An Investigation of the Effects of Total School Flexible Cluster Grouping on Identification, Achievement, and Classroom Practices

Marcia Gentry and Steven V. Owen
Gifted Child Quarterly 1999 43: 224
DOI: 10.1177/001698629904300402

An Investigation of the Effects of Total School Flexible Cluster Grouping on Identification, Achievement, and Classroom Practices

Marcia Gentry
Minnesota State University—Mankato

Steven V. Owen
University of Texas Medical Branch

ABSTRACT

This paper presents the findings of a longitudinal, causal comparative investigation of an elementary school cluster grouping program. Both quantitative and qualitative methodologies were used. Although the cluster grouping program was originally designed to provide differentiation of content and instruction for gifted students, positive effects were also found on the achievement of all students in the school. During the three program years, students involved in the school using cluster grouping were more likely to be identified as high achieving or above average. Fewer students were identified as low achieving. A significant increase in achievement test scores of all students was found when these students were compared to similar students from a comparison school district. Qualitative analyses yielded three core categories—the use of grouping, the impact of teachers, and the general school environment—that helped to provide an understanding of the quantitative findings.

Background

Cluster grouping is a widely recommended and often used strategy for meeting the needs of high achieving[1] students in the regular elementary classroom. Its use has gained popularity in recent years because of the move toward inclusive education, budget cuts, and heterogeneous grouping policies that have eliminated programs for gifted students (Purcell, 1994). However, little research exists on the effects of using cluster grouping on the achievement of gifted students, and none exists that has examined its effects on students of other achievement levels (Gentry, 1996; Hoover, Sayler, & Feldhusen, 1993).

Gentry (1996) noted many variations in definitions and applications of cluster grouping and identified three common themes in the existing literature on this topic. First, a group of students (varying in number from 3 to more than 10) identified as gifted, high achieving, or high-ability is placed into a classroom with students of other achievement levels. Second, the reason for cluster grouping is to differentiate curriculum.

PUTTING THE RESEARCH TO USE

This research describes the implementation of a total school application of cluster grouping over time with two entire graduation classes of students. It provides a rich example of the effects that a gifted program can have on an entire school when that program is integrated with the general education program and considers the needs of all students and teachers. It reinforces the notion that grouping, when done flexibly and with appropriately adapted curriculum and instruction, can help students of all achievement levels grow academically while assisting teachers in their efforts to better meet the individual needs of their students. Based on the findings of the study, the implications for practice include: (1) professional development in gifted education should not be restricted to just those teachers responsible for students identified as gifted because the use of gifted education "know how" has the potential to improve general education practices; (2) placing a cluster of high achievers in one classroom can increase the chance that their needs will be met while offering the opportunity for talent to emerge in the other classrooms; and (3) unlike suggestions by many reformers, the elimination of ability grouping may not be beneficial to students and teachers.

Third, the teacher of the high ability cluster should have background, training, and experience in working with gifted students.

Current research indicates that there are several major benefits of cluster grouping: Gifted students regularly interact with their intellectual peers *and* age peers (Delcourt & Evans, 1994; Rogers, 1991; Slavin, 1987a); cluster grouping provides full-time services for gifted students without additional cost (Hoover et al., 1993; LaRose, 1986; Winebrenner & Devlin, 1994); curricular differentiation is more efficient and likely to occur when a group of high-achieving students is placed with a teacher who has expertise, training, and a desire to differentiate curriculum than when these students are distributed among many teachers (Bryant, 1987; Kennedy, 1995; Kulik, 1992; Rogers, 1991); removing the highest achievers from most classrooms allows other achievers to emerge (Kennedy, 1989; Winebrenner, 1992); and, finally, cluster grouping reduces the range of achievement levels that must be addressed within the classrooms of all teachers (Coleman, 1995; Delcourt & Evans 1994; Rogers, 1993). Conversely, the literature reveals several concerns about the use of cluster grouping, and these concerns parallel those raised regarding the use of ability grouping in general. These include the effect that removing the brightest students from classrooms has on the students and teachers in these classrooms (Hoover et al. 1993; Oakes, 1985; Slavin 1987a), the methods for selecting teachers for the high-achieving cluster classroom (Oakes; Slavin, 1987b), and whether cluster grouping provides appropriate differentiation for the high-achieving students (Delcourt & Evans 1994; McInerney, 1983; Rogers, 1991; Westberg, Archambault, Dobyns, & Salvin, 1993).

Recent studies that examined cluster grouping include survey research conducted by Hoover et al. (1993) and two studies by Delcourt and her colleagues (Delcourt, Loyd, Cornell, & Goldberg, 1994). Hoover et al. reported that classroom teachers believed that cluster grouping benefited both gifted and nongifted students. However, these researchers also concluded that "despite clear potential benefits of cluster grouping, there have been no empirical studies of the prevalence of cluster grouping nor of its effects, perceived or actual, on gifted children" (p. 13). Delcourt et al. examined four programming arrangements for gifted students,including special schools, separate classes, pull-out programs, and within-class programs, and their effects on achievement and affective outcomes. Of the 11 districts included in the study, one used cluster grouping, which was classified as a within-class program. However, across all programs, gifted students from within-class programs received the lowest scores in all areas of achievement when compared to their gifted peers in the other programming options. Delcourt et al. concluded that "since within-class programs are a popular model in gifted education, their curricular and instructional provisions for the gifted must be carefully maintained lest they disintegrate into a no program format" (p. 77). Yet, in a follow-up study (Delcourt & Evans, 1994) that examined exemplary programs in gifted education, a school using cluster grouping was selected as the best example of a within-class program. Key variables that distinguished these exemplary programs were leadership, atmosphere and environment, communication, curriculum and instruction, and attention to student needs. In addition, the exemplary programs were found to influence student achievement and motivation through exposure to challenge and choices. The extent to which these themes are evident within a cluster grouping program may help explain both its success and its impact on student achievement.

Rogers (1991) recognized that the research base on cluster grouping is limited and cautioned that the cluster teacher must be trained and motivated to work with gifted and talented students and that the curriculum must be appropriately differentiated. In a meta-analysis of the research on ability grouping, Kulik (1992) found that youngsters of all achievement groups benefited from ability grouping programs when the curriculum was appropriately adjusted to the aptitude levels of the groups. As a result, he recommended that schools use various forms of flexible ability grouping.

Although many experts advocate the use of cluster grouping (Balzer & Siewert, 1990; Brown, Archambault, Zhang, & Westberg, 1994; Coleman, 1995; Davis & Rimm, 1985; Hoover et al., 1993; Juntune, 1981; Kaplan, 1974; Kulik & Kulik, 1991; LaRose, 1986; Renzulli, 1994; Rogers, 1991; Winebrenner, 1992), surprisingly little evidence exists regarding its effectiveness. Clearly, research that can provide evidence about the effects of cluster grouping on students is needed.

Research Questions

This study examined the use of cluster grouping during a four-year period in a small, rural school district in the Midwest. The following research questions guided the study:

1. Is cluster grouping related to teacher perceptions of student achievement as measured by teacher identification categories?
2. How do students in the cluster grouping school compare with students from a similar school who are not involved in cluster grouping after adjustment for initial differences with regard to achievement?
3. What factors exist within the classrooms and the school using cluster grouping that may influence student achievement?

Background of the Treatment Program

Schools, classrooms, teachers, and students are complex, interactive entities, making their study challenging at best, and the results from such study ambiguous. Yet, programs must be studied in their full context to provide insight into their workings and their possible effects. It is from in-depth examination of real programs in real schools that the opportunity to learn about schools presents itself. It is for these reasons that this study examined the existing use of cluster grouping in a small, rural school district, purposefully selected because of its innovative use of cluster grouping with students of all achievement levels in all classrooms.

Cluster grouping in the treatment district began in grade 3 and continued through grade 5, with a flexible identification process beginning at the end of second grade that included information from teachers, parents, and achievement tests. Teachers were involved with the identification and placement of students into the classrooms, which was done using grade level conferencing. Each year in May, second-, third-, and fourth-grade teachers

1. rated their students' academic performance as *high-achieving*, *above-average*, *average*, *low-average*, or *low* (students' academic performance as observed by the teacher and the Scales for Rating the Behavioral Characteristics for Superior Students (Renzulli, Smith, Callahan, White, & Hartman, 1977) functioned as a basis for these ratings);
2. indicated those students who received special education or Chapter 1 services; and
3. noted students who had behavior problems or who should be separated.

Teacher ratings were compared with achievement scores on the Iowa Tests of Basic Skills (Hieronymus, Hoover, & Lindquist, 1984), and discrepancies were discussed. By using both teacher ratings and achievement scores, it was possible for a student who tested poorly or for a student whose classroom performance did not reflect his or her ability to be identified as *high-achieving* or *above-average* without the use of *cutoff* scores. This system of checks and balances is similar to that suggested by Renzulli and Reis (1985) in the Revolving Door Identification Model. Although there were no cutoff scores for identification, the process was done consistently by the same teachers on a yearly basis.

There were five classrooms per grade level in which students were placed yearly on the basis of their identification categories. One classroom had the cluster of *high-achieving* students, with the remainder of the class comprised of *average*, *low-average*, and *low-achieving* students. The other four classrooms each had students who achieved at *above-average, average*, *low-average*, and *low* levels. Additionally, two of these classrooms had clusters of special needs students who received Chapter 1 or special education assistance. In each of these two rooms, an aide or a teacher-consultant worked with the classroom teacher for the majority of the day. By arranging classes in this manner, each heterogeneous classroom had a group of *above-average* achieving students, but one class had the specific cluster of *high-achieving* students. In this way, the use of resource personnel was maximized. Behavioral problems—from all achievement levels—were evenly distributed among the five classrooms.

When the cluster grouping program was adopted, all teachers were provided with a general overview of gifted education and talent development based on the Schoolwide Enrichment Model (Renzulli & Reis, 1985) and were involved in two half-day in-service training sessions regarding the above described approach to cluster grouping. Annual in-services in gifted education (e.g., curriculum compacting, curricular and instructional differentiation, and thinking skills) and opportunities to attend regional, state, and national conferences on gifted education were made available to all teachers. The teachers responsible for teaching the *high-achieving* cluster volunteered and were selected by the staff and administration. Each of these teachers took classes in gifted education and attended several workshops to improve their methods for working with high-achieving students. It should also be noted that, as in any school, cluster grouping was not the only type of grouping or treatment that occurred. In fact, in this school, there were a variety of grouping arrangements that took place, including regrouping between classes for math and reading. Also, because of the increased number of students who were identified as high-achieving, a class of these students existed by fifth grade. The complexity of the grouping arrangements made it impossible to analyze the relative effects of each arrangement on student achievement. For further discussion of the treatment program, including its philosophy and practices, refer to Gentry (1996).

Methods and Procedures

Research Design

The research design was causal-comparative and longitudinal, employing both quantitative and qualitative methodologies. The first two research questions were addressed using descriptive and inferential statistics, and the third research question was addressed with qualitative methods. The combination of quantitative and qualitative methods allowed a more thorough description of how cluster grouping was imple-

Table 1

Demographic Factors Upon Which Treatment and Comparison Schools Were Matched

Factor	Treatment School	Comparison School
Geographic Region	Rural Midwest	Rural Midwest
Ethnic composition	White, < 1% minority	White, < 1% minority
Student Population*	1,499	1,202
Socioeconomic status*	Low	Low
School Configuration	1 elementary school K–5 5 classes/grade level	1 elementary school K–6 4 classes/grade level
Pupil to teacher ratio*	20:1	21:1
Per pupil revenue*	$3,704	$4,071
Rank in state for spending on basic needs programs* (out of 524 districts)	503rd	491st

Note. **Source: 1992–93 Bulletin 1014 (Michigan Department of Education, 1994).*

mented within the complex context of a real school. Although it was not possible to isolate the effects of a single variable—cluster grouping—this study provided a realistic picture of how cluster grouping worked in concert with other variables found within schools. Causal-comparative research is done after the fact using existing data, and it does not seek to attribute causality; rather it seeks to establish relationships and trends from which future research can be conducted.

Sample

Purposive sampling was used in this study. The treatment sample included all students from two graduation class years who attended the school from grades 2 through 5 (Class of 2000: $n = 97$; Class of 2001: $n = 100$). The comparison school was selected based on its demographic similarity to the treatment school (see Table 1) and because its students had not been involved in cluster grouping or gifted programming (Class of 2000: $n = 68$; Class of 2001: $n = 69$). Any students for whom achievement data were unavailable for grades 2, 3, 4, and 5 were eliminated from the analyses. The Classes of 2000 and 2001 were selected because longitudinal data could be obtained from both the treatment site and comparison site to compare the students' academic achievement.

The sample also included teachers and administrators from the treatment site who were involved in the program. Follow-up interviews were conducted with 14 of 15 grade 3–5 teachers and with three of five administrators who were originally involved with the program and were involved for the entire time during which the program took place.

Instrumentation

To examine student achievement effects, the present study used existing achievement data from both the treatment and comparison schools. Normal curve equivalent (NCE) scores were collected for each student (grades 2–5) in the areas of total math and total reading from standardized achievement measures used by the schools. The treatment school used the Iowa Tests of Basic Skills (ITBS), Form G (Hieronymus, Hoover, &

Lindquist, 1984), while the comparison school used the California Achievement Test (CAT), Form E (1984) to measure yearly student achievement. Because of the ex post facto nature of this study, available instrumentation was used. Airasian (1989) stated that the CAT "compares very favorably to other achievement batteries of its genre such as . . . the *Iowa Tests of Basic Skills*" (p. 128). Thus, although the content of these two standardized tests was not identical, the NCE scores provided an achievement standing relative to the respective test's norm in a group and allowed comparison in achievement to be made on the basis of normed scores.

To address Research Question 3, a semi-structured interview protocol was developed based on themes identified by Delcourt and Evans (1994) (leadership, atmosphere and environment, communication, curriculum and instruction, attention to student needs) and factors identified by Westberg et al. (1993) (questioning and thinking, providing challenges and choices, reading and written assignments, curriculum modifications, enrichment centers). Interviews were taped and transcribed.

Analyses

BMDP statistical software (Dixon, 1992) was used to screen and analyze the data (one outlier was eliminated from the Class of 2001: Mahalonobis D-squared distance value $p < .0003$). Descriptive statistics were used to address Research Question 1 and inferential statistics—including multivariate repeated measures ANCOVAs and planned contrasts—were used to address Research Question 2. Although discriminant function analyses is a preferred follow-up for MANOVA and MANCOVA, it cannot be used with repeated measures; therefore, univariate ANOVA and ANCOVA were used to examine the multivariate main effects. Grade 2 NCE scores in math and reading were used to adjust the groups for initial differences. Separate analyses were run for each graduation year (Class of 2000, Class of 2001). Assumptions for the analyses for each research question were examined (namely normal distribution, homogeneity of variance, and sphericity), and no violations were found.

To address Research Question 3, data from interviews with teachers and administrators ($n = 17$) and documents were gathered, and qualitative procedures were employed (Spradley, 1980). Together with the quantitative findings, the interview transcriptions enabled triangulation of data, a technique that provides checks for both reliability and validity of data through the comparison of multiple sources and data collection methods (Mitchell, 1986). Interview transcriptions and document reviews were coded and analyzed for patterns and themes (Strauss & Corbin, 1990). Trustworthiness was enhanced by using a "devil's advocate," triangulating the data, and checking and questioning the data.

Results and Implications

Identification Findings and Implications for Research Question 1: Is Cluster Grouping Related to Teacher Perceptions of Student Achievement as Measured by Teacher Identification Categories?

Descriptive statistics provided insight into the identification of students in the treatment school during the three program years. Overall for both data sets (Class of 2000 and Class of 2001), more students were identified as *high achieving* each successive year, while fewer students were identified as *low achieving*. By fifth grade, each of these classes had one entire classroom of students identified as *high achieving*, yet all other classrooms still contained groups of students identified as *above average*. Figure 1 depicts the changes in the number of students identified as *high achieving* for both data sets during the three program years, and Figure 2 depicts changes in the frequencies of students identified as *low achieving* during the three program years.

Changes in students' identification categories during the three program years were classified as *increased*, *decreased*, *no change*, or *varied*, and these changes were tabulated. *Increased* was defined as moving up, for example, from *average* to *above average*, and *decreased* was defined as moving down, for example, from *high achieving* to *above average* during the course of three years. No change was used to describe those students whose identification category remained constant for each of the three program years. Students whose identification category changed, but did not increase or decrease as described above, were counted as *varied*. A large percentage of students' identification categories in both classes increased (Class of 2000: 47%; Class of 2001: 34%),or saw no change (Class of 2000: 40%; Class of 2001: 45%), whereas only a small number of students' identification categories decreased (Class of 2000: 3%; Class of 2001: 9%). Students in this program were regarded by their teachers as higher achievers as they progressed from third to fifth grade in the program, a result that led to the analyses of achievement scores.

Qualitative follow-up to these findings yielded interesting results that might explain the trend of identifying more students achieving at higher levels during the course of the three program years. Many teachers ($n = 13$; 93%) and all administrators ($n = 3$) believed that the increase in the number of students identified at higher levels was directly related to the grouping practices used in this school. For example, as Teacher 4C explained:

> Maybe cluster grouping has a lot to do with it. The cluster grouping may give the lower achieving students more self-confidence, because I think they become more involved in class when the high [achieving] kids are removed. And you know that those high kids are competitive

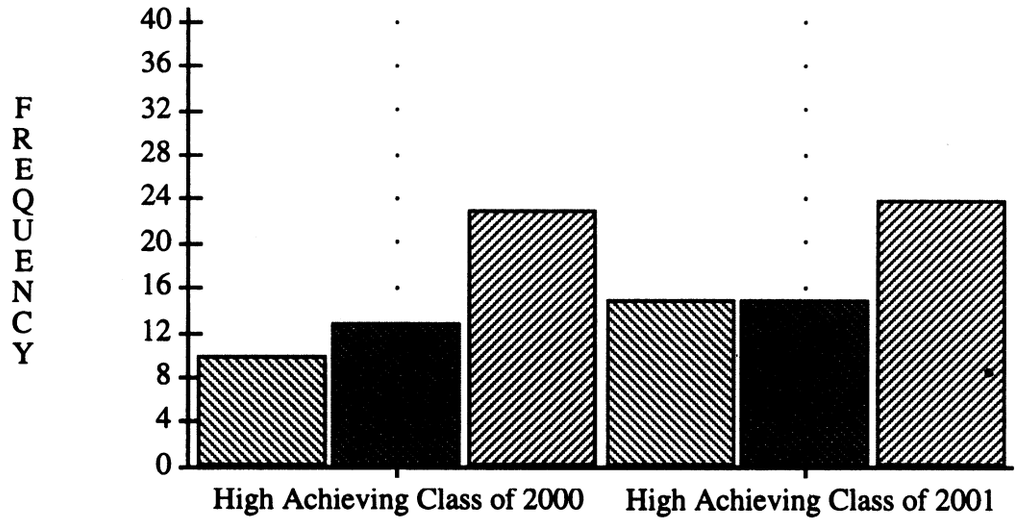

Figure 1. Changes in High Achievement Identification From Grade 3 to Grade 5 for Students in the Class of 2000 and the Class of 2001

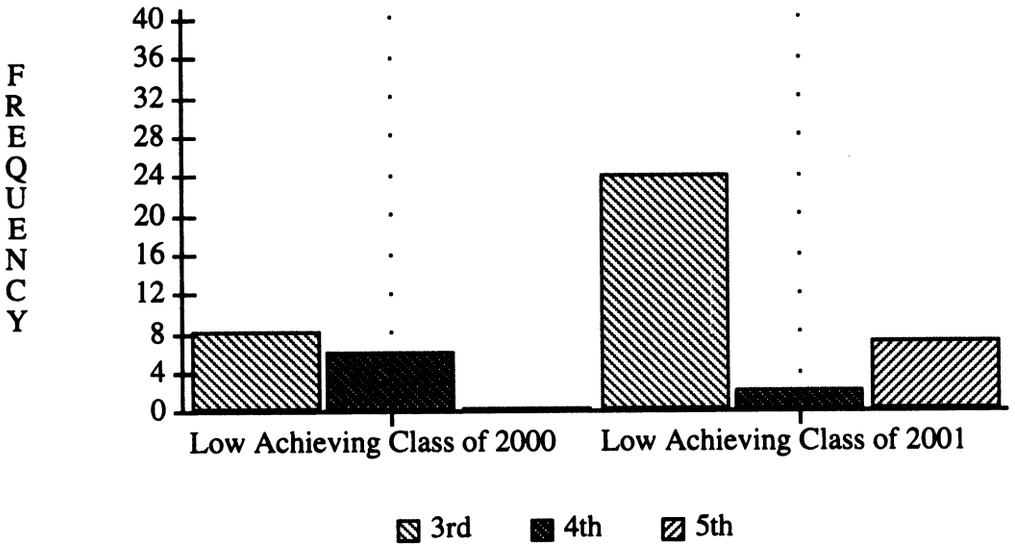

Figure 2. Changes in Low Achievement Identification From Grade 3 to Grade 5 for Students in the Class of 2000 and the Class of 2001

and tend to dominate class sometimes. Also, the average student or high-average student really blossomed, too, which may be due to cluster grouping.

From the perspective of Teacher 3E:

> We've talked about why we find more higher achieving students for several years. Part of it, I feel, is that when you pull those really high kids out—those who always have their hand up first and jump in with the answers—when you get rid of those students by putting them together in a cluster classroom—the other kids have a chance to shine. They take risks more often and see themselves as being leaders of the group. They are no longer frightened to offer answers.

As Teacher 3A discussed:

> I think the low and average children really benefited, because we only spent as much time on things as they needed to and then we moved on. Even if they moved at a slower pace, they were feeling successful. I feel even the low students had good self-esteem because they were constantly successful. I think that's why we also had fewer low students. There's a fine line between high average and high achieving,

> and I think a child who is in a classroom where there are not other children taking front stage has that opportunity and starts to shine. Their confidence builds, and I think that turns a high-average child into a high-achieving child.

The teachers in this study believed that removing the highest achievers from four of the five classrooms gave other students the opportunity to grow and achieve at higher levels than they might have if the highest achieving students had remained in the classroom. This result is consistent with the findings of Kennedy (1989), who found that when the gifted students were absent from the regular classrooms, new talent emerged from other students.

Additionally, other teachers (n = 8; 64%) and administrators (n = 2; 67%) suggested that the increase in achievement was due to efforts by the teachers to facilitate achievement among all of their students. These efforts included maintaining high expectations, creating a positive learning environment, and using a variety of strategies to challenge individual students. As Teacher 5A explained:

> One thing that caused more students to be identified was our expectations. I think that when kids are expected to achieve at a higher level, they try to do that. And I think that high expectations help students to try, and this effort boosts their scores. I think when students are exposed to higher level thinking skills and challenging work, it helps them achieve. When they are with other kids who are working at high levels, I think that helps them. I know I found that with the cluster grouping.

Teacher 5C described her thoughts:

> The high students were all with [Teacher 5A], and we expected more [from the students] we had. By removing some of the higher kids, it may have influenced the others to work harder ... and maybe teachers expected more because we didn't have the higher students and treated it as a regular classroom and expected the average students to rise to the occasion.

Summary of Findings for Research Question 1

A combination of grouping and teacher practices may have been responsible for the changes in identification of students in this study. As achievement of students increased within the classrooms, they were more likely to be identified as higher achieving. The cluster grouping program may have provided these students with more opportunity to achieve by removing the highest achievers from four of the five classrooms in each grade level. Teachers believed in the program and in their students' abilities. The teachers indicated that the grouping and placement used in the cluster program helped them to better meet individual needs, and, with the highest achieving students removed from their classrooms, other students gained in achievement and confidence. These findings, which are contrary to popular views in the reform movement that grouping somehow damages the low-achieving students (George, 1993; Oakes, 1985; Slavin, 1987a; Wheelock, 1992), should be considered together with analyses of ability grouping (e.g., Kulik & Kulik, 1992; Rogers, 1991) when decisions are made regarding how children will be placed in elementary classrooms. Cluster grouping may offer students opportunities for academic growth as well as recognition by their teachers, and its use should be seriously considered by elementary schools.

Achievement Findings and Implications for Research Question 2: How Do Students in the Treatment School Compare with Students from the Comparison School with Regard to Achievement?

To investigate the trends in identification, standardized student achievement was compared with the achievement of students from the comparison school. Repeated-measures MANOVA was used with school as the independent variable and grade 3, 4, and 5 NCE math and reading scores as dependent variables. Second-grade math and reading scores were used as covariates, because purposive sampling made randomization impossible. The covariates were significant for both the Class of 2000, $F(4, 300) = 81$, $p < .05$, and the Class of 2001, $F(4, 326) = 79.92$, $p < .05$, but not highly correlated with the independent variables, with canonical correlations of .16 for both data sets (accounting for only 2.6% of the variance that might otherwise explain the dependent variables).

After adjustment by the covariates, there were significant differences in the main and interaction effects for the Class of 2000. The main effect of school, $F(2, 150) = 16.98$, $p < .001$, and the interaction of school and time, $F(4, 610) = 8.01$, $p < .001$, were significant. Because a large sample can more easily yield statistical significance, effect sizes for these results were examined. The effect sizes (R^2) were calculated as 1 minus Wilks' Lambda (Tabachnick & Fidell, 1989). For the significant main effect of school, the effect size was $R^2 = .18$, and the interaction had an effect size of $R^2 = .10$, both practically significant with medium and small effects, respectively (Cohen, 1988). Therefore, after adjustment for initial differences in achievement scores between the schools, the school that the students attended accounted for 18% of the variation in achievement scores measured by the combined supervariable of math and reading. The interaction of school and time accounted for 10% of this variation.

For the Class of 2001, after adjusting for initial differences, significant differences were found in main effects of school, $F(2, 162) = 10.14$, $p < .001$; and time, $F(4, 662) = 6.65$, $p < .001$; and in the interaction of school by time, $F(4, 662) = 4.32$, $p < .002$. Effect sizes for these significant differences were $R^2 = .11$, $R^2 = .08$, and $R^2 = .05$, respectively, all small, but practically significant (Cohen, 1988). The school that the students attended accounted for 11% of the variation in

achievement scores measured by the combined supervariable of math and reading; the repeated occasions of testing accounted for 8% of this variation; and the interaction of school and time accounted for 5% of the variation, after adjustments by the covariates.

Explaining the Omnibus Test: Between School Contrasts

The first set of contrasts used BMDP program 2V ANCOVA procedures to compare the achievement of students in reading and math by grade level between schools. In each contrast, the reading or math score was the dependent variable, school was the independent variable, and the grade 2 reading or math score was the covariate.

Table 2 includes the results of these contrasts for the Class of 2000, and Table 3 portrays results of the same analyses for the Class of 2001. As indicated in Table 2, after adjustment for initial differences, on average, students in the Class of 2000 comparison school scored significantly higher (M_c = 53.84) than students in the treatment school (M_t = 47.94) in grade 3 reading scores, with a small effect size of R^2 = .055. Yet, by grade 5, students in the treatment school were averaging significantly higher scores (M_t = 53.25) than students in the comparison school (M_c= 48.16), with a small effect size of (R^2 = .042) (Cohen, 1988).

Also indicated in Table 2, with regard to mathematics scores of the students from the Class of 2000, significant differences were found between the schools at grade 3, grade 4, and grade 5 after adjusting for initial differences. The treatment school students averaged higher adjusted scores than the comparison school students on each occasion (grade 3: M_t = 55.98; M_c = 49.79; grade 4: M_t = 55.60; M_c=50.87; grade 5: M_t = 58.01; M_c = 48.25), with the largest difference between the scores occurring at grade 5. The effect sizes for differences at grades 3 (R^2 = .039) and 4 (R^2 = .032) were practically significant, yet small, and the effect size for the grade 5 differences was medium R^2 = .099 (Cohen, 1988).

As depicted in Table 3, after adjusting for initial differences, comparison school students from the Class of 2001 averaged significantly higher scores than treatment school students in reading achievement when in grade 3 (M_c = 52.62; M_t = 46.79), but there were no significant differences in reading achievement in either grade 4 or grade 5. The medium effect size, R^2 = .066, was practically significant and accounted for 6.6% of the variation in adjusted reading scores. By grade 5, the treatment students had increased the mean adjusted score in reading achievement (M_t = 49.27) to a level that eliminated statistical differences between the groups.

As indicated in Table 3, with regard to the mathematics scores of the students from the Class of 2001, significant differences in achievement were found for grades 4 and 5 after adjusting for initial differences. In each case, the students in the treatment school scored higher than the students in the comparison school (grade 4: M_t =57.12, M_c = 51.49; grade 5: M_t =55.54, M_c=47.92). The difference at grade 4 had a small, but significant effect size of R^2 = .044, and the grade 5 difference had a medium and practically significant effect size of R^2 = .076.

Explaining the Omnibus Test: Within Schools Contrasts

The second set of planned contrasts compared student achievement in reading and math by the repeated measure of time. Within each school, scores were contrasted between grades 3 and 4; grades 4 and 5; and grades 3 and 5 to determine when significant changes occurred. For each contrast, the independent variable was time and the dependent variable was math or reading score. These scores were contrasted for each data set using BMDP program 2V ANOVA procedures. Because these contrasts were performed within the individual school data sets, no covariates were used.

For the treatment school Class of 2000, significant differences were found in mean reading achievement scores between grades 3 (M = 46.11) and 4 (M = 49.00), $F(1,85)$ = 4.59, $p < .05$, and grades 3 (M = 46.11) and 5 (M = 51.23), $F(1,85)$ = 12.87, $p < .05$. The difference in reading scores between grade 3 and grade 5 had practical significance and a large effect size of R^2 = .13, or 13% of the variance between the grades (Cohen, 1988). However, no differences existed in math achievement for these students.

The comparison school Class of 2000 contrasts included significant differences in mean reading achievement between grades 4 (M = 53.59) and 5 (M = 50.69), $F(1,66) = 5.14$, $p < .05$, and grades 3 (M = 56.22) and 5 (M = 50.69), $F(1,66)$ = 20.49, $p < .05$, with student scores decreasing. The decrease between grade 3 and grade 5 had large practical significance (R^2 = .23), which explained 23% of the variation in scores between these grades. As with the treatment school, no differences were found in math achievement between any grade levels.

Treatment students from the Class of 2001 showed significant increases in average reading achievement between grades 3 (M = 46.11) and 4 (M = 50.21), $F(1,97)$ = 10.30, $p <$ .05, a practically significant increase with a medium effect size of R^2 = .095. No significant differences in math achievement were found between any of the grades.

Planned contrasts for the comparison school students from the Class of 2001 indicated no differences in reading achievement with respect to time, but all contrasts of math

Table 2

Class of 2000 Planned Contrasts: Means, Standard Deviations, Adjusted Means, and F-values for Reading and Math Achievement Measures in Grades 2, 3, 4, and 5

	Treatment School		Comparison School			
Achievement measure	Mean (SD)	Adjusted mean	Mean (SD)	Adjusted mean	*F* value	E.S.
(Covariate) Reading Grade 2	49.42 (23.83)		56.71 (28.84)			
Reading Grade 3	46.12 (17.15)	47.94	56.22 (18.49)	53.84	8.89*	R^2 = .055
Reading Grade 4	49.13 (17.27)	50.98	53.59 (18.62)	51.13	.01	
Reading Grade 5	51.30 (18.05)	53.25	50.69 (17.90)	48.16	6.75*	R^2 = .042
(Covariate) Math Grade 2	53.98 (21.24)		59.00 (19.96)			
Math Grade 3	54.58 (21.32)	55.98	51.62 (17.12)	49.79	6.41*	R^2 = .039
Math Grade 4	54.44 (16.47)	55.60	52.40 (17.31)	50.87	5.12*	R^2 = .032
Math Grade 5	56.48 (22.96)	58.01	50.35 (18.81)	48.25	16.73**	R^2 = .099

*Note. n = 155; * p < .05; **p < .001*

achievement were found to be statistically significant with decreases in mean scores between grades 3 (M = 57.84), 4 (M = 53.42), and 5 (M = 50.10). This steady decrease between grades 3 and 5, $F(1,67) = 18.58$, $p < .05$, had a large effect size of $R^2 = .214$.

Summary of Findings for Research Question 2

Even though students in the treatment schools began with lower reading scores than did students in the comparison school, after three years in a flexible cluster grouping program, the treatment school students outperformed or equaled their comparison school counterparts. Additionally, the growth in reading achievement had both practical and statistical significance for the treatment school students. Qualitative findings revealed that treatment school students from both the Class of 2000 and the Class of 2001 were regrouped between classes for reading instruction on the basis of performance in reading during each year of the program (grades 3–5). Administrators from the comparison school

Table 3

Class of 2001 Planned Contrasts: Means, Standard Deviations, Adjusted Means, and F-values for Reading and Math Achievement Measures in Grades 2, 3, 4, and 5

	Treatment School		Comparison School			
Achievement measure	Mean (SD)	Adjusted mean	Mean (SD)	Adjusted mean	*F* value	E.S.
(Covariate) Reading Grade 2	50.67 (21.91)		53.22 (15.03)			
Reading Grade 3	46.11 (19.38)	46.79	53.60 (14.56)	52.62	11.71**	R^2 = .066
Reading Grade 4	50.21 (17.00)	50.90	51.62 (15.20)	50.64	.01	
Reading Grade 5	48.59 (18.57)	49.27	52.78 (15.14)	51.80	1.78	
(Covariate) Math Grade 2	52.04 (18.33)		57.78 (16.45)			
Math Grade 3	56.04 (19.68)	57.70	57.84 (18.45)	55.43	.96	
Math Grade 4	55.79 (17.75)	57.12	53.42 (13.49)	51.49	7.72*	R^2 = .044
Math Grade 5	54.03 (17.93)	55.54	50.10 (16.20)	47.92	13.56**	R^2 = .076

Note. $n = 155$. * $p < .05$. ** $p < .001$

indicated that heterogeneous and whole-group instruction were used for teaching reading in their elementary classrooms, and they were not involved in achievement grouping. These findings may indicate that the effects of the cluster grouping combined with regrouping for reading instruction had a positive impact on the reading achievement of treatment school students in the three years.

With regard to math achievement for both data sets, students in the treatment school scored significantly higher than the comparison school students in mathematics during the three program years, with the largest differences in grade 5 after adjusting scores for initial differences. However, unlike reading achievement, treatment school students did not show significant changes in math achievement, possibly because the math scores were already high (as indicated by mean NCE scores above 50 for students from each data set). For the three program years, the average NCE math score for the Class of 2000 was 53.75, and for the Class of 2001 it was 55.28.

Students were regrouped for math instruction between classes by achievement levels. As with reading, the collection of high teacher expectations, the use of grouping, and the use of challenging instructional strategies may have been responsible for the high achievement in mathematics of the students from the treatment school.

As indicated in the qualitative part of this analysis, many teachers (n = 11; 79%) and all administrators (n = 3) thought that the restriction of the range of achievement in classrooms, as well as the between-class grouping by achievement levels in reading and math, helped teachers meet the individual needs of students in their classrooms. Qualitative findings also indicated the use of a variety of instructional strategies around the themes of challenge, choice, and interests. High teacher expectations and the use of grouping may also have influenced student achievement in the treatment school. Again, contrary to the popular anti-grouping sentiment, these findings reinforce the idea that the use of flexible grouping coupled with appropriate instruction may positively influence student achievement. The implication for elementary schools is that flexible achievement grouping used in conjunction with challenging curriculum should be considered when designing educational programs. As Teacher 3C explained:

> By using achievement grouping, we are able to challenge the high achievers and meet the needs of the low achievers without having the low achievers or the high achievers feel like they had been singled out. We are able to adjust our curriculum and instruction to meet the individual needs of the students at their levels.

Qualitative Findings for Research Question 3: What Factors Exist in the Classrooms and the School Using Cluster Grouping That May Have Influenced Student Achievement?

The findings discussed in this section emerged as core categories after open, axial, and selective coding had been applied to the data as recommended by Strauss and Corbin (1990). This coding yielded three core categories: *the use of grouping*, *the apparent impact of the teachers*, and *the general school environment*. Response frequencies greater than or equal to six (over half of the teachers not responsible for the high-achieving cluster students) were considered to represent a general consensus and are reported as a theme within a core category. With regard to the administrators, two of three responses indicated a theme.

The Use of Grouping

Because cluster grouping implies ability grouping, both the program documents and the teacher interviews focused on the use of various forms of grouping in the program between grades 3 and 5. Grouping occurred within classrooms and between classrooms and, in both cases, was flexible. Like the identification procedures that were used to place students into classrooms, grouping was employed in a variety of ways, and students were not locked into specific groups for the duration of the day. Additionally, many teachers (n = 13; 93%) reported that they thought the cluster grouping was directly related to the increase in the number of *high-achieving* students identified during the three program years. Others (n = 11; 79%) reported that they believed the cluster grouping program helped them better meet the needs of the individual students within their classrooms.

Between-class groups. It became evident after the first few interviews that even though cluster grouping was used for placing students in classrooms, students were regrouped by achievement for reading and math instruction in grades 3, 4, and 5. This meant that, within each grade level, the teachers regrouped the students for reading and math instruction by achievement levels, and different teachers instructed students who were not necessarily the students from their regular class. This regrouping applied to the Class of 2000 for reading in grades 3–5 and for math in grades 4 and 5. The Class of 2001 was regrouped for instruction for reading and math in grades 3–5. Teachers at each grade level chose to do this to better meet the needs of students. One teacher would take the low-achieving students, another would take the advanced students, and the remaining three teachers would have students who were achieving near average in reading and math. The teacher who had the *high-achieving* cluster did not necessarily teach these students for reading and math. Therefore, other teachers had the opportunity to work with the most advanced students. The teachers explained that more students than those in the *high achieving* cluster could be in the *high reading* or *high math* sections, and that these sections did not necessarily include the same students. As Teacher 3A explained:

> We had so many high math students who weren't in the high cluster, we thought, to really meet the needs of the grade level, we would have a cluster group strictly for math. We also had the high cluster reading group to meet the needs of other children who may not have been identified or who had strengths that weren't evident across the board. We were able to target more children for high reading by regrouping within the grade level for reading.

Teacher 3B, who taught the low math class, explained:

> I teach the low math group, which includes the learning disabled students and those identified for Chapter 1 assistance. With these students, I am able to teach in different ways and go at a slower pace, but they think that they are great at math because the better students are not in the room to make them feel slow. We do a lot of hands-on things like base-10 blocks, patterning, and touch math, because many of them can't get it in the traditional ways. We do a lot of problem solving, mental math ... I challenge them at grade level ... I don't dumb down the curriculum, I just teach it differently so they can be successful, too.

This teacher also taught the advanced reading section where she used a literature-based curriculum, and students worked beyond grade level, had their curriculum compacted, and were involved in many different writing activities. When teaching the low-achieving students who had been regrouped for math, she had the assistance of a teacher consultant and a Chapter 1 aide.

In addition to using between-class grouping by achievement for math and reading, the three *high-achieving* cluster teachers indicated that they used between-grade grouping when it was needed to meet the needs of individual students. For example, Teacher 5A explained:

> Some students went to sixth grade to take math in the middle school, because they were even beyond where I was with the high math group. After math in the middle school, they would return for the rest of the day in fifth grade.

Within-class groups. The types of within-class grouping that were reported included interest grouping (n = 8), cooperative grouping (n = 7), and flexible grouping (n = 6). Six teachers explained that they used flexible grouping and, depending on the lesson, students often chose their groups or partners. Teacher 4D described her use of grouping in the following way:

> I do all those things, cooperative learning, interest groups, peer teaching, whole group instruction. We're driven ... to do what works for children and use a variety of methods. So, anything that we feel we can use in our classrooms to facilitate whatever the needs are, we do that.

Teacher 3A described,

> I used all forms of grouping, including cooperative learning, flexible groups, and interest groups. Sometimes I chose the groups, sometimes it was by interest, and sometimes groups were chosen independently by the students.

Flexible groups. Flexibility emerged as a key component of both the within-class grouping and the between-class grouping. Seven teachers explained that the grouping between classes was always flexible, and that if a student needed to be in another section, the cooperation and flexibility existed within each grade level to move students around as needed. With the exception of the four teachers who said that their primary mode of instruction was whole-group instruction, it was evident from the many comments that the use of grouping within the classrooms varied and was flexible in nature. For example, Teacher 4C explained:

> The types of groups that I use in my class depend on the activity; sometimes I use cooperative learning, or peer tutoring, other times I use interest grouping, or I group students by ability. The main thing with my use of grouping is that it is flexible.

Even though students were identified in various achievement levels for placement into classrooms, it became evident that these identification categories were not fixed, nor were they used consistently to group students for instruction. Rather, students were grouped and regrouped in many flexible ways designed by the teachers to help them be successful.

Cluster grouping and meeting the needs of individual students. Eleven teachers (79%) indicated that cluster-grouping placement strategies made it easier for them to meet the needs of individual students in their classrooms. The superintendent and the assistant principal agreed. Eight teachers (57%) said that cluster grouping allowed them more time to work with lower achieving students at a level appropriate for these students, whereas all three of the teachers who worked with the high achievers indicated that they were able to do much more to challenge and promote growth in these students than had previously been possible. In the 1990 program evaluation report submitted to the Board of Education, four teachers explained how cluster grouping had helped them better address individual student needs in their classrooms.

The teachers who taught the high-achieving clusters said that it was beneficial to the high achievers to be clustered together because they challenged each other and didn't always get to be the best. Teacher 4A explained, "They [high-achieving students] challenge and motivate each other, and with just one or two kids, I don't think that would happen."

Nine teachers (64%) indicated that the restricted range of achievement levels created by cluster-grouping placements made meeting individual needs easier for them.

> Yes. That's one thing teaching for the first 10 years, I always felt guilty, like I always felt I wasn't giving enough time to the low kids, and I also felt like I wasn't challenging the high kids enough. Because I think the gap is narrower so I can zero in on their needs. (Teacher 3B)

> The kids were more deliberately placed, so we didn't have as broad of a range and didn't have to deal with the extremes. I also had an aide and a teacher consultant, which helped to meet the needs of the students who were struggling. (Teacher 4B)

The majority of the teachers agreed that cluster grouping helped them meet the needs of individual students in their classrooms. The restricted range of achievement levels created more time for the teachers to work with students in their classrooms. They also reported that cluster grouping was beneficial to students because it allowed students of like achievement levels to work together and challenge each other.

The Apparent Impact of Teachers

Positive classroom environments. Teachers and administrators reported positive classroom environments and said that school was a place where students wanted to be. This finding was confirmed by school climate surveys, completed in 1990 and 1991 as part of school improvement planning, in which students in the upper elementary school indicated that they were, on average, happy with their classrooms and felt that

school was friendly and safe. Parent satisfaction surveys were high each year, as reported in the annual reports to the Board of Education and to the State Department of Education. During the semistructured interviews, teachers were asked to describe the atmosphere of their classrooms.

> Excited. The kids don't want to miss school, even when they're sick. I never have a motivation problem, because they like what they are doing and are challenged and feeling successful. (Teacher 3A)

> I think it's safe for them to be who they are, to be different, and to disagree with me. If they offer suggestions on my teaching, I listen to them. I think they understand that I really care, but that there are high expectations for learning. (Teacher 5A)

Both teachers and administrators discussed how the teaching strategies and the curriculum modifications were used to benefit the students. Many teachers (n = 12) discussed adjusting assignments, helping students to feel successful, and making their classrooms places that students wanted to be. A theme of concern and caring was continuously discussed by teachers.

High, yet realistic teacher expectations. All of the teachers reported that their expectations for students were high, and two said that they expected more than one year's growth from their students. Two other teachers stated that their expectations were the same as when the high-achieving students had been in their classrooms. Fifth-grade teachers discussed preparing their students for success at the middle school. The general tone of the interviews indicated that the teachers believed in the need to challenge, but at the same time, help the students experience success. Three teachers said they had been accused of having standards that were "too high." None of the teachers said that removing the high-achieving students from their classrooms had, in any way, influenced their expectations. Comments from teachers regarding their expectations included:

> I always keep my expectations of the high achievers high and work and make sure I really push the kids to do more than they want to do. I give work back to them, tell them they have good ideas, and encourage them to expand [their ideas], because they are capable of more. (Teacher 3A)

> I truly expected all students to achieve ... regardless of where they are or who they are. I want to meet the needs of students and feel my standards or expectations are high. (Teacher 5E)

> I don't believe because a child has an LD or EMI label means that they are low. I think that's a problem with education—just because a child is identified with a disability or something, some people tend to think, Well, they're low. I expect a lot from them—I don't think they are dumb. I think they can do just as many things as gifted kids, maybe not to the full extent, but in some things they can go beyond. If it's their interest, they can excel just as much as anybody else can. (Teacher 3B)

As the quantitative analyses of the identification categories and achievement data indicated, students seemed to be successful in these classrooms.

Strategies for challenging students and meeting students' needs in the cluster-grouped classroom. Most teachers indicated that they were concerned with meeting the needs of individual students. The strategies these teachers used to challenge and meet the needs of students in their cluster-grouped classrooms are summarized in Table 4. The related themes of challenge, choice, and student interest emerged through all these strategies. Strategies are presented according to which of these themes or combination of themes they belong.

As indicated in Table 4, the majority of these strategies were reported to have been used by the teachers who taught the classrooms with the high-achieving cluster students. However, many strategies were used in other classrooms with students of all achievement levels. For example, curriculum compacting was used by all teachers who had the high-achieving students, but also by five other teachers. Four teachers with regular classrooms had implemented the choice of independent study with their students, and seven teachers regularly provided enrichment experiences beyond the curriculum to their students. Teachers in many classrooms reported using thinking, questioning, and problem-solving strategies, and over half of the teachers reported that they gave students choices in group assignment and curriculum assignments. A variety of methods were used to incorporate student interests including the use of enrichment/interest centers (n = 10); curriculum compacting (n = 8); and independent study in an area of student choice and interest (n = 7). As Teacher 4A described:

> Because their ideas are implemented, their ideas become part of what we do. Students are pretty empowered in the classroom. For example, a couple of years ago, we had two girls really interested in special education. They did some research and worked once a week with the hearing impaired teacher and her students, and then they came back to class and taught us sign language and shared what they learned.

As indicated by the teachers of the high-achieving cluster students, there was a balance of acceleration and enrichment through appropriate challenges and choices.

> In the homeroom with the high cluster, I found with English and science we were able to move much faster and at a higher level. We didn't need to do spelling every day like other classrooms. I was able to use that time for independent studies and special projects with children. I really liked it because I thought I was challenging the students, and it was productive. I pretest and give them other choices instead. I also move at a faster pace, at a higher level, with higher expectations. (Teacher 3A)

> I use fifth grade math and spelling. I use all kinds of enrichment things—mind benders, we pull in Engin-Uity, inventions, Invent America stuff, they get involved in poetry writing, Science Olympid, Math Olympiad. (Teacher 4A)

> Academically talented [students] were allowed to move up to the sixth grade or in some way [work] independently. We did various types of activities. They would have choices ... they could put on a drama to present their material, they could write a book, they use poetry, they could sing a song.... Then, we had different enrichment programs

Table 4

Strategies for Challenging and Meeting Students' Needs in the Regular Cluster Grouped Classroom: Frequency of Use by Grade Level Responses

Strategy	Grade 3 Responses (n = 5)	Grade 4 Responses (n = 4)	Grade 5 Responses (n = 5)
CHALLENGE			
Integrating High Order Thinking Skills	5★	3★	3★
Developing Critical Thinking Skills	2	3★	3★
Using Creative Thinking Skills	2★	2★	2★
Integrating Problem Solving	3★	2★	3★
Assigning Projects	3★	1★	1★
Using Acceleration	1★	2★	1★
Adjusting Assignments	4	3★	3
CHALLENGE & INTEREST			
Spending Time with High Achievers	1★	1★	1★
Developing Curricular Extensions	5★	4	3★
CHOICE & INTEREST			
Providing Choice of Partners or Groups	2★	2★	4★
Providing Choice to Work Alone or Together	3★	2★	3★
CHALLENGE, CHOICE, & INTEREST			
Using Open-Ended questioning	5★	4★	4★
Offering Independent Study	2★	2★	3★
Using Challenge Questions	2★	2★	1
Implementing Curriculum Compacting	4★	1★	3★
Providing Enrichment Experiences	5★	2★	3★
Providing Choice of Problems or Assignments	2★	2	3★

Note. ★*Indicates that one of the respondents included the teacher of the high achieving cluster.*

> throughout the school that kids could apply for and attend. If they liked to write, they could go to a residence in writing; sometimes, we had a mentor in drama or art with them, and they were allowed to pursue those things. I would pretest; if they knew it, then we would cover only the things that they hadn't mastered. With independent studies, they had choices to select things they were interested in, but they were required to meet a certain standard, a certain way of writing; they had to produce a product, had to share with an audience ... that sort of thing. I tried to have writing assignments across the curriculum, plus their independent study was like a thesis-type paper. (Teacher 5A)

The General School Environment

Strong administrative leadership and support. Teachers supplied evidence of strong administrative leadership. Only one teacher said the administration had not been supportive, two others expressed that support had been mixed, and 11 indicated that there had been firm support on the part of the administration.

Professional development opportunities. Professional development was ongoing, and most teachers indicated that it was an important part of their success as teachers and with the cluster grouping program. Before choosing to implement the cluster grouping program, all staff attended a one-day workshop on the concept of cluster grouping, and seven teachers went on site visitations to a school that was successfully using cluster grouping. National, state, regional, and local profes-

sional development opportunities in gifted education were made available to staff, with all participating in at least the local opportunities (e.g., curriculum compacting; differentiating and individualizing curriculum and instruction; LD gifted and underachievers; meeting the needs of gifted math and science students). In all, a total of 64% of the teachers attended national, state, or regional professional development conferences or workshops in gifted education. Additionally, six teachers mentioned how helpful it was to have the teachers who teach the *high-achieving* cluster in the building as resources. As Teacher 3B explained:

> I've learned so much from [Teacher 3A], and I adapt many of the strategies that she uses with her high achievers and use them with my LD and low achievers. I don't think that gifted education is just for gifted students.

Belief in colleagues and collaboration. The administration and teachers demonstrated strong support of and confidence in the teachers. There was a general atmosphere in this school of quality and of caring by teachers who seemed to do their best to work with students. Fifty-five percent of the teachers who were not responsible for the *high-achieving* students indicated that they used strategies in their classrooms that they thought were typically "gifted education" strategies. All of the third-grade teachers, for example, indicated that they were glad that Teacher 3A had the cluster of high achievers because she had to put so much extra effort into meeting these students' needs, and she was talented in working with those children. The teachers had confidence in each other, worked together, and were regarded as competent by the administration.

Program benefits to all students and teachers. The program was viewed as successful because the teachers and administrators believe it was beneficial both to the teachers and to the students. The teachers liked the program, and many believed it helped them better meet the needs of the students in their classrooms. Teacher 3B explained how she came to view the program:

> One thing—I remember how skeptical I was at the beginning because I'm not a risk-taker. I thought the same thing a few other people thought—oh, you take those top kids out and I'm not going to have any spark. And that was so far from being true. I see lots of sparks in my room.... and having my daughter in [the program] ... there's such a difference in her attitude and her love for school is back ... before being placed in the high-achieving cluster, she wasn't being challenged in school, now to see her doing research projects as an eight-year-old ... she's doing projects so beyond what I ever thought and she is so excited about school.

The administrators who were interviewed expressed their belief that the cluster-grouping program had helped the teachers do their jobs. As the superintendent explained:

> Well, I think we've got some real benefits. I had a great deal of skepticism when we first started because I thought, Well, are we looking at an elitist program where we're taking the cream of the crop and separating them even though they may be within a classroom with other students that's going to "dummy down" the other classes. In fact, it's had just the opposite effect. We have been able to have leadership rise in other classes where we don't have the very bright students who have been in those classes. So, it's had a real bonus effect for more general education students, from what I can see ... and at the same time accomplishing more challenges for the gifted kids. Additionally, I think that the cluster grouping program actually makes the teachers' jobs easier.

Summary of Findings for Research Question 3

Qualitative findings provided further insight into the treatment school and classrooms. The teachers in this study created positive classroom environments in which high expectations were held for all of their students. They used a variety of strategies, including various forms of grouping, to challenge and meet student needs. The program was supported by strong administrative leadership, and teachers had continuing professional development and growth opportunities in which most teachers chose to become involved. Both teachers and administrators worked collaboratively and indicated confidence in their colleagues' abilities. These findings were similar to those found in the exemplary programs for gifted investigated by Delcourt and Evans (1994), who cited the following characteristics of these programs: strong leadership, supportive atmosphere and environment, and flexible curriculum and instruction matched to student needs.

Discussion

The quantitative findings (increased reading achievement, higher math achievement, and increased numbers of students identified as *high-achieving* in the treatment school) combined with the qualitative findings indicate that cluster grouping, when combined with high teacher expectations, the use of strategies to challenge and meet individual needs, and positive classroom environments, may have a positive impact on all students in a school. All teachers and administrators (n = 17) involved in the program believed that cluster grouping was beneficial to both students and teachers, because it helped students be successful by structuring classes in a manner that helped teachers better address individual needs. These findings support research-based suggestions by Kulik and Kulik (1992) and Rogers (1991), who suggested that grouping by ability, when used in conjunction with appropriate differentiated instruction, can be beneficial to student achievement.

Contrary to findings by Oakes (1985, 1995), the teachers in this study who did not have the cluster of high-achieving students were not regarded as the poor teachers, and they did not lower their expectations for their students. In fact, they reported that the opposite occurred, and they expected the same or more from their students as highlighted by the following teacher comments.

> I guess I have the same high standards for the average-and low-achieving student as I do for any other student. (Teacher 4 C)

> I think it's our expectations. I think I just thought that they could all do it and expected them to do it. (Teacher 5D)

As noted by Tomlinson and Callahan (1992), Renzulli (1994), Reis, Gentry, and Park (1995), and the U. S. Department of Education (1993), the use of gifted education "know-how" has the potential to improve general education practices. The cluster-grouping program investigated in this study was designed to simultaneously address the needs of high-achieving students *and* the needs of other students. As a result of this connection with the general education program, professional development opportunities in gifted education were made available to all staff, and dialogue between teachers of the high-achieving cluster students and the rest of the staff was encouraged. As a result, all teachers received professional development in gifted education strategies and reported using these strategies in their classrooms with all of their students.

Unlike the classrooms described by Archambault et al. (1993) and observed by Goodlad (1984) and Westberg et al. (1993), the classrooms in this school were characterized by a variety of challenging activities and varied instructional strategies. Renzulli (1994) noted that the practice in many schools of diagnosing and remediating weaknesses should be replaced with a talent development approach to enrichment learning and teaching that recognizes student interests, strengths, and talents as a basis for their education. In this study, integration of the cluster-grouping program with the general education program seemed to impact all teachers and students in the school. The treatment school teachers applied many strategies from gifted education to their daily teaching, something that might not have happened had professional development in gifted education been reserved only for the teachers of the high-achieving students. The implication is that all staff and, consequently, all students can benefit from in-service in gifted education strategies. Therefore, schools should be careful not to limit their professional development in gifted education to just those teachers who work with identified gifted students. By offering more teachers opportunities to learn and to apply gifted education know-how, perhaps student achievement can be raised in schools.

The findings of this study should interest school districts that are struggling with how to meet the needs of gifted students in the regular classroom. Although current reform trends suggest that heterogeneous grouping is preferred (George, 1993; Hopfenberg & Levin, 1993; Oakes, 1985; Slavin, 1987a; Wheelock, 1992) when developing elementary classroom configurations, the findings of this study suggest that the deliberate placement of a narrower range of achievement groups in teachers' classrooms, including the placement of a group of *high-achieving* students together in one room, is beneficial to both students and teachers. It stands to reason that if the *high-achieving* students are placed with a teacher who has the background and willingness to adjust curriculum and instruction to meet these students' special needs, their needs are more likely to be met than if they are randomly placed into all teachers' classrooms for the sake of heterogeneous grouping. Further, as was done in the program in this study, if the placement of students in the other teachers' classrooms is done thoughtfully, and includes a group of students who are above average, then districts might see growth in identification and achievement similar to that observed in this study. The implication for districts is that a well-developed cluster-grouping program, such as the one in this study, can offer gifted education services to *high-achieving* students while helping teachers better meet the needs of all students.

Elementary classroom teachers might find the results of this study interesting as they struggle to meet the individual needs of students. Of special interest are the reports by teachers in this study that removing the highest achievers from four of five classrooms per grade level did not affect the way teachers viewed the students in their classrooms. There was no report of "losing the spark" by teachers who were not responsible for the *high-achieving* students. On the contrary, these teachers reported that having the *high-achieving* students removed from their classrooms helped them better meet the needs of individual students, while encouraging new talent to emerge. As Teacher 5B suggested:

> I really believe that those high-achieving kids are not models for the other kids. The other kids know where they are ... so they don't model themselves after those kids. When [the high achievers are] taken out and able to move at their own rate, then these other kids who are good and could be better begin to surface and begin to shine and not sit back and let those extremely high achievers take control of the classroom.

Other teachers may want to consider the views of the teachers involved in this study when deciding whether to try a cluster-grouping approach to programming and classroom placements.

The varied uses of grouping found in this study have implications for teachers who have questions regarding its

appropriate uses. The teachers in this study used achievement grouping in math and reading, reporting that this made it easier for them to challenge the student at appropriate levels. They also used other forms of flexible grouping. The implications are that flexible achievement grouping has the potential to produce academic gains for all students.

Although the weaknesses of causal comparative research are well documented (Gall, Borg, & Gall, 1996), a study such as this is valuable in other ways. First, it enabled the investigation of a practice that was implemented and carried out in an actual school setting. Second, it investigated a school-initiated innovation, as opposed to an innovation demanded by external sources, such as federal funding or special mandates. Because the innovation was school-based, local control and ownership were invested in the program. Although there are problems associated with the use of intact groups, the use of intact groups provided a distinct advantage in this research. The intact groups examined were stable over time and facilitated longitudinal comparisons of students between and within groups during the course of a three-year program. Finally, this study examined an often-recommended practice for which little research exists, and the findings can serve as the basis for further, more carefully controlled experimental or quasi-experimental research. A discussion of limitations follows.

Limitations

Internal validity of the study was limited by instrumentation, history, differential selection, the use of intact groups, and multiple treatments (Gall, Borg, & Gall, 1996). Regarding instrumentation, existing measures of achievement had to be used, and these measures were on two different tests. The use of NCE scores, the similarity of the ITBS and the CAT, the large number of students involved in the study, and the use of ANCOVA enabled the use of these two instruments to compare student achievement over time. The use of two data sets and repeated measures helped to control for the threat of history and increase confidence that results were not simply due to chance. However, results must be interpreted with caution because other events (of which the researchers were unaware) that occurred during the time the program took place may also have influenced results. The similarity of the comparison school and the use of a covariate helped to control for the threat of differential selection.

Huck and Cormier (1996) warned that a covariate must be used with caution when there are intact groups. However, the covariate was not highly correlated with the independent variable, and the use of nationally normed instruments on which students scored near the national mean helped to increase confidence in the use of the covariate. Additionally, whenever intact groups are used, there is the problem of intraclass correlation (ICC), meaning that the subjects' scores on dependent variables are often correlated simply because they are in the groups (Barcikowski, 1981; Scariano & Davenport, 1987), which violates the assumption that data are independent and inflates the alpha level. One method for addressing the problem of ICC is to make the classroom the unit of analysis. This was impossible here for several reasons. First, students were examined over a three-year period, and their classrooms placements changed. Second, flexible grouping also changed the groups that students worked in throughout the day. Finally, no information regarding classroom placements was available for the comparison school. To control for the inflated probability of Type I error, alpha levels were examined within the context of this problem. Multivariate ANCOVA analyses produced significance at $p < .0001$, and the contrasts produced significance ranges from $p < .05$ to $p < .001$, indicating that findings were likely significant despite the increased probability of Type I error. As Cohen (1994) suggested, examining practical significance estimated by effect sizes is more meaningful than reporting statistical significance. Effect sizes were reported throughout this study. These alpha levels and effect sizes increase confidence in the results when considering the problems associated with nonindependence of scores.

The effect of multiple treatments must be acknowledged. This study was not intended to isolate one variable, study that variable, and attribute causality to that variable. Rather, the intent was to acknowledge the complexity of a real program that existed in a real school. There was more going on than just "cluster grouping," as there would be in any school. It was not intended, nor was it possible, to isolate the effects of cluster grouping from the effects of several other variables, including regrouping by achievement for math and reading, the clustering of special needs students as well as high-achieving students, and the entire class of high-achieving cluster students that existed by fifth grade. However, viewed in total, the findings are powerful, and much can be learned about classroom practices, identification, and student achievement by examining the ways in which cluster grouping was integrated with and applied to an elementary school program, its curriculum, and instruction. Cluster grouping was the basis from which the school programs developed, but it was not the sole program existing in this school. Therefore, it would not be appropriate in this study of cluster grouping to make a claim that simply placing students in a cluster group will increase achievement among students without the flexible

grouping within and between classes, the staff development and ownership, high teacher expectations, differentiation of curriculum and instruction for all levels of students, and a reduction of range of achievement levels that the teachers had to teach.

The qualitative portion might have been strengthened by including interviews from parents and students; however, given that this study occurred after the students were in the program, and they had since progressed to middle school, it was decided that such interviews might not provide valid recollections of the elementary school experiences. In further study of cluster grouping that occurs during the program, it would be advisable to include perspectives from both students and parents.

Conclusion

The key to the findings of this study is that the use of cluster grouping facilitated many other positive changes in this school—as perceived by the teachers—such as rich staff development opportunities, ownership in a program that they developed, high teacher expectations, and a reduced range of achievement levels in their classrooms that helped facilitate teachers' desire to better meet the individual needs of all students. The use of grouping is a rich and complex issue, and far too many researchers have attempted to isolate and oversimplify its use. The intention of this study was to understand the working dynamics of cluster grouping in a school that saw consistent increases in achievement and identification of their elementary students. To this end, it seems clear that cluster grouping played a role in this school's success. ¶

References

Airasian, P. W. (1989). Review of the *California Achievement Tests, Forms E and F.* In J. C. Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (pp. 126–128). Lincoln, NE: Buros Institute of Mental Measurement.

Archambault, F. X., Westberg, K., Brown, S. B., Hallmark, B. W., Emmons, C. L., & Zhang, W. (1993). *Regular classroom practices with gifted students: Results of a national survey of classroom teachers.* Storrs, CT: The National Research Center on the Gifted and Talented.

Balzer, C., & Siewert, B. (Eds.). (1990). *Program and service models: Suggested programs and services for identified talented and gifted students, K–12.* (Technical assistance paper 3, Revised.). Salem, OR: Oregon Department of Education.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics, 6,* 267–285.

Brown, S. B., Archambault, F. X., Zhang, W., Westberg, K. (1994, April). *The impact of gifted students on the classroom practices of teachers.* Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.

Bryant, M. A. (1987). Meeting the needs of gifted first grade children in a heterogeneous classroom. *Roeper Review, 9,* 214–216.

California Achievement Test, Form E. (1984). Monterey, CA: CTB/McGraw-Hill.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1994). The earth is round. *American Psychologist, 49,* 997–1003.

Coleman, M. R. (1995). The importance of cluster grouping. *Gifted Child Today, 18*(1), 38–40.

Davis, G. A., & Rimm, S. W. (1985). *Education of the gifted and talented.* Englewood Cliffs, NJ: Prentice-Hall.

Delcourt, M. A. B., & Evans, K. (1994). *Qualitative extension of the learning outcomes study.* Storrs, CT: The National Research Center on the Gifted and Talented.

Delcourt, M. A. B., Loyd, B. H., Cornell, D. G., & Goldberg, M. D. (1994). *Evaluation of the effects of programming arrangements on student learning outcomes.* Storrs, CT: The National Research Center on the Gifted and Talented.

Dixon, W. J. (Ed.). (1992). *BMDP statistical software manual* (Vols. 1–2). Berkley, CA: University of California Press.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.

Gentry, M. L. (1996). *Cluster grouping: An investigation of student achievement, identification, and classroom practices.* Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.

George, P. (1993). Tracking and ability grouping in the middle school: Ten tentative truths. *Middle School Journal, 24*(4), 17–24.

Goodlad, J. I. (1984). *A place called school.* New York: McGraw-Hill.

Hieronymus, A. N., Hoover, H. D., & Lindquist, E. F. (1984). *Iowa tests of basic skills (Form G).* Chicago: Riverside Publishing.

Hoover, S., Sayler, M., & Feldhusen, J. F. (1993). Cluster grouping of elementary students at the elementary level. *Roeper Review, 16,* 13–15.

Hopfenberg, W. S., & Levin, H. A. (1993). *The accelerated schools resource guide.* San Francisco: Jossey-Bass.

Huck, S. W., & Cormier, W. H. (1996). *Reading statistics and research.* New York: Harper-Collins.

Juntune, J. (1981). *Successful programs for the gifted and talented.* Hot Springs, AR: National Association for the Gifted and Talented.

Kaplan, S. N. (1974). *Providing programs for the gifted and talented.* Ventura, CA: Office of the Ventura County Superintendent of Schools.

Kennedy, D. M. (1989). *Classroom interactions of gifted and non gifted fifth graders.* Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Kennedy, D. M. (1995). Teaching gifted in regular classrooms: Plain talk about creating a gifted-friendly classroom. *Roeper Review, 17*, 232–234.

Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives.* Storrs, CT: The National Research Center on the Gifted and Talented.

Kulik, J. A., & Kulik, C-L. C. (1991). Ability grouping and gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 178–196). Boston, MA: Allyn & Bacon.

Kulik, J. A., & Kulik, C-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*, 73–77.

LaRose, B. (1986). The lighthouse program: A longitudinal research project. *Journal for the Education of the Gifted, 9*, 224–32.

McInerney, C. F. (1983). *Cluster grouping for the gifted, the bottom line: Research-based classroom strategies. A Series for teachers.* St. Paul, MN: LINE.

Michigan Department of Education. (1994, August). *1992–93 Bulletin 1014: Michigan K–12 school districts ranked by selected financial data.* Lansing, MI: Author.

Mitchell, E. S. (1986). Multiple triangulation: A methodology for nursing science. *Advances in Nursing Science, 8*(3), 18–26.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J. (1995). More than meets the eye: Links between tracking and the culture of schools. In H. Pool & J. A. Page (Eds.), *Beyond tracking: Finding success in inclusive schools* (pp. 59–70). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Purcell, J. (1994). *The status of programs for high ability students.* Storrs, CT: The National Research Center on the Gifted and Talented.

Reis, S. M., Gentry, M. L., & Park, S. (1995). *Extending the pedagogy of gifted education to all students: The enrichment cluster study.* Storrs, CT: The National Research Center on the Gifted and Talented.

Renzulli, J. S. (1994). *Schools for talent development: A comprehensive plan for total school improvement.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., & Reis, S. M. (1985). *The schoolwide enrichment model: A comprehensive plan for educational excellence.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., Smith, L. H., Callahan, C., White, A., & Hartman, R. (1977). *Scales for rating the behavioral characteristics of superior students.* Mansfield Center, CT: Creative Learning Press.

Rogers, K. B. (1991). *The relationship of grouping practices to the education of the gifted and talented learner.* Storrs, CT: The National Research Center on the Gifted and Talented.

Rogers, K. B. (1993). Grouping the gifted and talented: Questions and answers. *Roeper Review, 16*, 8–12.

Scariano, S., & Davenport, J. (1987). The effects of violations of the independence assumption in the one way ANOVA. *American Statistician, 41*, 123–129.

Slavin, R. E. (1987a). Ability grouping: A best-evidence synthesis. *Review of Educational Research, 57*, 293–336.

Slavin, R. E. (1987b). Grouping for instruction. *Equity and Excellence, 23*(1–2), 31–36.

Spradley, J. P. (1980). *Participant observation.* New York: Holt, Rinehart & Winston.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics.* New York: Harper-Collins.

Tomlinson, C. A., & Callahan, C. M. (1992). Contributions of gifted education to general education in a time of change. *Gifted Child Quarterly, 36*, 183–189.

U.S. Department of Education, Office of Educational Research and Improvement. (1993). *National excellence: A case for developing America's talent.* Washington, DC: U.S. Government Printing Office.

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). *An observational study of instructional and curricular practices used with gifted and talented students in regular classrooms* (Research Monograph 93104). Storrs, CT: The National Research Center on the Gifted and Talented.

Wheelock, A. (1992). *Crossing the tracks: How untracking can save America's schools.* New York: New Press.

Winebrenner, S. (1992). *Teaching gifted kids in the regular classroom.* Minneapolis, MN: Free Spirit Publishing.

Winebrenner, S., & Devlin, B. (1994). *Cluster grouping fact sheet: How to provide full-time services for gifted students on existing budgets.* Lombard, IL: Phantom Press.

Author Notes

1. The definitions of the terms *high-achieving*, *high-ability*, and *gifted* vary in the literature. The students in this study were identified as *high-achieving*.