealed to the teacher at the time of scheduling, thus possibly biasing the results in a positive direction. Teachers were told that we were interested in how students respond to instruction and were asked to respond as they typically would. Teachers introduced the data collectors by explaining they were here to see what second, third, or fourth graders do in school. Although the observers were not told the student's classification, it was impossible to keep them blind about the handicapped vs. nonhandicapped distinction because of the manner in which services were delivered (i.e., special education in a separate, smaller room).

Inter-rater Agreement

Checks for inter-rater agreement on the CISSAR system were conducted 15 times, for 15 minutes each, during the study. Average agreement was 98.1% for task codes and 95.2% for student responding codes. The desired agreement level was 90%. IES checks for inter-rater agreement were conducted 18 times during the study. Inter-rater agreement for IES was computed in two ways: (1) Grouped agreement, in which ratings of 1 and 2, and ratings of 3 and 4 were combined and instances of agreement counted (e.g., rating of 1 by one observer and rating of 2 by other was counted as agreement), and (2) Exact agreement, which involved counting only cases where ratings were exactly the same. Since IES is a qualitative rating scale involving global, integrative judgments about a complex area, desired agreement was 50% for exact agreement and 75% for grouped agreement. Average exact agreement was 60% (range = 35-85%) and average grouped agreement was 84% (range = 70-95%).

To maintain adequate levels of inter-rater agreement throughout the study, observers discussed their areas of disagreement after each inter-rater agreement check, on the same day. In addition, semi-monthly meetings were held to discuss coding problems and disagreements.

Data Analysis

CISSAR data were converted to total minutes spent in each coded event. Further, when comparisons were made in different settings, data were converted to proportions because of unequal times the students spent in different settings. Several post hoc composites were formed for comparison purposes (see Table 2). One-way analyses of variance were used to test differences in the amounts of time handicapped (LB/EBD/EMR) and nonhandicapped students spent in different tasks and student responses over the entire school day. Repeated measures ANOVAs were used to make comparisons across settings and categories. In these analyses, EMR students were divided into EMR-resource and EMR-self-contained groups because of differences in the settings where they received special education instruction.

IES data were organized into six instructional factors on a theoretical and conceptual basis, rather than an empircal basis, and an adaptation was made to handle the NA and ? ratings. The six instructional factors used in data analysis were: Instructional Planning, Instructional Presentation, Checking for Student Understanding, Task Relevance, Practice, and Feedback. Brief descriptions are provided in Table 3.

Insert Table 3 about here

Ratings ? and NA were counted as missing items. If the total number of missing items was more than five, the case was dropped from the analysis. If the total number of missing items was five or less, the mean for the cluster scale was assigned for each missing item. There was no instance where more than one missing item occurred in an instructional cluster. Separate analyses were conducted on IES ratings in each setting (regular, special). Repeated measures ANOVAs were used to make comparisons across settings and categories.

Results

Quantitative Nature of Instruction

Entire school day comparisons. Table 4 is a summary of the number of minutes students were observed in different composite tasks and responses during an entire school day. Of the nine comparisons that were made, none emerged as significant. Thus, no differences were found in the types of tasks used by the different categories of students, nor in the nature of the students' responses to instruction. Considerable variability was evident in times for all categories. In fact, the ranges were quite similar in their spans and end points (see Table 4).

Insert Table 4 about here

Category X setting comparisons. Students in different categories spent different amounts of time in regular education and special education settings. Therefore, proportions of time were analyzed in categorical comparisons across settings. In addition, because the special education setting was different for some EMR students, analyses separated those served in resource settings (EMR-R) from those served in self-contained settings (EMR-S). NH students were not included in these analyses since they did not spend time in the special education setting. Only those handicapped students actually observed in both settings were included in analyses (LD=28, EBD=21, EMR-R= 10, EMR-S=5). Table 5 is a summary of the proportion of time in each setting that students were involved in various composite tasks and responses. Of the nine categorical comparisons made, one emerged as significant. Follow-up Student-Newman-Keuls tests indicated that EMR students in self-contained settings spent more time making management responses than all other groups. In addition, there were four category by setting interactions, out of a possible nine. The significant category by setting interactions were: (a) Textbooks: F (3,60) = 3.93, p <.01, LD students used textbooks a greater proportion of the time in regular education settings than did EMR resource students, no differences in special education settings; (b) Other media: F (3,60) = 5.45, p <.01, EMR self-contained students spent more time using media in regular education than all other groups, no differences in special education settings; (c) Fetch/Put Away: F (3,60) = 6.91, p <.001, EMR self-contained students spent more

time on fetch/put away tasks during special education than students in all other groups, no differences in regular education settings; and (d) Inappropriate responses: F (3,60) = 4.05, $p < .01$, EMR self contained students spent less time making inappropriate responses than all other groups in regular education settings, no differences in special education. In summary, all significant differences except one (textbooks) involve EMR students served in self-contained settings.

Insert Table 5 about here

Of the nine setting comparisons, five were significant: (a) Paper tasks: F (1,60) = 7.32, $p < .01$, greater proportion of time spent on paper tasks in special education, (b) Active academic responses: F (1,60) = 67.99, $p < .0001$, greater proportion of time in active academic responses in special education; (c) Academic engaged time: F (1,60) = 71.45, $p < .0001$, greater proportion of time in academic engaged time in special education; (d) Inappropriate student responses: F (1,60) = 28.17, $p < .0001$, greater proportion of time in inappropriate responses in regular education; and (e) Management responses: F (1,60) = 31.00, $p < .0001$, greater proportion of time in management responses in regular education. In summary, setting effects favored special education in terms of students' academic responses.

Qualitative Nature of Instruction

Regular education. One-way analyses of variance were used to compare ratings of the qualitative nature of instruction for LD, EBD, and NH students in regular education settings. The means, standard deviations, and ranges for ratings of the qualitative nature of instruction for the three groups of students are listed in Table 6. The average ratings are on the higher end of the 4-point qualitative rating scale. There was one significant difference in the qualitative nature of instruction for the three groups of students: Ratings for nonhandicapped students on the Task Relevance cluster were higher than ratings for LD and EBD students, F (2,86) = 3.30, p <.05.

Insert Table 6 about here

The variability within student category on each instructional cluster was considerable. The ranges, which appear in Table 6, are comparable for LD, EBD, and NH students in regular education settings. For each cluster and the total IES score, one student's instruction was "very much like" the variable rated (e.g., Instructional Planning) while another student's instruction was "not at all" or "not much like" the same variable.

Special education. One-way analyses of variance were used to compare ratings of the qualitative nature of instruction for LD, EBD, and the two groups of EMR students in special education. The means, standard deviations, and ranges for ratings of the qualitative nature

of instruction for the four groups of students are listed in Table 7. Again, the average ratings are on the higher end of the 4-point qualitative rating scale. There were four significant differences in the qualitative nature of instruction in special education settings for the four groups of students. EMR students served in resource rooms (EMR-R) were rated lower on the qualitative nature of instruction than any another handicap group on three instructional clusters and the total IES score. Specifically, EMR-R students' ratings on the Instructional Presentation cluster were lower than the ratings for LD, EBD, and EMR-S students, F (3,80) = 7.05, p <.001. Both groups of EMR students' ratings on the Checking for Student Understanding cluster were lower than ratings for LD students, F (3,80) = 4.82, p <.01. In addition, EMR-R students' ratings on the Checking for Student Understanding cluster were lower than the ratings for EBD students. For the Task Relevance cluster, EMR-R students' ratings were lower than the ratings for EBD and LD students, F (3,80) = 4.20, p <.01. Finally, EMR-R students' total score rating on the qualitative nature of instruction was lower than those of the three other groups, F (3,80) = 5.00, p <.01.

Insert Table 7 about here

There were no significant differences between the four groups on three qualitative nature of instruction clusters: Instructional Planning, Practice, and Feedback. There were no significant

differences between LD and EBD students' and EBD and EMR-S students on the qualitative nature of instruction in special education settings, and LD and EMR-S students' instruction differed only on the Checking for Student Understanding cluster.

The variability for the four groups of handicapped students was considerable, with the exception of Instructional Planning for LD and EMR-S students. LD students and EMR students served in self-contained classes were more similar in how their instruction was planned. There was more similarity in diagnostic procedures, curriculum sequence followed, establishment of individual goals and mastery criteria, and classroom management strategies for these students.

Setting differences. Only LD and EBD students were observed in both regular and special education settings. The average ratings for these students in regular education and special education settings are shown in Table 8. Results from two-way analyses of variance indicated no categorical or interaction (category x setting) effects. There was a main effect for setting, although LD and EBD students' instruction was rated on the higher end of the qualitative rating scale for both settings (see Table 8). Consistently, the qualitative nature of instruction for LD and EBD students was rated higher in special education on the total score and all instructional clusters, with the exception of the Practice cluster.

Insert Table 8 about here

Discussion

The purpose of this investigation was to examine the extent to which instruction differed for mildly handicapped students classified as LD, EBD, and EMR, and the extent to which handicapped students' instruction differed from nonhandicapped students' instruction. We found few categorical differences in how handicapped students were instructed. There were no differences over the entire school day in the tasks students used nor in the nature of their responses (academic, management, or inappropriate). Differences found in quantitative variables were a function of the setting rather than the category assigned to the student. For the qualitative variables, there were no differences between LD and EBD students in either regular or special education settings; differences found in the qualitative variables for LD and EBD students were a function of the setting. In addition, LD and EBD students' instruction was similar to nonhandicapped students' instruction, with the exception of one qualitative cluster. Within special education settings, EMR students' instruction differed from LD and EBD students on half of the instructional clusters. On each cluster, instruction was rated lower for EMR students, particularly those served in resource room settings.

On the quantitative measures, category by setting interactions suggest a need to look in more depth at the manner in which services are provided to EMR students, particularly those whose services are provided in the self-contained classroom. EMR students whose services were provided in this way spent more time on fetch and put away tasks in the special education setting and more time using special media in the regular classroom. However, they also spent less time in the

regular education setting making inappropriate responses. These findings, while not conclusive, might lead one to wonder about the efficiency of the self-contained classroom model and/or about our expectations for these students in general.

The qualitative measure, IES, describes the extent to which instruction for an individual student is characterized by principles of effective instruction. Although we found few categorical differences in the qualitative nature of instruction for the handicapped groups, the differences, where they exist, have important implications for the overall effectivness of instruction. Within special education settings, EMR students' instruction was rated lower on the three variables that comprise the "meat" of instruction, that is, Instructional Presentation, Checking for Student Understanding, and Task Relevance. The degree to which instruction was characterized by an explicit lesson explanation with substantive teacher-student interaction, followed by adequate success on an academically relevant task with sufficient teacher monitoring and frequent checking for student undestanding was lower for EMR students, particularly students served in resource rooms. In contrast, there were no differences between the groups in the more mechanical, or perhaps more thoroughly trained, aspects of instruction, that is Instructional Planning, Practice, and Feedback. All students' instruction, regardless of categorical designation, was similar in determining the students' instructional needs, kinds and amount of practice, and frequency and

feedback. The difference for EMR students' instruction lies in the prescriptive and interactive aspects of instruction.

Within the regular education setting there were no differences in the qualitative nature of instruction for LD and EBD students. However, these two groups of students' instruction was rated significantly lower on Task Relevance than was instruction for nonhandicapped students. Nonhandicapped students' instruction was characterized by a greater degree of high success on academically relevant tasks. There was a greater degree of congruence between the instructional goal and assigned tasks for nonhandicapped students. It appears that there is a better learning prescription for nonhandicapped than for mildly handicapped students in regular education. Given the amount of time LD and EBD students spend in regular education settings, this difference may mean differences in instructional outcomes for handicapped and nonhandicapped students. Research has demonstrated the importance of low achieving and handicapped students working on an academically relevant task with a high success rate (e.g., Rosenshine & Stevens, 1986). This may not be happening for many LD and EBD students during instruction in the regular education classroom.

Another major finding of this study was considerable variability within categories on both quantitative and qualitative aspects of instruction in both regular and special education. Clearly, students' instruction within the same categorical designation is not similar. For example, one LD student's academic responding time for an entire

school day was 25 minutes, while another LD student made active academic responses for 105 minutes. In addition, placement in special education does not mean instruction is similar. One EBD student's instruction was not characterized by checking for student understanding, while another EBD student's instruction was characterized very much by checking for student understanding. These kinds of differences in instruction result in differences in students' opportunity to learn. It is critical that educators note individual differences in how students respond to instruction. In fact, Stainback and Stainback (1984) have argued that individual differences, not categorical labels, should be the basis for providing special education services to students. Our data indicate that the categorical label assigned to a student does not delineate anything about the student's nature of instruction. Educators need to analyze the instructional environment (quantitative and qualitative aspects) for individual students and use this information in intervention planning.

There are significant setting effects for both quantitative and qualitative aspects of instruction for LD and EBD students. Quantitative differences in students' responses clearly favor the special education setting. While there were no setting differences on the Practice cluster, ratings on the qualitative aspects of instruction are significantly higher on the other instructional clusters in special education than in regular education settings. The specific reasons for these setting differences remain unverified. One hypothesis is that smaller student-teacher ratios and greater emphasis on meeting individual differences in special education allows for different opportunities for feedback, checking for student understanding, etc.

The challenge for researchers and educators is to design specific interventions that increase students' academic responding and can be used by regular education teachers with handicapped learners. Mildly handicapped students spend the majority of their school day in the regular education classroom. These teachers need helpful interventions - ones that allocate additional resources are necessary. Otherwise regular education teachers find that giving to one student means taking away from another (Gerber & Semmel, 1985). The allocation of additional resources may be particularly important for improving the instructional match or learning prescription for mildly handicapped students in regular education.

There are some limitations to the current investigation. Data on the quantitative and qualitative nature of instructional time were not collected at the same time. Further, the observational data are based on school-identified handicapped students and on a "snapshot" view of the student's total instructional program. Finally, the relationship between these instructional variables and achievement must be investigated. Recognizing these limitations, we believe special educators and child study team members need to address issues related to the value of classifying students. Whose needs does the categorical approach to special education serve? Clearly there are handicapped students, and these students, regardless of categorical designation, have different instructional experiences. Do we as educators want to spend valuable time determining how to divide up and what to call students? Or... do we want to use this time to instruct

and monitor closely the effectiveness of the instruction on the student's progress and performance? Carroll (1984) poignantly reminds us that it is what happens during the time allowed for learning that is important, and that student aptitudes for learning can possibly be improved through proper use of time.

In closing, we recognize that the investigation of categorical differences among LD, EBD, and EMR students is a relatively new area; most studies have been conducted since 1985. Although many of the recent studies appear to provide support for categorical differences (see Table 1), we believe that the findings need to be examined more closely. It is important to note that identified categorical differences across the reviewed studies occur most often when a dimension of the student's behavior (e.g., adjustment, teacher expectations) is rated by another individual. In recent studies, the support for categorical differences comes more from rated than from observed differences.

We expect that research on categorical differences among groups of handicapped students will continue. We speculate that some categorical differences will be found. A critical issue is where the differences exist and how such differences are used in educational practice. For example, Epstein et al. (1986) found that EBD students were rated as more hyperactive than LD or EMR students. We must ask whether this finding is important because the degree of hyperactivity may help to differentiate LD, EMR, and EBD students. Or, is it more important to learn how to teach any student who has hyperactive

characteristics? If identified group differences are not evident in the instructional environment for students, do they have practical significance for educators? Why debate differences among groups of handicapped students if nothing different occurs in instructional intervention? The variability within each of the handicap groups on the instructional environment measures suggests to us a need to attend more fully to individual differences. There are handicapped students, and each handicapped student's instructional experience is unique.

References

Algozzine, K., Morsink, C. & Algozzine, B. (1985). Classroom ecology in categorical special education classrooms: And so, they counted the teeth in the horse. Journal of Special Education, 20, 209-217.

Carroll, J.B. (1984). Educational research: Reflections on the past. In Carroll, J.B. (Ed.). Improving education: Perspectives on educational research. Pittsburgh, PA: National Academy of Education.

Cullinan, D. & Epstein, M. (1985). Adjustment problems of mildly handicapped and nonhandicapped students. RASE, 6(2), 5-11.

Epstein, M., Cullinan, D. & Gadow, K. (1986). Teacher ratings of hyperactivity in learning-disabled, emotionally disturbed, and mentally retarded children. Journal of Special Education, 20, 219-229.

Gerber, M., & Semmel, M.I. (1985). The microeconomics of referral and reintegration: Towards a new paradigm of special education evaluation. Studies in Educational Evaluation, 11, 13-29.

Greenwood, C.R., Delquadri, J., & Hall, R.V. (1978). Code for instructional structure and student academic response: CISSAR. Kansas City, KS: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.

Harrison, P. (1987). Research with adaptive behavior scales. Journal of Special Education, 21, 37-68.

Lang, R. & Kahn, J. (1986). Teacher estimates of handicapped student crime victimization and de nquency. Journal of Special Education, 20, 359-365.

Marston, D. (1987). Does categorical teacher certification benefit the mildly handicapped child? Exceptional Children, 53, 423-431.

O'Sullivan, P., Marston, D. & Magnusson, D. (1987). Categorical special education teacher certification: Does it affect instruction of mildly handicapped pupils? RASE., 8(5), 13-18.

Reynolds, M.C., Wang, M.C. & Walberg, H.J. (1987). The necessary restructuring of special and regular education. Exceptional Children, 53 (5), 391-398.

Rolison, M. & Medway, F. (1985). Teachers' expectations and attributions for student achievement: Effects of label, performance pattern, and special education intervention. American Education Research Journal, 22, 561-573.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), Handbook of research on teaching (3rd ed.) (pp 376-391). New York: McMillan.

Sindelar, P., Smith, M., Harriman, N., Hale, R., & Wilson, R. (1986). Teacher effectiveness in special education programs. Journal of Special Education, 20, 195-207.

Singer, J., Butler, J., Palfrey, J., & Walker, D. (1986). Characteristics of special education placements: Findings from samples in five metropolitan school districts. Journal of Special Education, 20, 319-337.

Stainback, W., & Stainback, S. (1984). A rationale for the merger of special and regular education. Exceptional Children, 51, 102-111.

Stanley, S.O., & Greenwood, C.R. (1980). CISSAR: Code for instructional response and student academic response: Observer's manual. Kansas City, KS: Juniper Garden's Children's Project, Bureau of Child Research, University of Kansas.

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

Ysseldyke, J.E., Christenson, S.L., McVicar, R., Bakewell, D., & Thurlow, M.L. (1986). Instructional environment scale: Scale development and training procedures (Monograph No. 1). Minneapolis: University of Minnesota, Instructional Alternatives Project.

Table 1. Summary of Outcomes of 21 Reviewed Studies Reporting Categorical Comparisons

Comparison Area	Number Reviewed	Outcomes		
		No Diff	?	Diffs
Academic performance	3	1	1	1
Behavioral characteristics	5	1	0	4
Medical problems	1	1	0	0
Social-emotional characteristics	2	1	0	1
Teacher and instructional characteristics	8	3	0	5
Post-school outcomes	2	0	1	1
Total	21	7	2	12

[a]Outcomes were judged to indicate support of categorical differences (Diffs), questionable support (?), or no support for differences (No Diff).

Table 2. Observation Event Areas, Specific Events Coded, and Composites Formed to Summarize Data[a]

Area	Events Coded	Post Hoc Composites
Instructional Task	Textbook	
	Workbook Worksheet Paper and Pencil	Paper Tasks
	Listen to Lecture Teacher-Student Discussion	Teacher Tasks
	Other Media Fetch/Put Away	
Student Response	Writing Playing Academic Game Oral Reading Silent Reading Talking Appropriately Answering Academic Questions Asking Academic Questions	Active Academic
	Active Academic Events Passive Responding	Academic Engaged
	Disruptive Playing Inappropriately Inappropriate Task Talking Nonappropriately Inappropriate Locale Looking Around	Inappropriate
	Raising Hand Looking for Materials Moving Playing Appropriately Waiting	Management

[a]Based on a modification of Stanley & Greenwood's (1980) CISSAR: Code for Instructional Structure and Student Academic Response: Observer's manual

Table 3. IES Instructional Clusters

Cluster	Description
Instructional Planning	Five items describe critical aspects of instructional planning, such as instructional match, curriculum sequence, goals, acceptable standards of performance, and classroom management procedures.
Instructional Presentation	Twelve items describe important aspects of developing and presenting an effective lesson, including instructional clarity (e.g., cues, modeling, clearly stated goals), explicitness of assigned tasks, opportunities for a student to respond, appropriate use of motivational techniques, and a well-paced lesson.
Checking for Student Understanding	Eight items describe important aspects of checking student understanding of how to perform the task accurately and monitoring student performance to ensure attention and progress toward achieving instructional goals.
Task Relevance	Six items describe important aspects of providing an academically relevant task to the student, including congruence between the lesson explanation and practice activity, appropriateness of success rate, adequacy of student understanding, and adequate academic engaged time.
Practice	Four items describe important aspects of practice activities, including amount of practice, variety and type of practice, and information on homework assignments.
Feedback	Five items describe important aspects of feedback including specificity and frequency, use of corrective procedures, and communication to the student.

Table 4. Numbers of Minutes Students Were Observed in Different Composite Tasks and Responses During an Entire School Day

Composite/Event	Category[a] LD	EBD	EMR	NH	Significance level[b]
Textbooks					
$\bar{X}$	43.7	35.1	26.9	42.3	ns
SD	24.8	22.4	18.8	22.4	
Range	13-112	0-90	0-69	0-81	
Paper Tasks					
$\bar{X}$	70.0	74.2	69.4	73.5	ns
SD	29.7	31.7	25.9	26.3	
Range	13-163	31-138	23-114	23-122	
Teacher Tasks					
$\bar{X}$	37.1	47.2	40.5	48.4	ns
SD	15.5	23.0	14.8	21.1	
Range	7-88	14-105	19-72	12-108	
Other Media					
$\bar{X}$	39.0	29.3	40.3	25.6	ns
SD	26.9	19.0	24.1	17.3	
Range	4-126	0-62	4-133	0-59	
Fetch/Put Away					
$\bar{X}$	24.0	28.3	28.1	25.7	ns
SD	11.5	12.4	12.4	11.1	
Range	6-51	8-57	9-55	10-57	
Active Academic Responses					
$\bar{X}$	69.8	64.4	60.5	72.8	ns
SD	22.7	20.5	20.1	15.2	
Range	25-105	37-123	31-124	40-110	
Active Engaged Time					
$\bar{X}$	133.7	126.9	121.2	141.6	ns
SD	31.3	32.5	28.4	28.0	
Range	66-184	63-211	72-195	82-184	
Inappropriate Responses					
$\bar{X}$	37.3	46.6	37.8	30.0	ns
SD	26.8	24.9	23.8	19.5	
Range	7-141	9-108	7-112	2-87	
Management Responses					
$\bar{X}$	42.6	40.3	46.0	43.6	ns
SD	16.9	17.0	19.4	15.2	
Range	12-86	8-81	16-83	21-80	

[a]Categories are LD = learning disabled (N=30), EBD = emotionally/behaviorally disturbed (N=32), EMR = educable mentally retarded (N=30), NH = nonhandicapped (N=30)

[b]Significance levels are based on one-way Anovas (df = 3,118) with p = .01 the criterion required for significance.

Table 5. Proportions of Time in Regular and Special Education Settings Students Were Observed in Different Composite Tasks and Responses[a]

Composite/ Event	Regular Education				Special Education				Sig. level[b]		
	LD	EBD	EMR-R	EMR-S	LD	EBD	EMR-R	EMR-S	Cat	Sett	CxS
Textbooks											
$\bar{X}$	.22	.16	.09	.20	.17	.14	.30	.04	ns	ns	.01
SD	.13	.10	.13	.14	.17	.23	.21	.03			
Paper Tasks											
$\bar{X}$	.30	.34	.34	.10	.39	.35	.49	.35	ns	.01	ns
SD	.14	.14	.18	.21	.20	.31	.19	.05			
Teacher Tasks											
$\bar{X}$	.21	.23	.26	.10	.11	.26	.06	.17	ns	ns	ns
SD	.10	.13	.13	.20	.10	.32	.05	.10			
Other Media											
$\bar{X}$	.13	.12	.16	.46	.28	.22	.11	.23	ns	ns	.01
SD	.14	.09	.10	.37	.19	.22	.14	.15			
Fetch/Put Away											
$\bar{X}$	.14	.15	.14	.14	.05	.03	.04	.21	ns	ns	.001
SD	.07	.08	.05	.09	.06	.03	.05	.08			
Active Academic Responses											
$\bar{X}$	.26	.26	.19	.20	.49	.45	.55	.34	ns	.0001	ns
SD	.10	.08	.09	.13	.17	.16	.12	.08			
Active Engaged Time											
$\bar{X}$	.51	.54	.50	.54	.79	.78	.80	.62	ns	.0001	ns
SD	.11	.14	.14	.11	.15	.14	.10	.12			
Inappropriate Responses											
$\bar{X}$	.18	.25	.03	.09	.08	.09	.10	.11	ns	.0001	.01
SD	.10	.16	.16	.07	.09	.10	.07	.10			
Management Responses											
$\bar{X}$	.23	.21	.19	.37	.12	.12	.11	.27	.0001	.0001	ns
SD	.08	.08	.08	.14	.10	.07	.05	.05			

[a]Categories are LD = learning disabled (N=28), EBD = emotionally/behaviorally disturbed (N=21), EMR-R = educable mentally retarded served in resource room (N=10), EMR-S = educable mentally retarded served in self-contained setting (N=5).

[b]Significance levels are based on repeated measures Anovas (Cat df=3,60; Sett df=1,60; CxS df=3,60), with p=.01 the criterion required for significance.

Table 6. Instructional Differences by Category in Regular Education[a]

Instructional Cluster	Category[a] LD N = 29	EBD N = 31	NH N = 29	Significance level[b]
Instructional Planning				
$\bar{X}$	3.4	3.3	3.6	ns
SD	.61	.74	.56	
Range	2.2-4.0	1.4-4.0	2.0-4.0	
Instructional Presentation				
$\bar{X}$	3.0	2.9	3.3	ns
SD	.72	.71	.60	
Range	1.3-4.0	1.2-3.8	1.8-4.0	
Checking for Student Understanding				
$\bar{X}$	2.7	2.8	2.8	ns
SD	.72	.76	.74	
Range	1.3-3.9	1.4-4.0	1.4-4.0	
Task Relevance				
$\bar{X}$	3.0	3.0	3.4	.042
SD	.65	.70	.59	
Range	1.5-4.0	1.6-4.0	1.3-4.0	
Practice				
$\bar{X}$	3.0	2.8	3.0	ns
SD	.71	.61	.48	
Range	1.3-4.0	1.5-3.8	1.7-3.8	
Feedback				
$\bar{X}$	3.2	3.0	3.1	ns
SD	.67	.66	.69	
Range	1.8-4.0	1.8-4.0	1.8-4.0	
Total Score				
$\bar{X}$	3.0	2.9	3.2	ns
SD	.57	.58	.50	
Range	1.9-3.7	1.6-3.8	1.8-3.8	

[a]Categories are: LD = learning disabled, EBD = emotionally/behaviorally disturbed, NH = nonhandicapped; the numbers are less than the total sample because of missing items.

[b]Significance levels are from one-way Anovas (df = 2,86), with $p = .05$ required for significance.

Table 7. Instructional Differences by Category in Special Education

Instructional Cluster	Category[a] LD N=30	EBD N=24	EMR-R N=10	EMR-S N=20	Significance level[b]
Instructional Planning					
$\bar{X}$	3.8	3.6	3.7	3.7	ns
SD	.26	.50	.54	.31	
Range	3.0-4.0	2.4-4.0	2.4-4.0	3.0-4.0	
Instructional Presentation					
$\bar{X}$	3.2	3.4	2.6	3.2	.000
SD	.41	.48	.63	.33	
Range	2.3-3.8	2.3-4.0	1.3-3.5	2.3-3.7	
Checking for Student Understanding					
$\bar{X}$	3.5	3.4	2.9	3.1	.004
SD	.43	.56	.74	.57	
Range	2.4-4.0	2.3-4.0	1.8-3.9	2.1-4.0	
Task Relevance					
$\bar{X}$	3.4	3.2	2.8	3.1	.008
SD	.37	.55	.70	.56	
Range	2.3-4.0	2.0-4.0	1.7-3.8	2.2-4.0	
Practice					
$\bar{X}$	3.1	2.9	2.8	3.1	ns
SD	.28	.53	.39	.54	
Range	2.3-3.5	2.0-4.0	2.0-3.3	2.0-3.8	
Feedback					
$\bar{X}$	3.6	3.7	3.3	3.6	ns
SD	.56	.47	.59	.52	
Range	1.8-4.0	2.4-4.0	2.0-4.0	2.4-4.0	
Total Score					
$\bar{X}$	3.4	3.4	2.9	3.3	.005
SD	.29	.39	.50	.34	
Range	2.6-3.9	2.5-3.9	2.2-3.7	2.6-3.7	

[a]Categories are LD = learning disabled, EBD = emotionally/behaviorally disturbed, EMR-R = educable mentally retarded students served in resource rooms, EMR-S = educable mentally retarded students served in self-contained classes; the number of subjects are less than the total sample due to missing items.

[b]Significance levels are from one way Anovas (df = 3,80), with p = .05 required for significance levels.

Table 8. Instructional Differences for LD and EBD Students in Regular and Resource Settings[a]

Instructional Cluster	Setting			
	Regular Education	Special Ecucation	F	P
Instructional Planning	3.4	3.7	7.11	.010
Instructional Presentation	2.9	3.3	14.75	.000
Checking for Student Understanding	2.7	3.5	31.65	.000
Task Relevance	3.0	3.3	11.65	.001
Practice	2.9	3.0	1.08	ns
Feedback	3.1	3.7	26.10	.000
Total Score	3.0	3.4	26.31	.000

[a]Results from the two-way anovas (df = 1,50) with 29 LD and 23 FBD students.