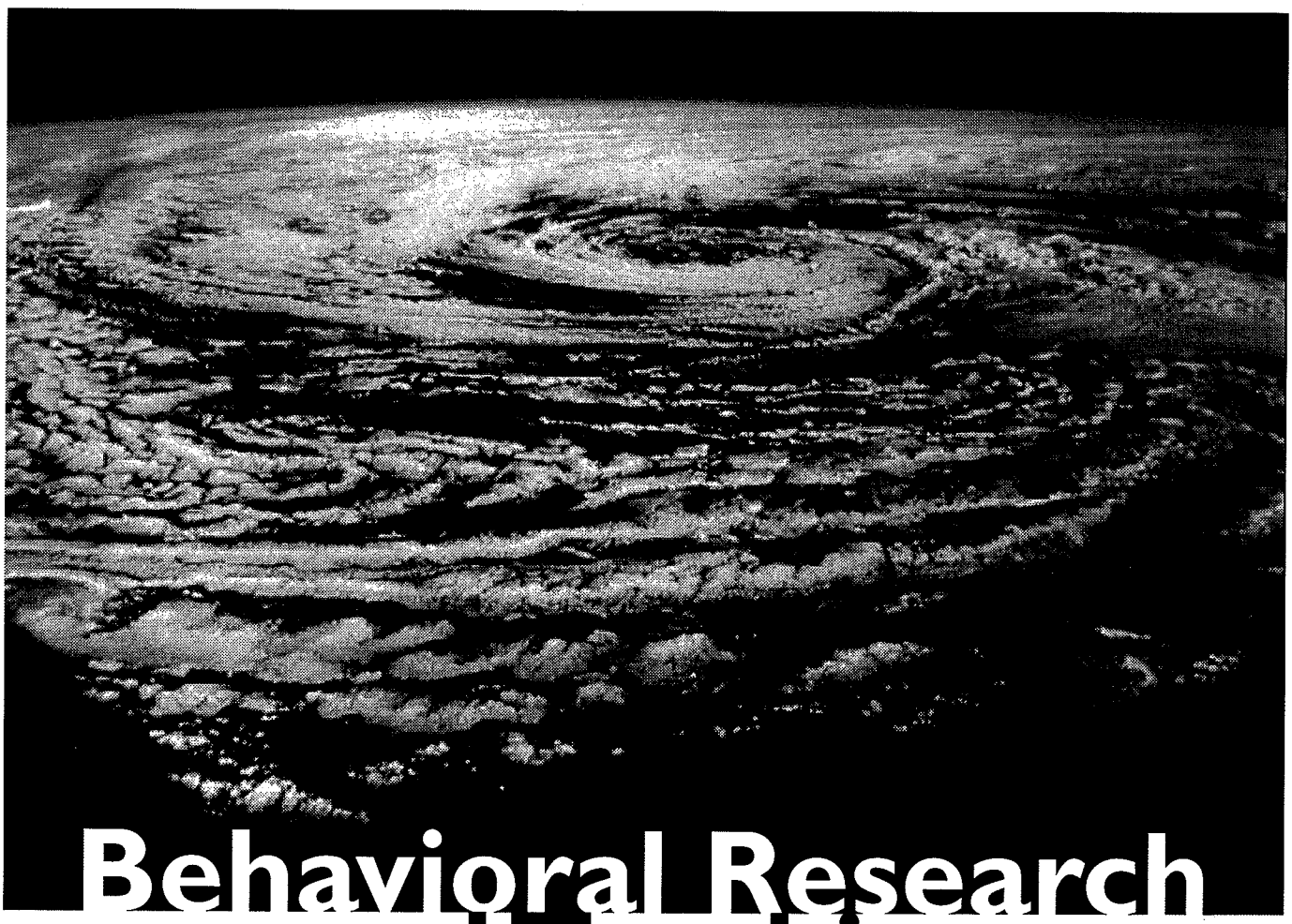

Behavioral Research and Teaching

Developmentally Appropriate Practice (DAP) from a Student Performance Assessment Perspective: A Study of Variance

Gerald Tindal, Ph.D.
Jerry Marr
Tracey Hall, Ph.D.
Carl Cole, Ph.D.
Nancy McCullum
Paul Goldman, Ph.D.

Research, Consultation, and Teaching Program
College of Education, University of Oregon, Eugene, Oregon

RR17

Published by:
Research, Consultation, and Teaching Program
Behavioral Research and Teaching
College of Education
University of Oregon

Staff
Gerald Tindal, Program Director
Gretchen Matis, Editor

Tindal, Gerald; Marr, Jerry; Hall, Tracey; Cole, Carl; McCullum, Nancy; Goldman, Paul;
Developmentally Appropriate Practice (DAP) from a Student Performance Assessment Perspective: A Study of Variance
Research Report 17

Preparation of this document was supported in part by the U.S. Department of Education, grant number H023R30026. Opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement by the Department should be inferred.

Cover design: Gretchen Matis

Abstract

A dual tension exists in many primary programs in which teachers strive to accommodate student differences at the same time that they organize students into grades and instructional groups. Recently, the term "developmentally appropriate practice" (DAP) has been used to describe a philosophy that presumably guides such an accommodation, typically in favor of age and individual appropriateness. In this study, we use three reading measures in a multi-age primary program (Grades 2 and 3) to investigate student differences: (a) across groups, (b) by age-grade, and (c) within individuals over time. We then document concurrent accommodations defined as referral for and placement into special education programs. The results reflect a distinct difference between an instructional perspective and a classroom reality.

Developmentally Appropriate Practice from a Student Performance Assessment Perspective: A Study of Differences

In this study, we focus on the assessment of student performance by teachers within a multi-age primary program to help define and operationalize the construct, *Developmentally Appropriate Practice (DAP).* We review the emerging literature in this area and offer our own views with supporting data. Then we focus on performance assessments in reading and concurrent special education referral-placement data. Using this information, we propose a model that may explain the effects of DAP in multi-age programs.

Developmentally Appropriate Practices

The development of early education programs to address the needs of *all* children, not just those who fit the normative and meritocratic ideals of Western industrialized culture, reflects the underlying values of a democratic society: active respect for diversity. It calls for pedagogical practices that accommodate not only ethnic differences, but social differences, developmental differences, and differences associated with disability (New & Mallory, 1994). This is the significance of the concept of developmentally appropriate practices. But, despite the ambitious efforts in the 1980s of the National Association for the Education of Young Children (NAEYC) to achieve consensus as to what DAP is or should be, debate continues (Bloch, 1991; Kessler & Swadener, 1992; Mallory & New, 1994; Carta, 1995).

As New and Mallory (1994) point out, the debate is not over whether developmental appropriateness should be practiced, but, rather, "*who* should make such a determination, and on the basis of *what* knowledge, values, and goals" (p. 3 [italics in original]). In simple terms, we would narrow the question of definition to this: How do we know DAP when we see it? How *should* we know it? What are the criteria for concluding that a given practice is truly developmentally appropriate? In approaching these necessary, if pedestrian, questions, we shall first examine the conventional approaches to the definition of DAP.

When the NAEYC issued its position statement and guidelines on DAP (Bredekamp, 1987), the idea was configured as an instructional strategy distilled from the input of thousands of early childhood practitioners (New & Mallory, 1994). These guidelines, rather than regarding developmentally appropriate practice as outcome based, a *function* of instruction, presented DAP as a constellation of *techniques* considered suitable to the formal education of young children. This focus on instructional techniques was due, in part, to the original conceptualization of DAP based upon Piagetian research in child development (O'Loughlin, 1991; Bredekamp, 1993).

One concept, *age appropriateness,* was derived from this body of knowledge. Age appropriateness refers to the understanding that "there are universal, predictable sequences of growth and change that occur in children during the first 9 years of life" (Bredekamp, 1987, p. 2). Another concept central to developmentally appropriate practice is *individual appropriateness.* This refers to the understanding that "each child is a unique person with an individual pattern and timing of growth, as well as individual personality, learning style, and family background" (Bredekamp, 1987, p. 2). However, these two dimensions of developmental appropriateness are not easily resolved into a practical instructional strategy for teachers in the classroom. As Carta (1995) points out, the NAEYC guidelines do not address individual appropriateness well: "As a consequence, teachers have been unclear about when to digress from age appropriate practices and how to modify teaching practices to address children's individual needs" (Carta, 1995, p. 1).

Again, from their original articulation to current thinking, DAP has been construed as an instructional strategy based on developmental stages and age appropriateness with curriculum planning based on "teachers' observations and recordings of each child's special interests and developmental progress" (Bredekamp, 1987, p. 3).

Thus, there is a disjunction in the articulation of the two concepts, age appropriateness and individual appropriateness. Developmental stages indicate that appropriate instruction must address "all areas of a child's development" (p. 3) and allow for "active exploration

and interaction" with the learning environment, as well as incorporate a broad array of activities and materials. Individual appropriateness indicates that such teaching strategies must be outcome based so that programs are sensitive to the needs of the individual student, which may be quite different at any stage or age. Furthermore, individual appropriateness is a dynamic construct in which the needs of the student change as a function of program success (or lack thereof), implying the need for continuous assessment.

In short, conventional understanding of DAP programs is based generally on an assessment of the developmental stage associated with a young child's age range (interest in others, self-awareness, physical, spatial, and temporal awareness, purposeful action and use of tools, motor skills, language development, and expression of feeling) and design of instruction to accommodate that developmental stage and any particular individual needs. This instructionally derived conceptualization of DAP, despite the NAEYC's comprehensive attempt to outline a continuum of teaching practices, results in a rather fragile and open-ended concept—one which has lead to misperceptions that tend to restrict the definition to certain practices. Carta (1995) identifies several of these "myths" regarding the principles of DAP, historically based as it is, on two distinct areas of research: early childhood education (ECE) and early childhood special education (ECSE). These myths are listed below. Carta identifies the first five as DAP myths and the second five as ECSE myths (Carta, 1995, pp. 3-4):

<u>DAP Myths</u>

1. Teachers following a DAP approach should never use direct instructional approaches.

2. Early childhood curriculum should be based entirely on children's interests and should not be goal directed.

3. Academics have no place in DAP programs.

4. DAP applies only to children who are typically developing, white, and middle class.

5. Programs adhering to the DAP guidelines must follow one specific approach if they are to be implemented correctly.

ECSE Myths

1. Most children with special needs must be taught in very structured activities in one-to-one settings with the teacher.

2. Activities that follow recommended practices in early childhood special education (ECSE) must be entirely teacher directed.

3. In typical ECSE practice, teachers must use artificial incentives to get children to perform behaviors.

4. Curricular content in ECSE focuses on academic subskills or individual behaviors that are taught out of context.

5. ECSE-recommended practices are completely behaviorally based and do not incorporate developmental principles.

The disjunction between DAP (ECE) and ECSE emerged because different underlying models guided their development: DAP was derived from research on child growth and maturation, and ECSE was originally based on behavioral analysis strategies and direct instructional approaches. More recently, ECSE practices have been increasingly based on developmental and developmental-cognitive models (Bricker & Cripe, 1992; Noonan & McCormick, 1993). According to Carta, this development in ECSE reflects a "convergence toward more naturalistic and functional approaches" (1995, p. 5).

Nevertheless, the juxtaposition of the above two sets of misperceptions highlights a fundamental confusion about developmentally appropriate practices. The confusion is either that DAP comprises a narrow set of instructional techniques applied in all cases, or that DAP is so broad as to be unsystematic in its application. It is a confusion we believe results from conventional definitions of DAP as sets of instructional techniques without adequate performance components. And it is a confusion about how to meld age

appropriateness (derived from developmental theory) with individual appropriateness (derived from behavioral theory). This confusion is manifested in the often ad hoc ways teachers apply DAP techniques, in the ways they assess their efficacy, and in the context-bound ways in which they report those assessments. For teachers, the question that we ask is a pragmatic one: How do we know DAP when we see it? How *should* we know it?

Talking to teachers underlines the need for arriving at a more systematic definition of DAP. The result is a definition that cannot be applied across settings. DAP implies both age and individual appropriateness. Yet, neither are anchored to any systematic definitions, analyses, or student performance outcomes. Many of the terms within the definitions are either tautological or idiosyncratic, using descriptors that imply outcomes which are not possible within a theoretical framework, or are so unique to a particular student that they fail to generalize to any overall principles. For example, to wait until students are ready to read without concurrently monitoring such a state of readiness implies that, at some unspecified time, students will simply initiate the behavior with no reference to previous models or guidance. Whence does such behavior occur? To assume that students differ is both obvious and ill-defined: On what common dimensions do they differ and by how much? While teachers may not be asking such questions directly, their perspectives and strategies can only be understood by answering them.

So, from a theoretical view, the construct of DAP must have attributes that move beyond individual teacher perceptions or student descriptors; instead, it needs a measurement system that can be researched using operational. Understanding age and individual appropriateness requires a common focus on inclusion and adaptation: How are groups of students to be grouped in a manner that is developmentally appropriate, that is sensitive to each student's emerging needs? If time is not the mark of change, then what is? What behavior can be used to mark change and provide a common index for understanding appropriateness for both age and individual? Finally, how do students progress through this development so that individual appropriateness remains somewhat

age appropriate? With such difficulty in understanding the construct of developmentally appropriate practices directly from teachers via interviews, from classroom observations, or from record reviews, we use this study to help define it empirically using student performance outcomes on a specific behavior.

Although developmentally appropriate practices are generally centered on assumptions about growth and appropriateness, rarely are the stages defined in empirical or quantitative terms. Rather, they are roughly assumed and rarely measured: Students arrive in schools with varying skills and backgrounds and have different trajectories for engaging in school programs. To investigate this issue, we developed a series of measures for students in a late primary program (Grades 2-3) and conducted several statistical analyses. In particular, we use three measures of reading to anchor both the level of performance as well as its change over time in defining age and individual appropriateness. We designed the study to ensure that the measures were consistent with teacher philosophy by incorporating teacher input in all phases of its development. We focus on the variance within a group of students to help operationalize the boundaries of age from a developmental perspective, to document the range of performance and to help us define what is "normal" within a bound of both standards and error.

In the second part of the study, we extend our analysis of individual appropriateness by focusing on referral to and placement in special education. Following our description of student reading performance, we present a dilemma viewing teacher responses as either accommodating the low performers to decrease the variance or placing students outside their environments.

Method

Teachers

The study took place in an elementary school serving a population of very low socio-economic status students. This school also had a very high transience rate with many families on welfare moving in and out during the year. It was ranked 337th from the top of the approximately 735 districts in the state.

Six teachers served in the late primary program with the following characteristics. The program was supported by four specialists: one self-contained teacher for severe behavior disorders, two resource teachers for students with learning disabilities, and one speech-language specialist. Teachers in the late primary program averaged over 6 years in the school and nearly 14 years in the district. They all were certified elementary teachers with one teacher also holding a special education and secondary certificate. Teachers had received their bachelor's degrees between 1973 and 1977; they had received additional professional credits from workshops and conferences accruing up to 60 credits each. Each of the teachers had nearly 10 years of elementary experience in Grades 2-3.

Subjects

We had 168 students in the six classrooms, although 42 students were not included in the analysis of reading because of missing scores in either the fall (25) or spring (17). An analysis of the demographic data for these 42 students revealed no differences in proportions broken down by gender, race, special education status, or attendance. The sample studied included 75 students in Grade 2 and 57 students in Grade 3. Although the vast majority of students were taught in the general education classroom (73), many students received some type of specialized service: Chapter 1 (23), Talented and Gifted (8), or special education (28). There was an even split of males and females (66 each), and most students were Caucasian with only 7% from varying ethnic backgrounds. Students missed very little school during the year, with 2-3 days being the average absence rate per trimester (see Table 1).

Table 1

Subject Descriptive Statistics for All Students

Grade Level	Count	Percent
Grade 2	98	57.310
Grade 3	73	42.690
Total	171	100.000

Status	Count	Percent
Chapter/Title	27	16.071
General Ed	95	56.548
Special Ed	38	22.619
TAG	8	4.762
Total	168	100.000

Assist	Count	Percent
Unknown	4	11.765
L	1	2.941
L/R/M/Spl	1	2.941
M/WE/Sp/R/MED	1	2.941
R	3	8.824
R/M/Sp/L	1	2.941
R/Sp/L	1	2.941
R/M/L	1	2.941
R/M/Sp	1	2.941
R/M/Spl/L	1	2.941
R/M/W/Spl	1	2.941
R/Sp	1	2.941
R/W	1	2.941
R/W/M	1	2.941
R/W/Sp	1	2.941
Sp	3	8.824
Sp,L	2	5.882
Sp/L	9	26.471
Total	34	100.000

Gender	Count	Percent
Female	88	51.765
Male	82	48.235
Total	170	100.000

Ethnicity	Count	Percent
African American	1	1.785
Asian Pacific	4	2.381
Hispanic	3	1.786
White	158	94.048
Total	168	100.000

Table 1 continued

Absence by Qtr*Grade	Q1-Abs, Total	Q1-Abs, 2	Q1-Abs, 3
Mean	1.748	1.718	1.788
Std. Dev.	2.231	2.223	2.257
Std. Error	.182	.241	.278
Count	151	85	66
Minimum	0.000	0.000	0.000
Maximum	14.000	14.000	10.000
# Missing	20	13	7

Absence by Qtr*Grade	Q2-Abs, Total	Q2-Abs, 2	Q2-Abs, 3
Mean	2.735	3.000	2.369
Std. Dev.	3.136	3.519	2.491
Std. Error	.252	.371	.309
Count	155	90	65
Minimum	0.000	0.000	0.000
Maximum	23.000	23.000	12.000
# Missing	16	8	8

Absence by Qtr*Grade	Q3-Abs, Total	Q3-Abs, 2	Q3-Abs, 3
Mean	2.750	2.844	2.621
Std. Dev.	3.380	3.301	3.507
Std. Error	.271	.348	.432
Count	156	90	66
Minimum	0.000	0.000	0.000
Maximum	17.000	17.000	17.000
# Missing	15	8	7

Development of Reading Measurement Systems

We worked with the teachers over the course of a year to develop several measures in reading. Through monthly workshops, tasks were developed that reflected important academic skills; in addition, scoring systems were generated for producing both quantitative and qualitative outcomes.

The reading tasks consisted of students reading from two passages (either from an easy passage and a medium difficulty passage, or from a medium difficulty and a hard passage). They read for one minute and three scores were generated: (a) oral reading fluency, the number of correct words per minute (see Deno, Mirkin, & Chiang, 1982); (b) a prosody rating of expression quality (see Tindal & Marston, in press); and (c) retell quality (see Hall & Baker, 1994). The last measure of retell quality was obtained by having students continue to read the medium difficulty passage silently after the one minute of oral reading and then retell it aloud to the administrator, with their response tape recorded for a portfolio and analyzed for the presence and quality of seven story grammar elements as well as an overall rating of retell quality. See Figure 1 for a description of the reading administration and scoring directions.

Results

We have divided the results section into two parts, reflecting the use of two types of data: (a) reading performance data, and (b) referral-placement data. In the first section, reading data include oral reading fluency, reading prosody, and retell quality for all students (in both grade levels and in both general and special education). In the second section, for both the nongraded primary program (NGP) and a comparable graded primary program (GP), referral and placement data are compared graphically for two time intervals, before and after introduction of the NGP.

Understanding Student Variance in a DAP Program

We used the same procedures to analyze reading performance data on all three measures: oral reading fluency, prosody, and retell quality. Two distinct analyses were completed: (a) comparisons between grade levels, and (b) comparisons between students in different status categories. In both sets of analyses, we compared student performance over time (from fall to winter to spring).

Differences Between Grades

In the first set of analyses, we compared the performance of students in the same late primary program using grade to block students into two cohort groups reflecting traditional grade designations. We conducted two types of analyses: (a) a repeated measures analysis of variance using both a between-group factor (Grades 2 and 3) and a within-subjects factor (time: fall, winter, and spring); and (b) stacked bar charts to look at actual intervals on the performance scale where students from different grade groups either overlapped or took outlying positions. In the stacked bar charts depicting student performance on oral reading fluency, prosody, and retell quality, the darker shaded portion reflects Grade 2 students, and the lighter shaded portion reflects Grade 3 students.

Oral reading fluency. As depicted in Table 2, the comparisons for oral reading fluency produced significant differences ($p < .0001$): between grade levels, over time, and as an interaction between grade and time. Because of this interaction, the main effects are rendered unimportant, with growth of second graders significantly greater than that attained by the third graders.

Table 2.

Repeated Measures Analysis of Variance Comparing Oral Reading Fluency between Grade 2 and 3 Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-ORF

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Grade	1	103979.090	103979.090	31.326	<.0001
Subject(Group)	127	421547.582	3319.272		
Category for FWS-ORF	2	39441.385	19720.693	164.939	<.0001
Category for FWS-ORF * Grade	2	2282.854	1141.427	9.547	.0001
Category for FWS-ORF * Subject(Group)	254	30369.095	119.563		

3 cases were omitted due to missing values.

Grade	Count	Mean	Std. Dev.	Std. Err.
2	222	57.716	38.903	2.611
3	165	90.861	31.153	2.425

3 cases were omitted due to missing values.

FWS-ORF	Count	Mean	Std. Dev.	Std. Err.
FALL-ORF	129	60.535	39.358	3.465
WINT-ORF	129	69.961	35.844	3.156
SPRING-ORF	129	85.047	39.076	3.440

3 cases were omitted due to missing values.

FWS-ORF * Grade	Count	Mean	Std. Dev.	Std. Err.
2, FALL-ORF	74	43.514	35.492	4.126
2, WINT-ORF	74	56.716	35.878	4.171
2, SPRING-ORF	74	72.919	39.940	4.643
3, FALL-ORF	55	83.436	32.215	4.344
3, WINT-ORF	55	87.782	27.229	3.672
3, SPRING-ORF	55	101.364	31.467	4.243

3 cases were omitted due to missing values.

In the stacked bar charts depicting student performance on oral reading fluency (see Figure 2), the fall sample shows a positively skewed distribution for second graders, with increasingly fewer students represented in the middle and high intervals. The data clearly reflect a distribution in which the performance of second graders overlaps considerably with that of third graders. Only in the lowest interval (0-10 words correct per minute) is a single grade level represented. Figure 2 provides a visual illustration of how this overlap in performance increased from fall to spring as second graders made greater gains than the third graders in oral reading fluency. Second grade scores increased from fall to spring by almost a full standard deviation from the fall mean, while third grade scores increased by little more than one half of a standard deviation. This difference is depicted visually in Figure 2.

Prosody. For prosody, the same results were found for both main effects (grade and time) and the interaction of grade x time (see Table 3). This interaction has been plotted as a bar chart in Figure 3. The second grade distribution is again positively skewed (only second grade is represented in the lowest interval, score 0), yet many second graders still read qualitatively like third graders as depicted by their numbers in the remaining interval scores.

As with oral reading fluency, second grade students made greater gains through the year than did third grade students on this measure of reading expression. Grade 2 student performance grew from a mean score of 1.9 in the fall (on a scale of 0-4) to 2.6 in the spring, while Grade 3 student performance remained stable as indicated by a mean score of 3.3 in both the fall and the spring.

Retell quality. For the retell, the only significant effect was with time ($p < .0001$). No differences between students from Grades 2 and 3 were found and no interaction of change over time by students from these grades was found. The separate distributions of each grade at each time are plotted in Figure 4. The stacked bar chart for the retell measure is nearly identical to that obtained with the prosody measure, reflecting a distribution that is

positively skewed with common performances across the distribution (see Figure 4). Again, Grade 2 students' scores increased by a standard deviation from fall to spring, while Grade 3 scores increased by only about one half of a standard deviation from the fall mean.

The data clearly reflect a distribution in which students' performance overlaps considerably as illustrated by the histograms in Figures 2-4. The formal statistical analysis comparing performance between grade levels at each of the times (fall, winter, and spring) generally reveals differences in the growth over the year for students from different grade levels, with Grade 2 students exhibiting the greatest gains.

Serving Differently-Abled Students in Developmental Classrooms

In our final analysis, we focus on the performance of differently-abled students in the general education distribution. We compared the performance of students in the same late primary program using status categories to locate students of varying abilities in the general distribution. Status categories include: Chapter I, Special Education, General Education, and Talented and Gifted.

We conducted a formal statistical analysis, using a repeated measures analysis of variance, to ascertain whether the level and change of reading fluency, prosody, and retell quality for students from specified categories was significantly different than those in general education. Most of the students included in the Special Education category are identified as having learning disabilities in reading (see Table 1) and are therefore of particular interest in these analyses.

Oral reading fluency. In oral reading fluency, the difference for both main effects (student status and time) is significant ($p < .0001$): Special education students read significantly fewer words correct per minute than their general education cohorts (about half) and both groups improve significantly from the fall to winter to spring. No interaction exists reflecting differential amounts of growth by these two groups over time. See Table 5 and Figure 5. While the stacked bar charts in Figure 5 clearly depict clustering

of Chapter I and Special Education students among the lower intervals, they also illustrate that some Special Education students are performing at the higher intervals.

Prosody. For prosody, the same results occur: Students with disabilities read qualitatively less well, all groups of students improve significantly over time ($p < .0001$), and no interaction exists between student status and growth (see Table 6 and Figure 6). As in oral reading fluency, most of the Chapter I and Special Education students perform at the lower intervals, but a few are represented at the higher intervals of the measure.

Retell quality. Finally, for the retell quality measure, significant effects occur again with status ($p = .0001$) and time ($p < .0001$): Differences exist in the quality of the retell as a function of student status and students, in general, improve from fall to winter to spring. No interaction is significant in the change of retell quality over time as a function of student status. See Table 7 and Figure 7. For retell quality, Chapter I and Special Education students are represented at all intervals but with greater numbers at the lower end of the scale.

Understanding the Educational Implications of Student Variance

At this point in the analysis, we have analyzed student reading performance in a multi-age primary program with students spanning Grades 2 and 3 and across different educational programs. In our analysis, we regrouped them into the grade levels to which they would have been assigned on the basis of age. In all of our analyses, we found considerable overlap in the distribution of skills on all of our reading measures. Some students in the younger grades performed as well as those in the older grades and vice versa. We also found differences in the manner in which this overlap was evident. With the use of stacked bar charts, the overlap was considerably more evident than in the statistical analyses of variance, whether between groups or over time. In light of the performance data reflected in the first part of this study, we now turn our analysis to the referral and placement patterns of two schools.

The experimental school was the nongraded primary program (NGP), and it was compared to a graded school program (GP) of similar size, catchment area, and socio-economic status. In Figure 8, we have plotted the referral rate over two years (1991-92 and 1992-93) for both the nongraded and graded programs; during this time, both schools were operating graded programs. In the early grades (1 and 2), the NGP school exceeded the GP school slightly (with a comparable overall ratio of 14:15, see Figure 8). In Figure 9, we have plotted the number of referrals after the experimental school began implementing the nongraded primary program (averaging 2 years' worth of data from 1993-94 and 1994-95). In the first three grades, particularly in Grades 1 and 2 (wherein the late primary 2-3 program would be analyzed), the number of referrals is almost 2:1 between the nongraded and graded schools (see Figure 9).

We have plotted referral data from 2-year intervals for the nongraded primary program across several grade levels to analyze those grades with the highest number of referrals. As depicted in Figure 10, the mode for referrals in the NGP program was first grade prior to establishing the NGP program; with NGP, the grade with the most referrals shifted to Grades 2-3, the very level which was nongraded (see Figure 10).

In Figure 11, we have plotted comparable data from the graded program (in 2-year intervals similar to the two years before establishing the NGP and after establishing the NGP). In this comparable school, the mode shifts from third grade prior to the NGP in the experimental school and following its implementation, the mode shifts to fourth and fifth grades (see Figure 11).

In Figure 12, we have plotted current referrals of both the non-graded and graded programs (1994-1995 data). Not only are the number of referrals in the NGP more than double the number in the graded program, but the very grades in which these increases are found are the nongraded levels (Kindergarten-Grade 1 and Grades 2-3). Comparable numbers appear in Grades 4 and 5 (see Figure 12).

Finally, we have plotted current (1994-1995) placement data for the two programs: nongraded and graded. Again, the numbers in the NGP program are much greater than those in the GP program, and the increases are located at exactly those grade levels addressed with NGP (see Figure 13).

Discussion

This study represents an investigation of Developmentally Appropriate Programs (DAP) within a multi-age, nongraded elementary program. The two major components of DAP address age appropriateness and individual appropriateness. In our discussion, we emphasize the complex nature of this construct and the need to build programs on such constructs more empirically so they are more aligned with outcomes. Finally, with respect to a few limitations, we offer an explanation and propose a further refinement of this research.

Age appropriateness implies both a level of skill or knowledge *and* a sequence of time within which the skills and knowledge are present *for most students.* Individual appropriateness, on the other hand, implies consideration of skill and knowledge of a person over time, with little concurrent consideration of the levels present with most students. For individuals, a time series approach is needed to ensure that the sequence of skills builds and develops in a somewhat orderly fashion, though not at precisely the same rate across individuals.

Among early childhood advocates, age appropriateness is frequently used to ensure that the cognitive emphases of school programs are aligned with the social and behavioral skills of students and the requisite skills of their cohort group. Though inherently norm-referenced, this definition is usually anchored to only loose notions of what is appropriate. For example, many DAP advocates would argue against an academic kindergarten program because the emphasis on literacy skills is not aligned with the behaviors characteristic of that age group and which must be present for them to learn (length of attention span, level of motivation, ability to work independently, etc.). In our study, one could argue that

emphasis on comprehension in second grade is inappropriate until students have firm practice and experience in simply learning to read. Two problems underlie this definition: First, no criteria exist for making the judgment about appropriateness, and second, not asking a question precludes attaining an answer. In the end, both problems leave the definition ill-equipped to help resolve educational problems.

Our results provide a more explicit set of criteria for making judgments about appropriateness for both the age (group) and the individual, actual performance scores and changes. From a normative perspective, it is obvious that DAP in reading is a useful construct for explicating individual but not age appropriateness. Our grade level analyses revealed that at the group level, the average performances of second and third graders were significantly different. At the individual level, many second graders performed well above third graders, and many third graders read more poorly than second graders. Thus, the program reflected both age and individual appropriateness. This finding was true for all of our measures, and such distributions would argue for DAP. On the other hand, an interaction was found between grade and time, with more growth occurring among the second graders than the third graders in fluency and prosody; this interaction was not significant with the retell measures. With this finding, the empirical support behind DAP practices is somewhat eroded. In essence, implementation of a DAP program reduced the differences between the grades, positioning more second graders at a higher level in the spring than in the fall. In other words, the program was not uniformly effective.

Within a DAP program, while teachers may provide accelerated opportunities for their younger students, they also need to ensure that the older students are not being held back. This argument is framed at the group level, not the individual level. So, while the program may appear to be individually appropriate, it may not be adequately framed as age appropriate. This perspective reflects an important distinction between nomothetic (general or group) and idiographic (individual) analyses (Allport, 1962). From a nomothetic perspective, groups are defined in terms of average performance and variation, yet from an

idiographic view, this average does not reflect any one individual. As we moved our analyses from the group to the individuals within the distributions, we see very different results. For example, the bar charts for each of the measures depict movement of individuals within a distribution over time and, though it is not possible to mark the movement of specific cases within these charts, some students improve while others do not. In general, more students in the earlier grade improve than students in the later grade. In these instances, individual appropriateness appears to be an important component of DAP and one which is well justified.

The lack of interaction between grades and time for the retell measures reflects the difficulties in operationalizing DAP in terms of age appropriateness. Presumably, the older students should have improved more on comprehension than the younger students, reflecting the shift in emphasis in reading programs from learning to read to reading to learn. An interaction would have been expected with differentially higher growth for the third graders from fall to spring; yet no interaction was found. Rather, only the main effect for grade and the main effect for time was found.

In addition to helping us explicate the construct of DAP, particularly in terms of age and individual appropriateness, the study also sheds light on some of the myths outlined previously (Carta, 1995). We saw academic growth in students' reading, clearly refuting the myth that academics have no place in DAP programs; the reading program appears to help students in our low SES population at all points of the achievement continuum, refuting any narrow applications of DAP to middle class, white students. Finally, the reading program appeared to work quite well across all teachers, who taught with very different strategies and practices, negating any need for singularity in approach. Obviously, however, across-teacher differences also were apparent (reflecting the nomothetic approach), making this overall gain apparent, while individual differences existed (reflecting the idiographic approach), a phenomenon which we did not study.

Several ECSE myths also were refuted: students need a very structured environment to learn, and activities need to be entirely teacher directed with artificial incentives used along with a task analyzed curricular focus. In these classrooms, the reading program was loosely structured around an activity-based approach using centers, with teaching more aptly described as facilitative than directive. Yet, many students learned to read in all aspects of fluency, prosody, and comprehension. On the other hand, given our concern with operationalizing individual appropriateness, for some students this outcome was not the case and, therefore, a more systematic approach to instruction would be warranted.

This last result leads us to the final set of data that were presented in this study, referral-placement rates. Generally, when teachers meet an unacceptable range of performance in their classrooms, they do not reduce it by bringing the bottom up; rather they eliminate it by sending the bottom out. As teachers confront difficult-to-teach students (presumably many of whom are experiencing reading problems), they refer them for special education placement.

Using Gerber and Semmel's perspective (1984, 1985), classroom variance can be understood from the demand it makes on teachers using a microeconomic analysis. According to this view, mild disabilities are developmental phenomena in which students' individual differences interact with the environment and acquire educational significance only by teachers' effective use of time and instructional effort, both of which place limits on the situation in the form of "tolerance" (the amount of variation allowed from the standard amount). Basically, students with disabilities are outside the band of tolerance, which, in turn, defines their "teachability," given the (limited) resources and time at hand. To increase tolerance, and therefore be more supportive of students with disabilities, teachers need to have more time or resources; alternatively, they can transfer responsibility to other professionals. This view is termed microeconomic because of the emphasis on limited and shared resource allocation, not unlike a supply and demand model embedded within the distribution of goods and services in the world of business.

Clearly, the multi-age classroom is more susceptible to this effect, with its inherent increased variance. A classroom of students from Grades 2 and 3, for example, is more likely to include higher and lower performing students (who may not necessarily be from upper and lower grades, respectively). Ironically, from a developmentally appropriate perspective, teachers have not only accommodated, but demanded this kind of increase in variance. The question, then, is what happens with failure over time? Do teachers refer these students for special education? Our data indicate that they do. According to Ysseldyke et al. (1982), a referral is likely to result in an assessment 90% of the time; the probability of the assessment resulting in special education eligibility and placement is 75%. Together, these odds reflect an overall probability of 67%, or two out of three chances that a referral results in placement. These data suggest a limited capacity of the general education classroom to accommodate individual differences in academic performance.

While the study begins to explicate DAP, advocates may well raise several arguments against our results. For example, we confined the study to reading, as opposed to all academic areas and social-behavioral areas. This perspective is possibly too narrow for analyzing constructs like age and individual appropriateness. On the other hand, any broader definitions make research designs difficult to implement. Finally, it's possible that our measures were more sensitive for a particular grade level, possibly because of its scaling properties. That is, the second graders showed more improvement because of the nature of the measure rather than the impact of the reading program. We would simply counter with an appeal to the vast technical adequacy literature that exists for reading fluency and prosody (Marston, 1989; Tindal & Marston, in press) as measures of broad reading programs. Such a criticism may exist for the retell, though technical adequacy of this measure has been reported by Fuchs, Fuchs, and Maxwell (1988).

In the end, the results should speak for themselves. DAP is a construct that may well be founded upon an ill-conceived basis. While individual appropriateness appears to be

supported, age appropriateness is more assumed than demonstrated. And yet, when put to the ultimate test of flexibility and responsiveness, individual appropriateness apparently has its limits too: When an unknown line is passed marking achievement below some level, students are referred and placed into special education programs outside of the DAP program.

To ensure that these findings are not spurious, then, this research needs to be extended to the construction and implementation of instructional programs. While we took an outcomes criterion (Reshly, Kicklighter, & McKee, 1988), with student performance determining the ultimate test of age and individual appropriateness, a process perspective may well address this issue. For example, with direct observations in classrooms, it would be possible to see if students and teachers engage in different activities within or across age groups; a 2-year study of the same students would help determine if students move through unique opportunities the second year or if they simply receive more of the same. Finally, observations of interactions between students and teachers would help determine if the program is indeed differential according to age and individual. If concurrent data were taken on outcomes, it might be possible to extend the product-process research paradigm of the 1980s into an explication of a specific construct, Developmentally Appropriate Practices.

Author Note

Gerald Tindal, Behavioral Research and Teaching; Tracey Hall, now at Penn State University; Jerry Marr, University of Oregon; Nancy McCullum, University of Oregon; Carl Cole, Bethel School District Number 52.

Publication of this paper was supported in part by the U.S. Department of Education, Office of Special Education Programs, grant number H023R30026. Opinions expressed herein do not necessarily reflect the position or policy of the U. S. Department of Education, and no official endorsement by the Department is implied.

This study was completed with the help of Denise Swanson and Nancy Bell, both of whom were instrumental in administering and scoring the reading assessments for the students. The teachers in this study also should be acknowledged for their support of our research and their pursuit of best practices. Data from the assessment were previously reported in paper published in the Oregon Conference Monograph (published in 1996, Abe Deffenbaugh, Gretchen Matis, and Carlie Neudeck, editors).

Correspondence concerning this paper should be addressed to: Dr. Gerald Tindal, Behavioral Research and Teaching, 5262 University of Oregon, Eugene, Oregon, 97403-5262. Electronic mail may be sent via Internet to: geraldt@darkwing.uoregon.edu.

References

Allport, G. W. (1962). The general and the unique in psychological science. Journal of Personality, 30, 405-422.

Bloch, M. (1991). Critical science and the history of child development's influence on early education research. Early Education and Development, 2(2), 95-108.

Bredekamp, S. (Ed.) (1987). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. Washington, DC: National Association for the Education of Young Children.

Bredekamp, S. (1993). The relationship between early childhood education and early childhood special education: Healthy marriage or family feud? Topics in Early Childhood Special Education, 13, 258-273.

Bricker, D., & Cripe, J. J. (1992). An activity-based approach to early intervention. Baltimore: Brookes.

Carta, J. J. (1995). Developmentally appropriate practice: A critical analysis as applied to young children with disabilities. Focus on Exceptional Children, 27(8), 1-14.

Deno, S., Mirkin, & Chiang. (1982). The use and standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study. Minneapolis, MN: Minnesota University, Institute for Research on Learning Disabilities.

Fuchs, L., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. Remedial and Special Education, 9(2), 20-28.

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. Educational Psychologist, 19(3), 137-148.

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

Hall, T., & Baker, S. (1994). Using classroom assessment data to monitor school reform: Evaluating student performance in nongraded classrooms. Diagnostique, 20 (1-4), 175-203.

Kessler, S. A., & Swadener, B. B. (Eds.) (1992). Reconceptualizing early childhood education curriculum: Beginning the dialogue. New York: Teachers College Press.

Mallory, B. L., & New, R. S. (Eds.) (1994). Diversity and developmentally appropriate practices: Challenges for early childhood education. New York and London: Teachers College Press.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), Curriculum-based measurement: Assessing special children. New York: Guilford Press.

New, R. S., & Mallory, B. L. (1994). Introduction. In B. L. Mallory & R. S. New (Eds.), Diversity and developmentally appropriate practices: Challenges for early childhood education. New York and London: Teachers College Press.

Noonan, M. J., & McCormick, L. (1993). Early intervention in natural environments: Methods and procedures. Pacific Grove, CA: Brooks/Cole.

O'Loughlin, M. (1991, October). Rethinking early childhood education: A socio-cultural perspective. Paper presented at Conference on Reconceptualizing Research in Early Childhood Education, Madison, WI.

Reschly, D. J., Kicklighter, R. H., & McKee, P. (1988). Recent placement litigation, part 3: Analysis of differences in Larry P., Marshall, and S-1 and implications for future practices. School Psychology Review, 17, 37-48.

Tindal, G. & Marston, D. (in press). Technical adequacy of alternative reading measures as performance assessments. Exceptionality.

Ysseldyke, Thurlow, Garden, Wesson, Deno, & Algozzine (1982). Generalizations from five years of research on assessment and decision-making [Research Report No. 100]. Mpls: University of Minnesota Institute for Research on Learning Disabilities.

Table 3.

Repeated Measures Analysis of Variance Comparing Prosody between Grade 2 and 3 Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-PROS

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Grade	1	92.279	92.279	37.933	<.0001
Subject(Group)	127	308.951	2.433		
Category for FWS-PROS	2	11.757	5.879	17.174	<.0001
Category for FWS-PROS * Grade	2	11.302	5.651	16.509	<.0001
Category for FWS-PROS * Subject(Group)	254	86.941	.342		

3 cases were omitted due to missing values.

Grade	Count	Mean	Std. Dev.	Std. Err.
2	222	2.279	1.123	.075
3	165	3.267	.925	.072

3 cases were omitted due to missing values.

FWS-PROS	Count	Mean	Std. Dev.	Std. Err.
FALL-PROS	129	2.481	1.288	.113
WINT-PROS	129	2.713	1.002	.088
SPR-PROS	129	2.907	1.114	.098

3 cases were omitted due to missing values.

FWS-PROS * Grade	Count	Mean	Std. Dev.	Std. Err.
2, FALL-PROS	74	1.851	1.069	.124
2, WINT-PROS	74	2.392	1.018	.118
2, SPR-PROS	74	2.595	1.158	.135
3, FALL-PROS	55	3.327	1.055	.142
3, WINT-PROS	55	3.145	.803	.108
3, SPR-PROS	55	3.327	.904	.122

3 cases were omitted due to missing values.

Table 4.

Repeated Measures Analysis of Variance Comparing Retell Quality between Grade 2 and 3 Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-RETELL

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Grade	1	6.240	6.240	3.537	.0627
Subject(Group)	108	190.515	1.764		
Category for FWS-RETELL	2	25.679	12.839	20.227	<.0001
Category for FWS-RETELL * Grade	2	2.545	1.272	2.005	.1372
Category for FWS-RETELL * Subject(Group)	216	137.110	.635		

22 cases were omitted due to missing values.

Means Table for FWS-RETELL
Effect: Grade

Grade	Count	Mean	Std. Dev.	Std. Err.
2	174	2.109	1.051	.080
3	156	2.385	1.032	.083

22 cases were omitted due to missing values.

FWS-RETELL	Count	Mean	Std. Dev.	Std. Err.
FALL-RETELL	110	1.955	.962	.092
WINT-RETELL	110	2.145	1.039	.099
SPR-RETELL	110	2.618	1.040	.099

22 cases were omitted due to missing values.

FWS-RETELL * Grade	Count	Mean	Std. Dev.	Std. Err.
2, FALL-RETELL	58	1.707	.879	.115
2, WINT-RETELL	58	2.069	1.057	.139
2, SPR-RETELL	58	2.552	1.046	.137
3, FALL-RETELL	52	2.231	.983	.136
3, WINT-RETELL	52	2.231	1.022	.142
3, SPR-RETELL	52	2.692	1.039	.144

22 cases were omitted due to missing values.

Table 5.

Repeated Measures Analysis of Variance Comparing Oral Reading Fluency between General and Special Education Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-ORF

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Status	3	214960.407	71653.469	28.840	<.0001
Subject(Group)	125	310566.265	2484.530		
Category for FWS-ORF	2	39441.385	19720.693	155.572	<.0001
Category for FWS-ORF * Status	6	961.403	160.234	1.264	.2745
Category for FWS-ORF * Sbjct(Grp)	250	31690.545	126.762		

Means Table for FWS-ORF
Effect: Status

	Count	Mean	Std. Dev.	Std. Err.
C	69	47.870	29.360	3.535
G	213	86.854	29.599	2.028
S	81	39.975	35.868	3.985
T	24	115.167	39.223	8.006

Means Table for FWS-ORF
Effect: Category for FWS-ORF

	Count	Mean	Std. Dev.	Std. Err.
FALL-ORF	129	60.535	39.358	3.465
WINT-ORF	129	69.961	35.844	3.156
SPRING-ORF	129	85.047	39.076	3.440

Means Table for FWS-ORF
Effect: Category for FWS-ORF * Status

	Count	Mean	Std. Dev.	Std. Err.
C, FALL-ORF	23	33.304	24.979	5.208
C, WINT-ORF	23	48.391	26.066	5.435
C, SPRING-ORF	23	61.913	30.542	6.368
G, FALL-ORF	71	76.423	30.663	3.639
G, WINT-ORF	71	84.310	26.471	3.142
G, SPRING-ORF	71	99.831	26.921	3.195
S, FALL-ORF	27	28.481	32.288	6.214
S, WINT-ORF	27	39.444	33.376	6.423
S, SPRING-ORF	27	52.000	38.915	7.489
T, FALL-ORF	8	106.000	42.058	14.870
T, WINT-ORF	8	107.625	37.187	13.147
T, SPRING-ORF	8	131.875	37.722	13.337

3 cases were omitted due to missing values.

Table 6.

Repeated Measures Analysis of Variance Comparing Prosody between General and Special Education Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-PROS

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Status	3	164.094	54.698	28.833	<.0001
Subject(Group)	125	237.136	1.897		
Category for FWS-PROS	2	11.757	5.879	15.526	<.0001
Category for FWS-PROS * Status	6	3.587	.598	1.579	.1537
Category for FWS-PROS * Sbjct(Grp)	250	94.656	.379		

Means Table for FWS-PROS
Effect: Status

	Count	Mean	Std. Dev.	Std. Err.
C	69	2.043	.962	.116
G	213	3.103	.889	.061
S	81	1.827	1.034	.115
T	24	3.958	1.160	.237

3 cases were omitted due to missing values.

Interaction Bar Plot for FWS-PROS
Effect: Status
Means Table for FWS-PROS
Effect: Category for FWS-PROS

	Count	Mean	Std. Dev.	Std. Err.
FALL-PROS	129	2.481	1.288	.113
WINT-PROS	129	2.713	1.002	.088
SPR-PROS	129	2.907	1.114	.098

Means Table for FWS-PROS
Effect: Category for FWS-PROS * Status

	Count	Mean	Std. Dev.	Std. Err.
C, FALL-PROS	23	1.609	.783	.163
C, WINT-PROS	23	2.304	1.020	.213
C, SPR-PROS	23	2.217	.951	.198
G, FALL-PROS	71	2.944	1.081	.128
G, WINT-PROS	71	3.070	.683	.081
G, SPR-PROS	71	3.296	.835	.099
S, FALL-PROS	27	1.593	1.083	.209
S, WINT-PROS	27	1.815	.962	.185
S, SPR-PROS	27	2.074	1.035	.199
T, FALL-PROS	8	3.875	1.553	.549
T, WINT-PROS	8	3.750	.886	.313
T, SPR-PROS	8	4.250	1.035	.366

3 cases were omitted due to missing values.

Table 7.

Repeated Measures Analysis of Variance Comparing Retell Quality between General and Special Education Students at Three Times (Fall, Winter, and Spring)

ANOVA Table for FWS-RETELL

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Status	3	34.450	11.483	7.500	.0001
Subject(Group)	106	162.304	1.531		
Category for FWS-RETELL	2	25.679	12.839	20.888	<.0001
Category for FWS-RETELL * Status	6	9.346	1.558	2.534	.0217
Category for FWS-RETELL * Sbjct(Grp)	212	130.309	.615		

22 cases were omitted due to missing values.

Means Table for FWS-RETELL
Effect: Status

	Count	Mean	Std. Dev.	Std. Err.
C	51	2.020	1.029	.144
G	198	2.343	.999	.071
S	57	1.737	1.027	.136
T	24	3.042	.908	.185

Means Table for FWS-RETELL
Effect: Category for FWS-RETELL

	Count	Mean	Std. Dev.	Std. Err.
FALL-RETELL	110	1.955	.962	.092
WINT-RETELL	110	2.145	1.039	.099
SPR-RETELL	110	2.618	1.040	.099

Means Table for FWS-RETELL
Effect: Category for FWS-RETELL * Status

	Count	Mean	Std. Dev.	Std. Err.
C, FALL-RETELL	17	1.471	.717	.174
C, WINT-RETELL	17	2.353	1.057	.256
C, SPR-RETELL	17	2.235	1.091	.265
G, FALL-RETELL	66	2.015	.868	.107
G, WINT-RETELL	66	2.182	1.036	.128
G, SPR-RETELL	66	2.833	.904	.111
S, FALL-RETELL	19	1.684	1.003	.230
S, WINT-RETELL	19	1.526	.841	.193
S, SPR-RETELL	19	2.000	1.202	.276
T, FALL-RETELL	8	3.125	1.126	.398
T, WINT-RETELL	8	2.875	.835	.295
T, SPR-RETELL	8	3.125	.835	.295

22 cases were omitted due to missing values.

Figure Caption

Figure 1. Scoring procedures for reading measures: Oral fluency, prosody, and retell quality.

Figure 2. Histograms of oral reading fluency for Second and Third grade students in the fall, winter, and spring.

Figure 3. Histograms of prosody for Second and Third grade students in the fall, winter, and spring.

Figure 4. Histograms of retells for Second and Third grade students in the fall, winter, and spring.

Figure 5. Bar charts reflecting growth of oral reading fluency, prosody, and retell for general and special education students from fall to winter to spring.

Figure 6. Bar charts reflecting growth of prosody for general and special education students from fall to winter to spring.

Figure 7. Bar charts reflecting growth of retells for general and special education students from fall to winter to spring.

Figure 8. Comparison of graded and nongraded referrals before program implementation (both schools are graded programs).

Figure 9. Comparison of graded and nongraded referrals after program implementation.

Figure 10. Referral levels in the nongraded program averaged with the 2 years before (1991-92 and 1992-93) and the 2 years after (1993-94 and 1994-95).

Figure 11. Referral levels in the graded program averaged with the 2 years before (1991-92 and 1992-93) and the 2 years after (1993-94 and 1994-95).

Figure 12. Current referral rates in both nongraded and graded programs.

Figure 13. Placement levels for both nongraded and graded programs.

Oral Reading Fluency

think ~~thick~~	**misidentification** (student decodes word incorrectly). Slash word and, if possible, write word student said.
she had smiled	**omission** (student leaves out word). Circle word omitted
H ~~left~~	**hesitation** (student doesn't decode word within 5 sec.) Tell student word and mark H over the word.
mm ~~mother~~	**word substitution** (student uses word or similar meaning). Slash word and write word substituted.
saw he said	**reversal** (student says "was" for "saw" or "said he" for "he said"). Mark transposed part with a loop.
SC ~~dog~~	**self-correct** (student says dot, then self-corrects and says dog). Write SC or C over the word.
once he ^ said	**insertion** (student adds word). Mark a caret and write in word added.
he was very	**repetition** (student repeats word or phrase more than once). Underline word or phrase repeated with wavey line.

Prosody Key

(1) Tele = Reads single words. No "flow." Telegraphic-like in sound. Word-by-word reading. Word calling.

(2) 2-3 Word = Some phrasing is noted (2-3 words). Very hesitant reading. Considerable pausing and drawn out blending.

(3) End/No Inflec = Pauses for ending punctuation. Inflection changes may not be present. Reading in phrases but missing the tone.

(4) Flow Most = Appropriate "flow" and phrasing is noted as well as attention to punctuation with pauses and appropriate inflection MOST of the time.

(5) Flow/Punc/Inflec = Reading generally "flows." Voice changes to reflect meaning changes. Appropriate ending inflections. Fluent and smooth.

Overall Storytelling Five-point scale for retell storytelling

5 points
- Complete retell; includes both the main idea and supporting detail.
- Retell is substantially in the same order as the story.
- Much elaboration, minor details that embellish the story are included; reporting of finer points that do not necessarily support main idea, but enrich the context of the story.

4 points
- Gist of the story is present and the retell generally is in same sequence as story.
- Substantial elaboration (more so than 3-point response); most detail is present.
- May contain some errors, but they are very minor (i.e., use of different adjectives).

3 points
- More details than 2-point retell, but generally following the same guidelines as the 4-point score, with the modifications noted in the 2-point score.
- May contain brief discussion of major points, but no supporting detail or elaboration.
- May be incomplete (does not mention all major points, but does include substantial elaboration of points mentioned).
- May contain some minor inaccuracies.
- May be reported out of sequence.

2 points
- May include a brief discussion of major points, but no supporting detail or elaboration.
- May be incomplete (does not mention all major points, but does include some elaboration of points mentioned.
- May be incomplete, but has gist of story (unlike 1-point retell).
- May contain some minor inaccuracies.
- Significant inaccuracy coupled with very brief response turns a 2-point response into a 1-point response.

1 point
- A few brief statements about the story.
- Contains some inaccuracies.
- May focus on just one aspect of the story.
- Not sequenced, nonsensical.

Figure 2

Histograms of oral reading fluency for Second and Third grade students in the fall, winter and spring

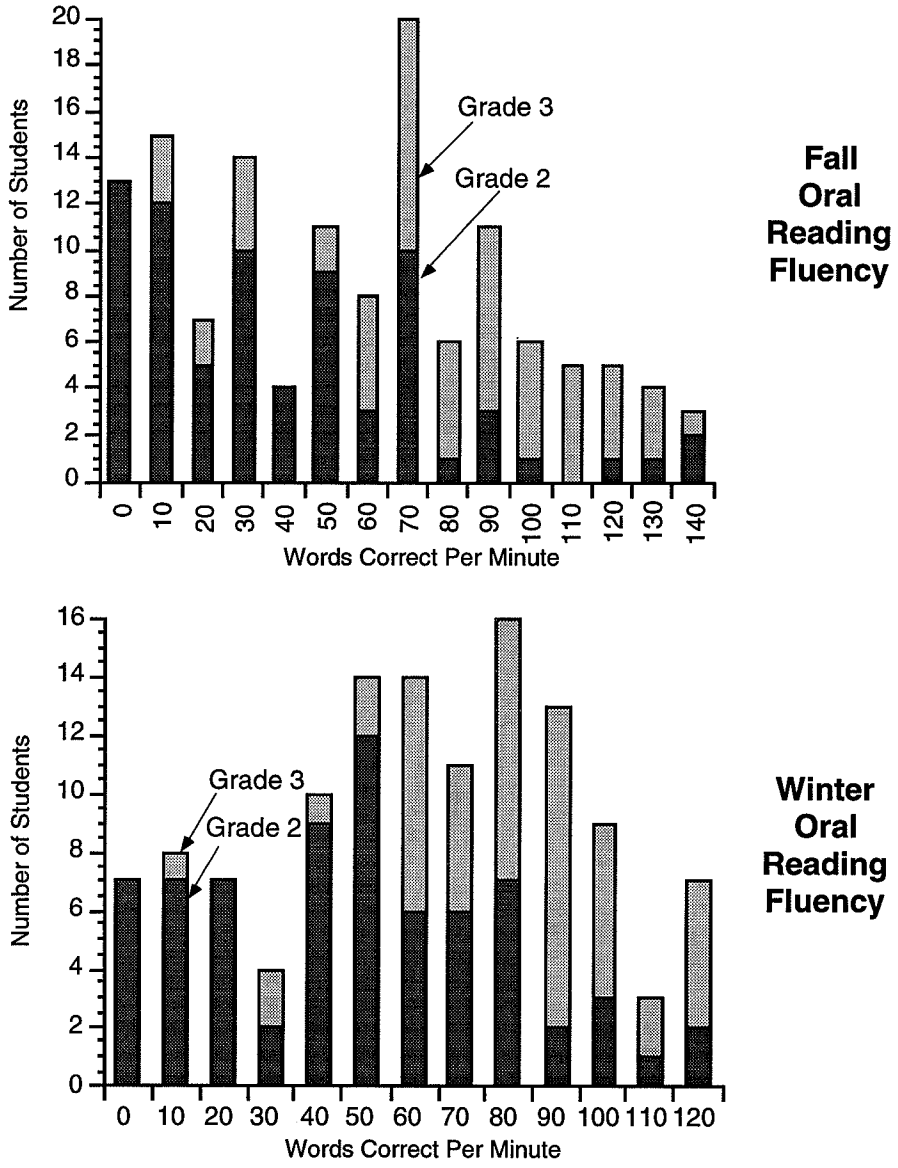

Figure 3

Histograms of prosody for Second and Third grade students in the fall, winter, and spring

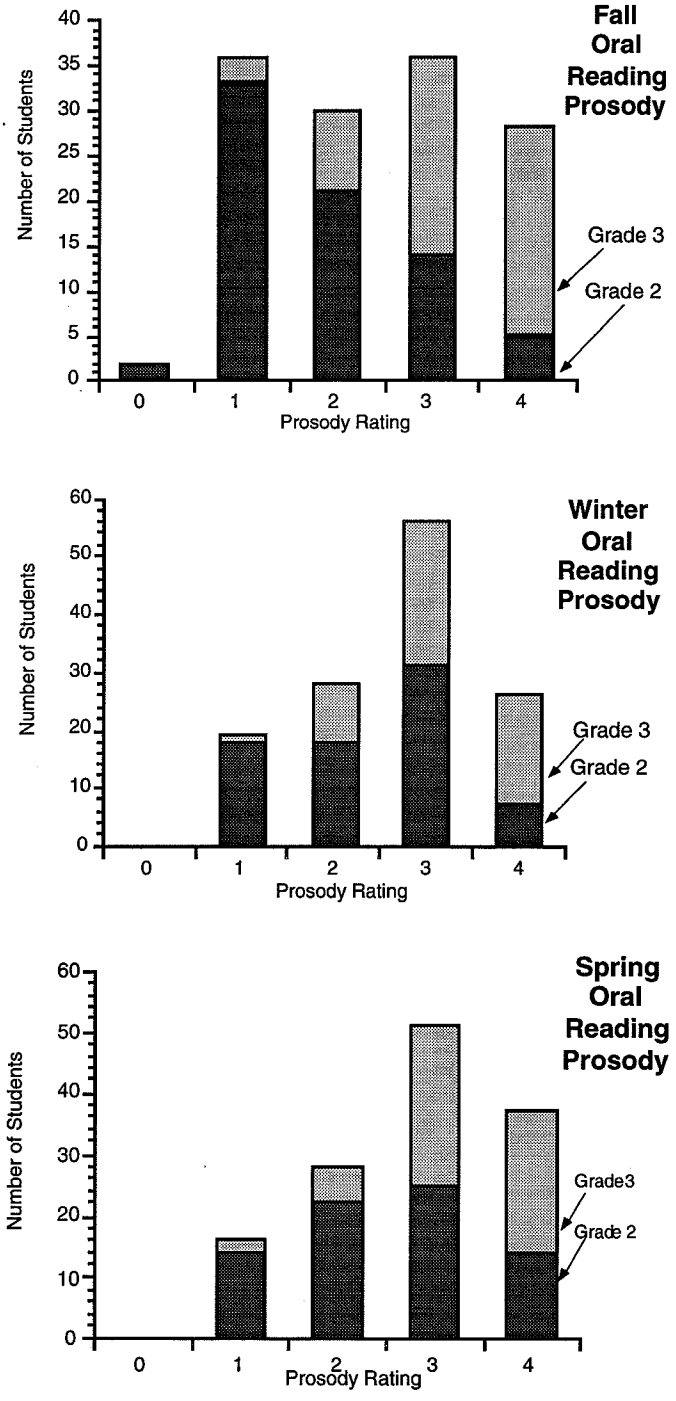

Figure 4

Histograms of retells for Second and Third grade student in the fall, winter and spring

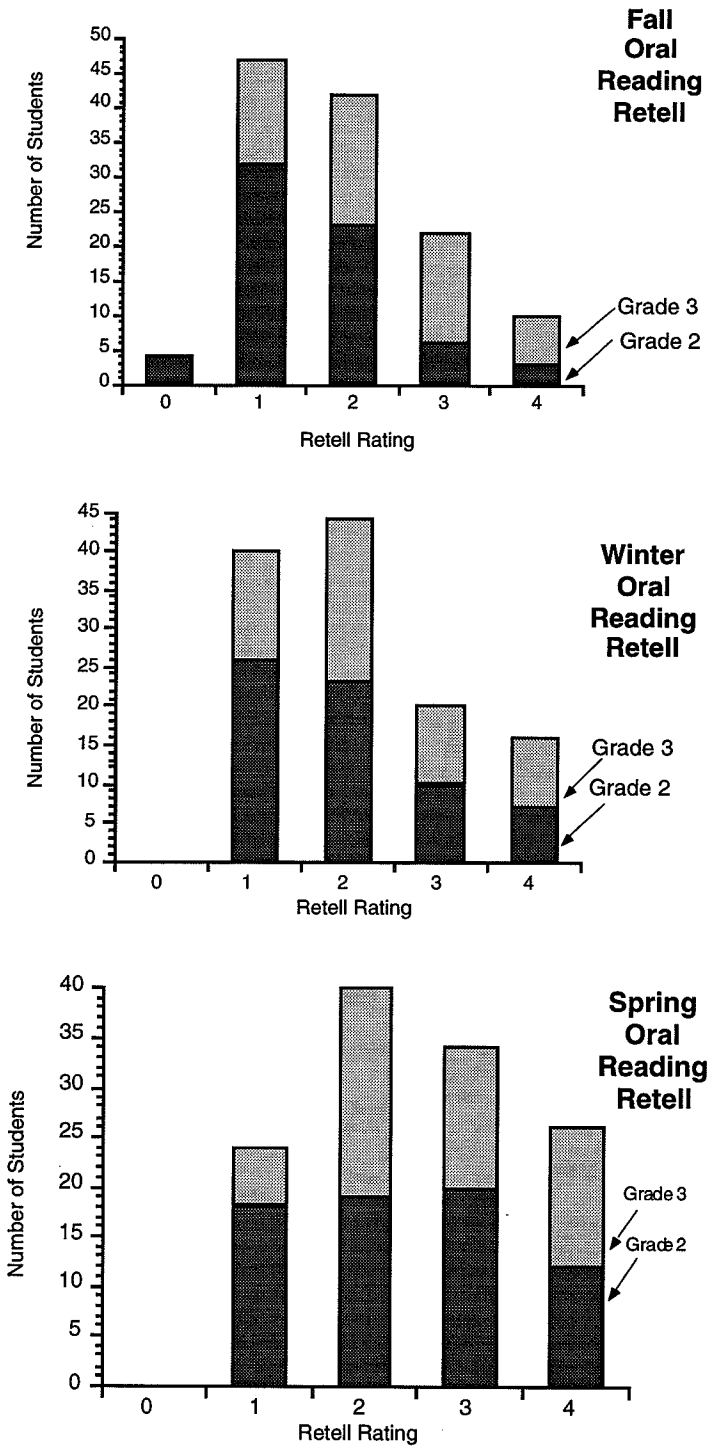

Figure 5

Bar charts reflecting growth of oral fluency, prosody, and retell for general and special education students from fall to winter to spring

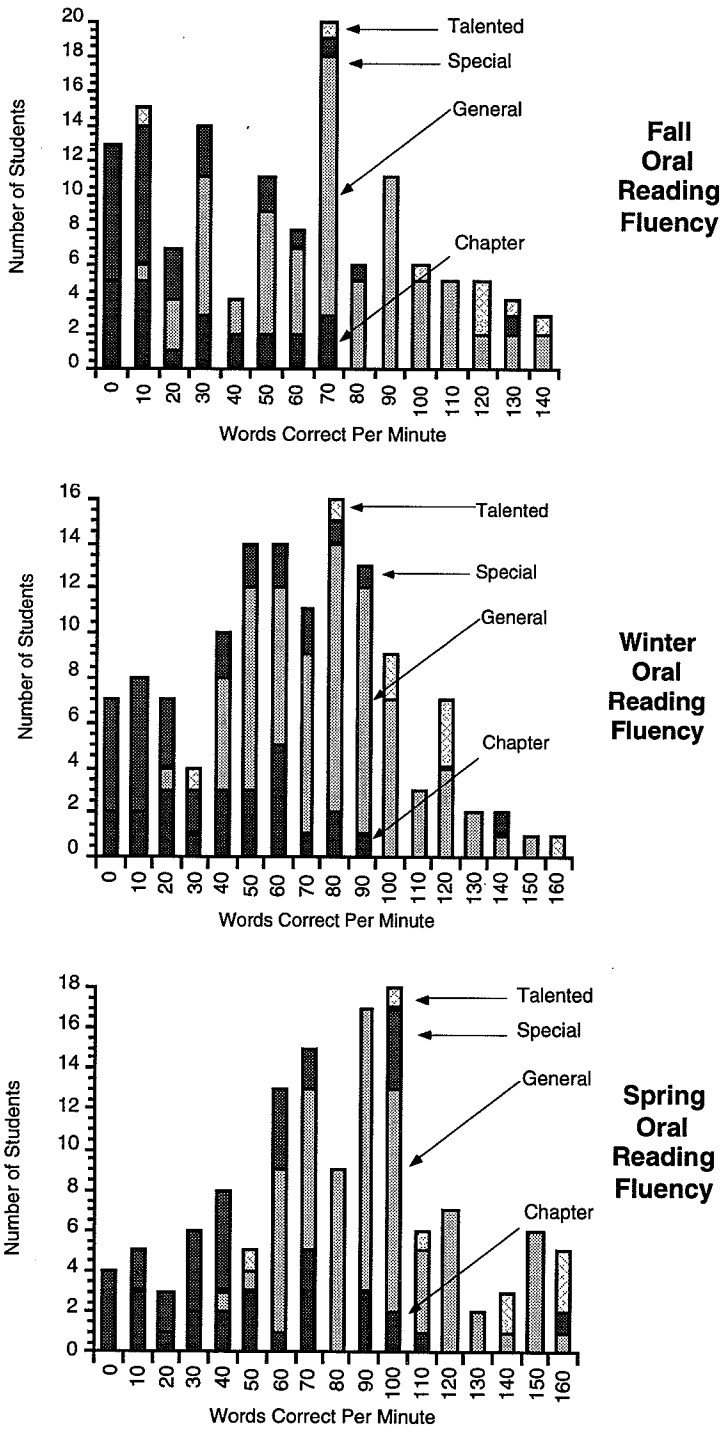

Figure 6

Bar charts reflection growth of prosody for general and special education students from fall to winter to spring

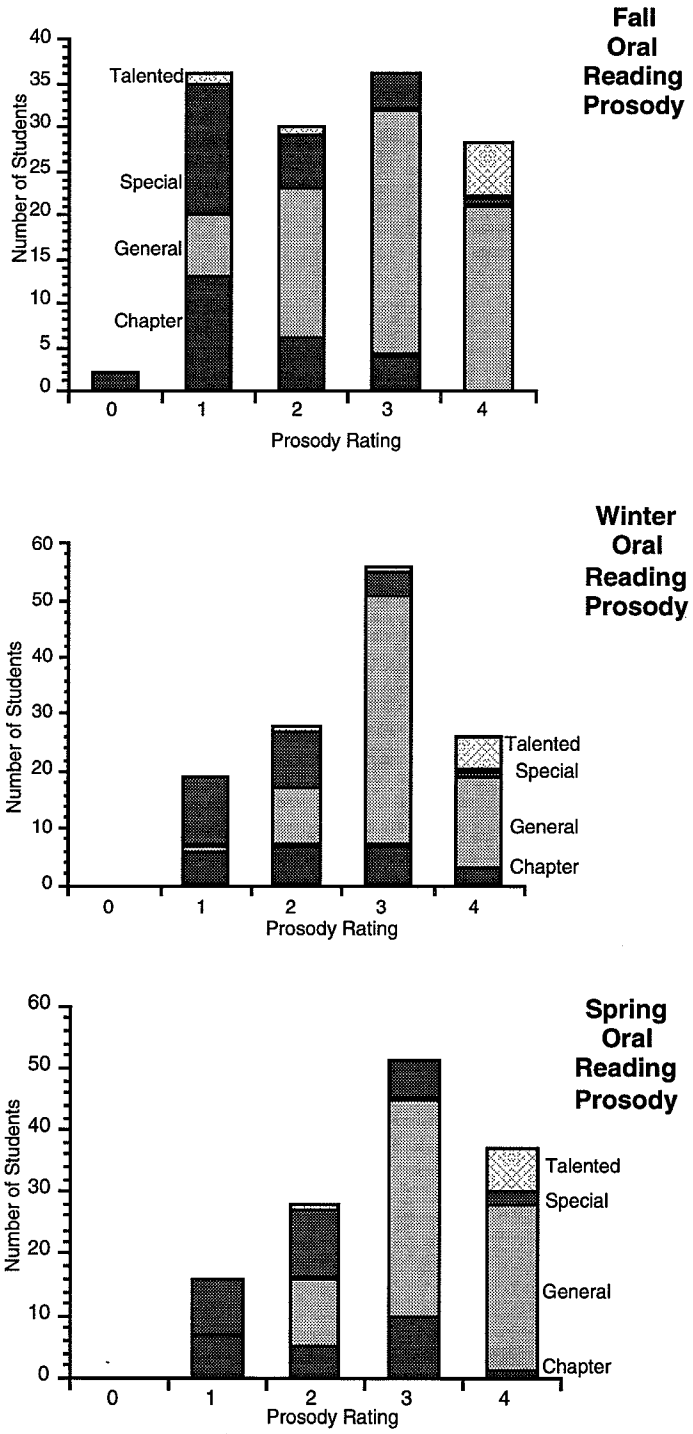

Figure 7

Bar charts reflecting growth of retells for general and special education students from fall to winter to spring

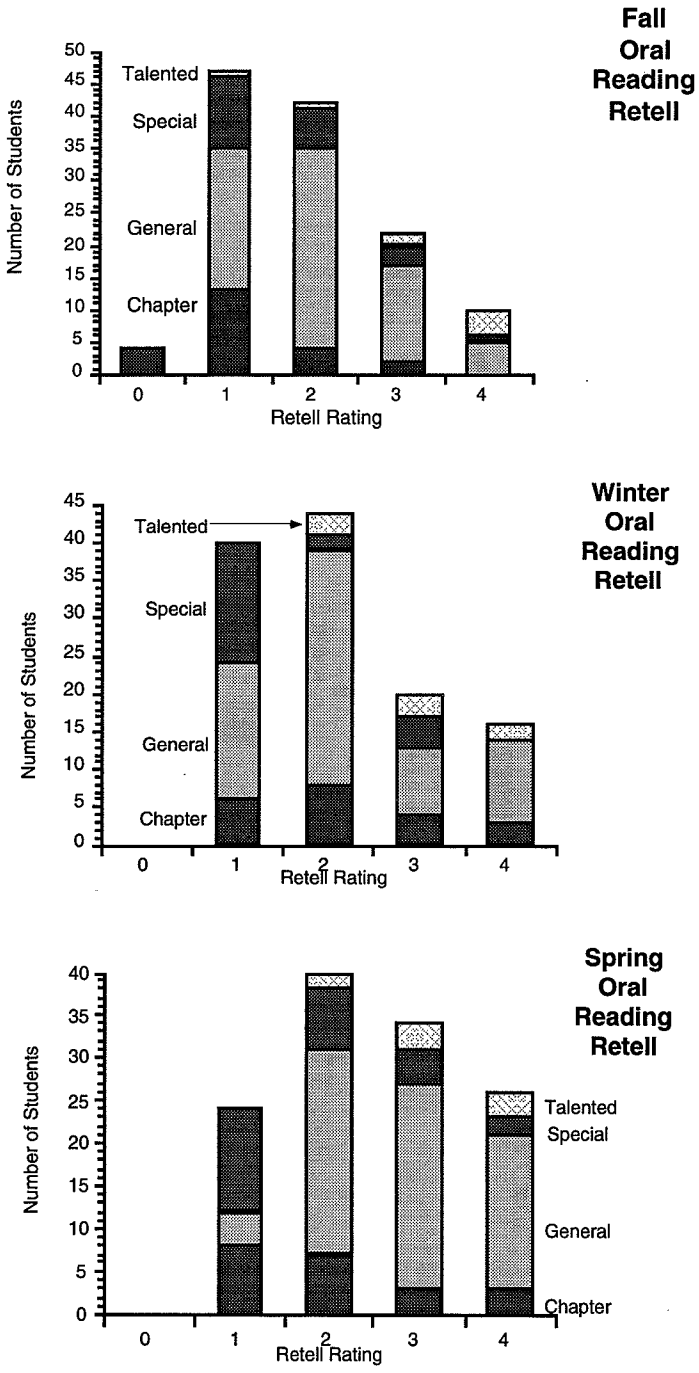

Figure 8

Comparison of graded and nongraded referrals before program implementation (both schools are graded programs).

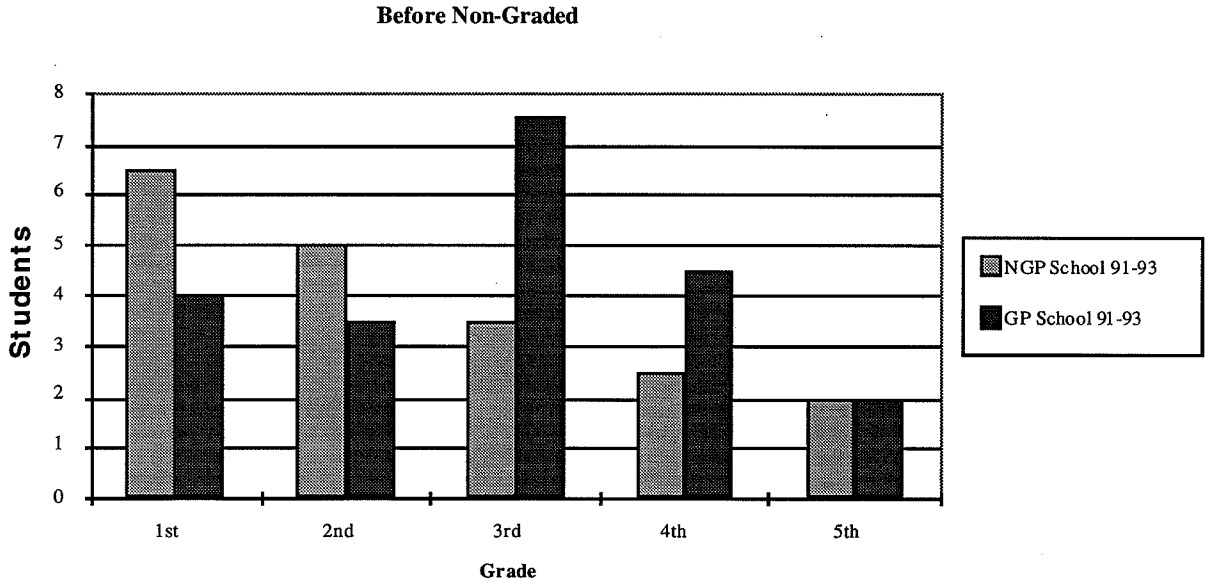

Figure 9

Comparison of graded and nongraded referrals after program implementation.

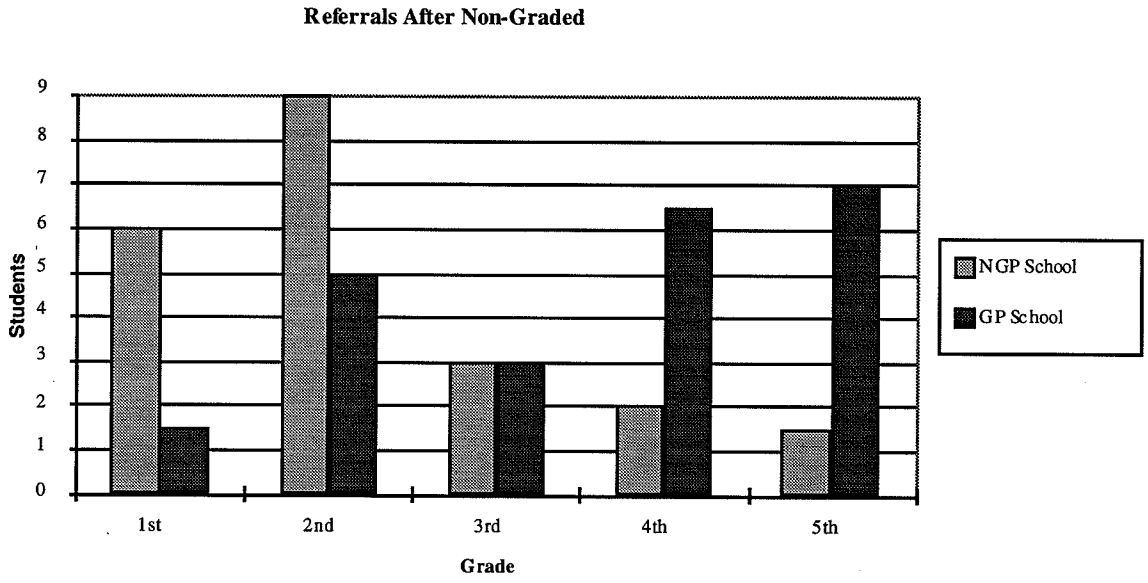

Figure 10

Referral levels in the nongraded program averaged with the 2 years before (1991-92 and 1992-93) and the 2 years after (1993-94 and 1994-95).

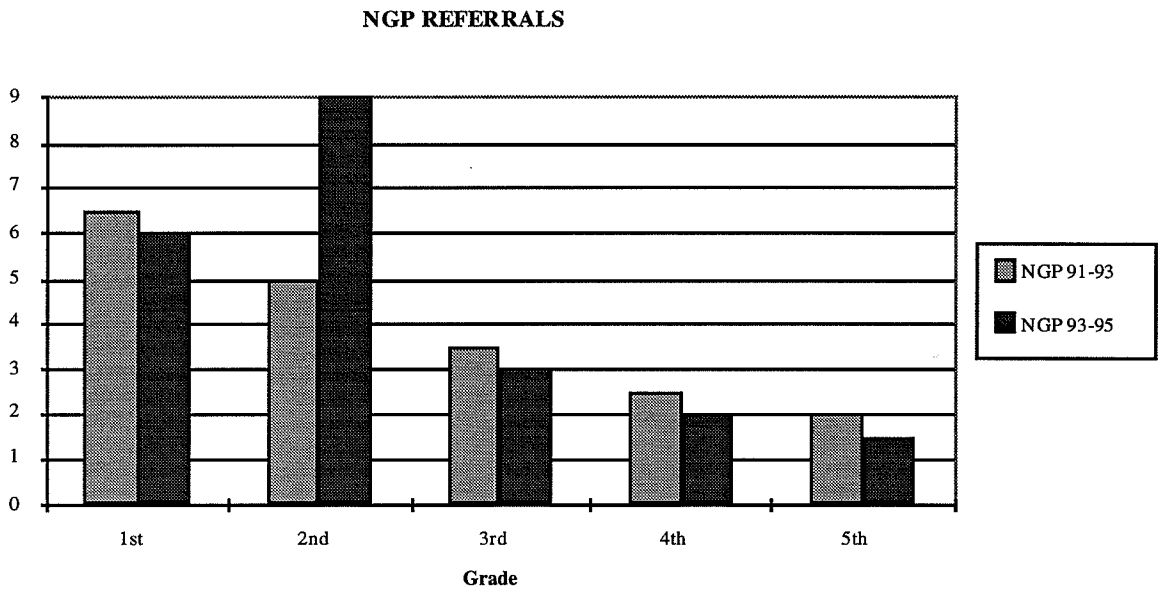

Figure 11

Referral levels in the graded program averaged with the 2 years before (1991-92 and 1992-93) and the 2 years after (1993-94 and 1994-95).

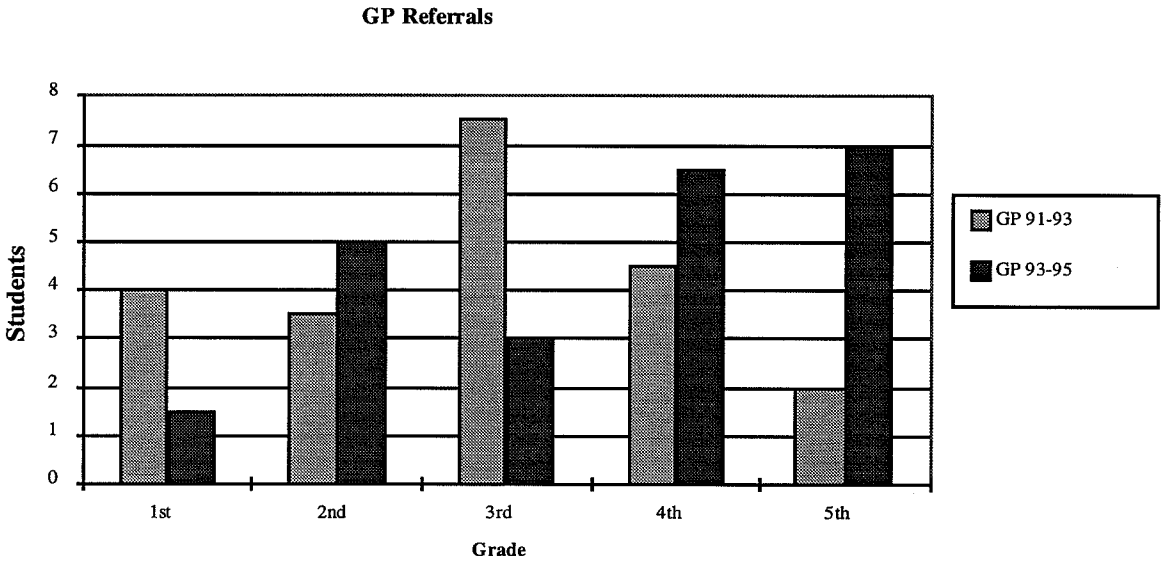

Figure 12

Current referral rates in both nongraded and graded programs.

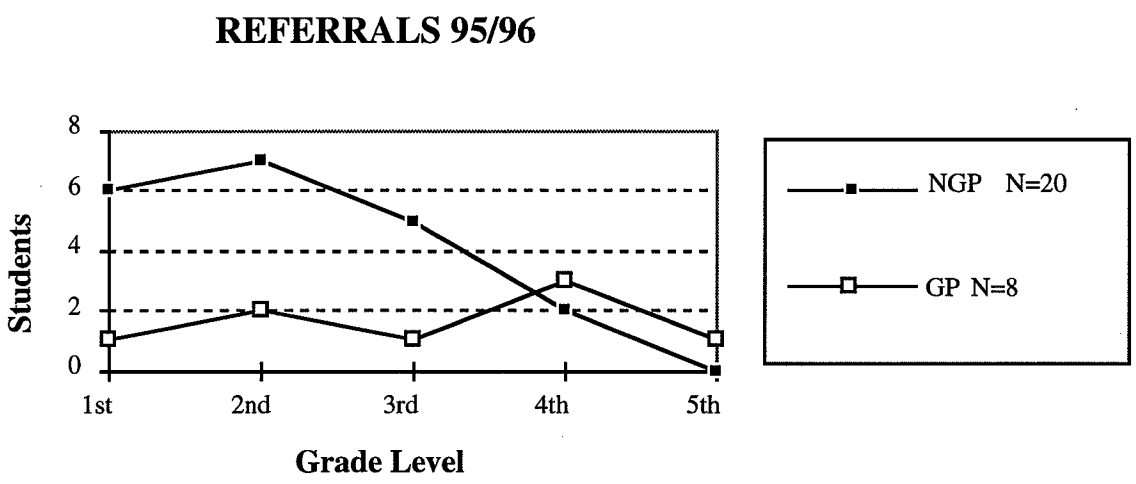

Figure 13

Placement levels for both nongraded and graded programs.