

Ability Grouping Effects on Academic Achievement and Self-Esteem: Who Performs in the Long Run as Expected

HELEN ABADZI

Ft. Worth Independent School District

ABSTRACT The regression-discontinuity analysis was used to study the effects of ability grouping on the academic achievement and self-esteem of a student cohort, grades 4-6 in a large Texas school district that used a cutoff score for group placement. Iowa Tests of Basic Skills (ITBS) scores were reported in normal curve equivalents; therefore, gains referred to amount of learning expected. Scores of high-ability students with pretest scores covaried showed a significant increase after a year of ability grouping, but the effect diminished in subsequent years. Regular student scores overall showed a nonsignificant drop. High-ability students' scores on the Coopersmith Self-Esteem Inventory tended to rise while regular student scores tended to drop. Differences between the two groups were generally accounted for by pretest ITBS score differences. Students who had scored within one standard error of measurement above or below the criterion and who were prone to misclassification showed the same results. In general, students closest to the cutoff score were more affected by placement. High-ability students initially performed above expectations, but the effect was temporary.

Ability grouping is practiced widely in the U.S. and in Europe (Hess, 1978; Morrison, 1976). Though interpretations of the terms "ability" and "grouping" vary considerably, the method generally consists of teaching together students who function similarly in learning achievement (Findley & Bryan, 1975).

Reviewers of the hundreds of ability grouping studies conducted since the 1920s (Esposito, 1971; Hess, 1978; NEA, 1968; Wilson & Ribovich, 1973) have concluded that superior students may benefit from this method, but lower ranking students may be hurt. Primary areas of concern are exposure to undemanding curriculum and the social stigma attached to students in low-ability groups. Psychological theories which relate performance changes to feedback from reference groups or individuals predict decreased self-concept, achievement motivation, and academic performance for members of

low-ability groups (Bandura, 1982; Carver & Scheier, 1982; Festinger, 1954; Jackson, 1959). This is particularly true when students see themselves as potential members of a higher group which is visible (Richer, 1976). High-ability students' test scores may eventually drop due to lack of competition (Findley & Bryan, 1971; Reid, 1977).

Teachers and administrators have generally supported ability grouping (Wilson & Schmits, 1978), and various forms of it are frequently implemented in schools around the country. A noted example was a large Texas school district which grouped students for separate instruction in the basal subjects at the beginning of the fourth grade. Exit criteria were practically nonexistent (parental permission was required to remove students whose grades had dropped from the high-ability group), and students generally continued in the same group throughout their school life. The sole criterion for entrance to the high-ability group was a score at or above the 77th percentile on the Iowa Tests of Basic Skills (ITBS) composite or a score at the 77th percentile in math total and reading subtests. The high-ability treatment generally consisted of separate instruction in reading, language, mathematics, science, and Spanish instruction not available to regular students, along with better teachers and optional materials to supplement the district curriculum.

The effects of this district's ability-grouping methods on students' academic performance were assessed at the end of the initial year of grouping (Abadzi, 1984). Third-grade spring ITBS scores, on the basis of which students were grouped, were compared to fourth-grade spring ITBS scores. Student self-concept was measured at the beginning and at the end of the fourth grade. Since a strict cutoff score had been used for classifica-

Address correspondence to Helen Abadzi, Psychology Department, University of Texas at Arlington, Arlington, TX 76014.

tion, students' pretest ITBS scores could be covaried from posttest scores through a powerful quasiexperimental design, the regression-discontinuity analysis. In order to assess treatment effects on different achievement levels, high-ability and regular student groups were each divided into upper, middle, and lower performance subgroups.

A fixed-effects, levels-within-groups analysis of variance on fourth-grade composite ITBS scores with third-grade scores covaried showed that group assignment did have a significant effect which could not be attributed to pregrouping scores (Abadzi, 1984). Interactions of the levels within groups with the covariate were not significant, and a single slope could describe both regular- and high-ability groups.

Results were analyzed in the form of normal curve equivalents (NCEs), a transformation used to represent percentile scores on an interval scale. Like percentiles, NCEs range in values from 1 to 99, but lack direct empirical meaning. Conversion of raw scores to NCEs is achieved by first calculating percentiles and then applying a fixed conversion table (Tallmadge, 1976). Accordingly, the 1st, 50th, and 99th percentiles coincide with the 1st, 50th, and 99th NCEs while intermediate values differ (e.g. the 77th percentile corresponds to the 67th NCE).

Overall, high-ability students' scores rose by three NCEs while regular student scores dropped by one NCE unit. Gains and losses were most pronounced in the lower high-ability and the upper regular groups. The correlation between self-esteem and academic achievement was .34 ($p < .01$), but the difference in self-esteem between high-ability and regular students was reduced to nonsignificance when third-grade ITBS scores were covaried.

Ability-grouping assignment appears to have promoted different achievement profiles for upper regular and lower high-ability students (see Abadzi, 1984, Figure 1), particularly if those were a few NCEs apart before grouping. Abadzi (1984) examined the achievement and self-esteem scores of students who in the third grade had scored within one standard error of measurement (6.32 NCEs from the cutoff score, the 77th ITBS percentile). Scores of students who had been classified as high ability rose by 6.4 NCEs, while scores of regular students dropped by 3.1 NCEs from third to fourth grade. Self-esteem score differences between the two groups were significant at the posttest but not at the pretest. In response to these results and to various other problems with ability grouping, the school district substituted the program with flexible honors classes after the present cohort finished the fifth grade.

The strict classification cutoff score permitted the use of a robust design, but the magnitude of treatment effects in subsequent years could not be determined. For this reason, test scores were obtained from the same students in the spring of the cohort's fifth grade, before the program was discontinued, and in the spring of the sixth

grade when cohort students were in heterogeneous classes of various middle schools.

Method

The test scores of 284 high-ability and 383 regular students were studied in eight randomly selected schools which did not receive compensatory education funds (see Abadzi, 1984). These students had taken the Iowa Tests of Basic Skills each spring in grades 3, 4, 5, and 6, and the California Achievement Test (CAT) in grade 2. Ability-grouping decisions were made on the basis of ITBS test scores at the end of grade 3. Students were also given a slightly shortened version of the Self-Esteem Inventory (Coopersmith, 1981) in a 45-minute session one month after the beginning of grade 4, one month following ability grouping, one month before the end of grade 4, and one month before the end of grade 5 (Abadzi, 1984). Since students were scattered among middle schools throughout the city after program termination, the Self-Esteem Inventory was not administered to grade 6. (Permission for test use was obtained from the Consulting Psychologists Press, Palo Alto, CA.)

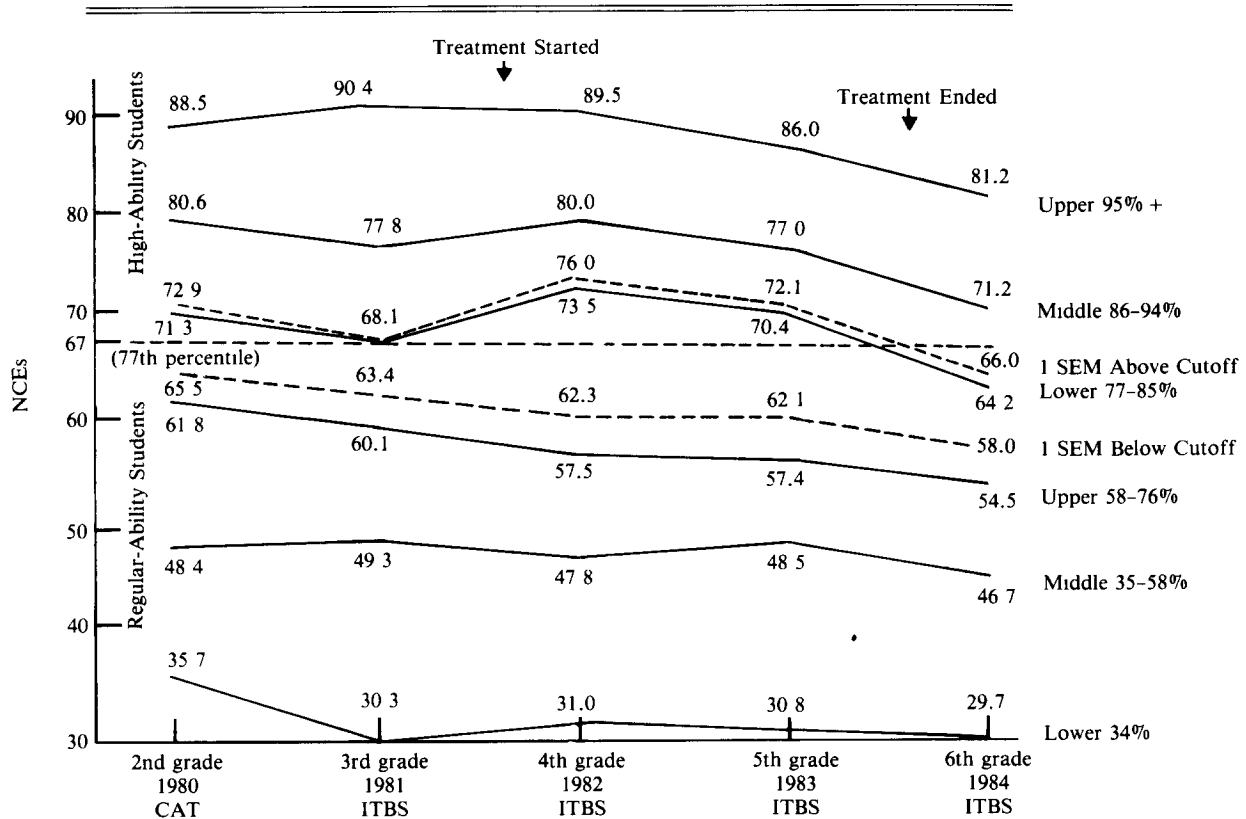
The data were analyzed at the student level in the form of NCEs, because NCEs are accurate in reflecting achievement levels, sensitive to small gains, and meaningful to users (Tallmadge, 1976). Students with some missing scores were included in statistical tests whenever possible. Self-Esteem Inventory responses showed an annual 12–13% mortality rate from grade 3 to grade 5; annual ITBS score mortality rate was 8–9% from grade 3 to grade 6. Attrition was slightly higher for regular students. However, the student dropout pattern in both groups apparently was not systematic because scores of students without missing observations in most measurements differed by less than one NCE unit from scores of students with missing observations. Results reported include students who may have missing values in some variables.

Results

The ITBS scores of all students generally showed a downward trend through five years of school (see Figure 1), which reflected a failure to learn as much as was expected, but high-ability student scores declined more than the scores of regular students. Ability grouping apparently did not alter the performance of the highest and lowest students; it mostly influenced the scores of students closer to the cutoff point, and even the magnitude of that effect gradually diminished.

A fixed-effects, levels-within-groups analysis of variance on fourth-grade composite ITBS scores with third-grade scores covaried had shown that group assignment did have a significant effect that could not be attributed to pregrouping scores, while the same analysis of second-

Figure 1.—Five-Year Means by Group and Performance Level



grade scores had shown no significant main effect for subsequent grouping. Similarly, an analysis of variance on fifth-grade composite ITBS scores (see Figure 1) with third-grade scores covaried produced a significant main effect for group, $F(1, 487) = 18.49, p < .01$, and a significant main effect for levels within groups, $F(4, 487) = 2.61, p < .05$. The analysis of variance on sixth-grade scores, after program termination, with third-grade scores covaried also produced a significant main effect, $F(1, 428) = 12.18, p < .01$, and a significant levels-within-groups effect, $F(4, 422) = 2.78, p < .03$. Though pretest and posttest self-esteem measures (see Figure 2) did not show a significant main effect for group after third-grade ITBS score variance was removed, the follow-up measure at the end of the fifth grade did so, $F(1, 340) = 3.81, p < .05$. The slopes of the high-ability and regular groups were not significantly different in any analysis, so in each year a common regression line could describe both groups.

After significant ITBS score gains during the first year in grouped classes, high-ability students showed statistically significant losses of 2.7 NCEs from grade 4 to grade 5, $t(257) = 5.67, p < .01$, and 5.1 NCEs from grade 5 to grade 6, $t(219) = 11.42, p < .01$. Regular students, who had showed nonsignificant losses in the first year of grouped classes, held the same scores in the

second year and showed smaller, though significant losses of 2.2 NCEs from grade 5 to grade 6, $t(284) = 4.39, p < .01$.

Students who scored a little above or below the cutoff point could have been classified in either group as a result of measurement error. Therefore, the performance of students who had scored within one standard error of measurement, 6.32 NCEs, on each side of the cutoff score was analyzed in order to assess effects on students who in previous years had had similar scores. After a year of grouping, high-ability student scores rose by three NCEs while regular student scores dropped by one NCE unit (see Figure 1). The gap continued to exist during the next two years, although it diminished as high-ability student scores dropped. Differences in 1983 and 1984 between high- and regular-ability students in this group became nonsignificant after 1981 ITBS scores were covaried. At the same time, high-ability students lost 3.4 NCEs from grade 4 to grade 5, $t(82) = 3.7, p < .01$, and 6.1 NCEs from grade 5 to grade 6, $t(75) = 7.8, p < .01$, while regular students showed a nonsignificant gain of .4 NCEs from grade 4 to grade 5 and a loss of 4.0 NCEs from grade 5 to grade 6, $t(41) = 2.9, p < .01$. Self-esteem scores of high-ability and regular students closest to the cutoff score differed significantly at the posttest, $t(111) = 2.17, p < .03$, though they had not

differed at the pretest and showed even greater differentiation at the follow-up test, $t(81) = 2.66, p < .01$, when high-ability scores rose by 4.9 percentage points, $t(42) = 3.1, p < .01$, and regular student scores rose by 2.1 points (t n.s.).

Discussion

High-ability students did not maintain in the long run the performance gains they made in the first year of grouping. At the end of the fifth grade, the second year in grouped classes, upper-level students had lost 3.6 NCEs in comparison to their preprogram scores, and middle-level students had returned to preprogram performance. Only lower-level high-ability students, who seemed to have benefitted the most from the treatment still showed some gains. On the other hand, upper- and middle-level regular students maintained their scores in the second year of the program after a drop following grouping. Score decline became more rapid for all subgroups after program termination, and high-ability students continued to show greater losses than regular students in the sixth grade. Program effects did not disappear, however. The upper regular students never recovered their initial score losses, and at the end of the sixth grade their score pattern was more similar to that of middle regular students, while in preprogram grades

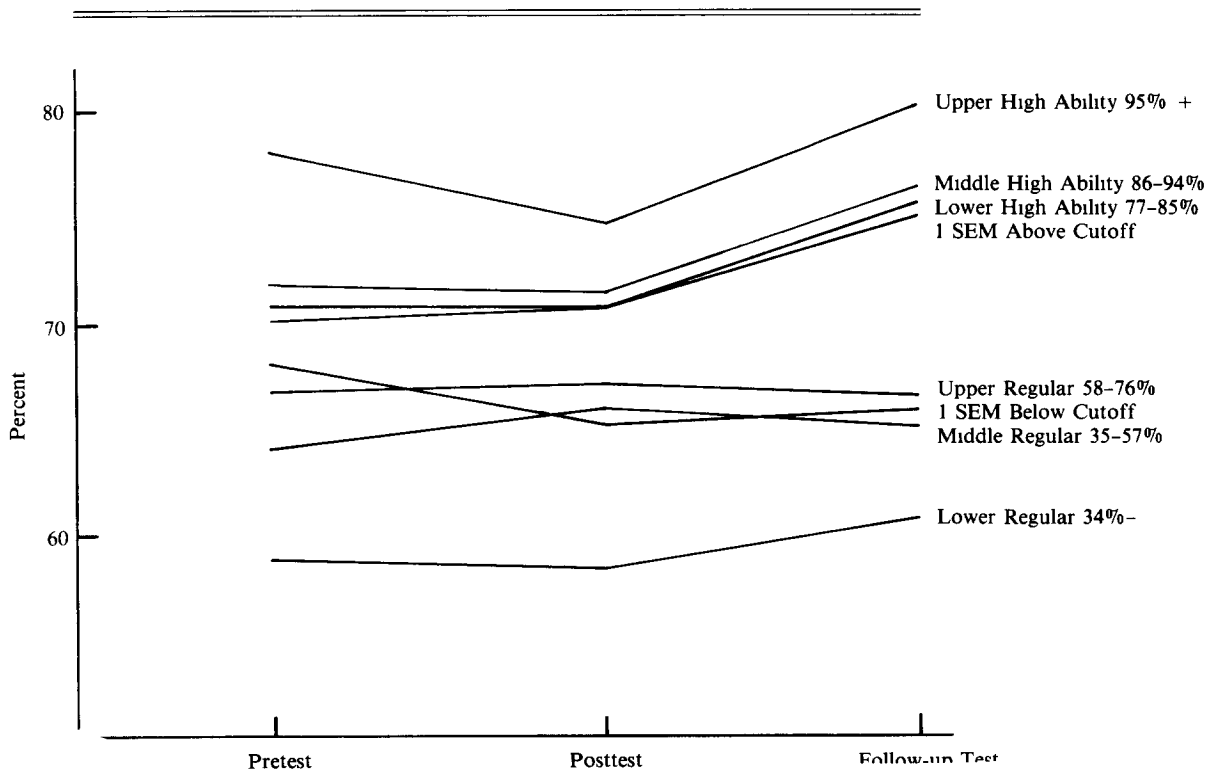
2 and 3 it had resembled the pattern of lower-level high-ability students (see Figure 1).

The general downward trend in performance expectations at the end of the elementary school is characteristic of the school district's test scores and, indeed, of the test scores in many other districts. The high-ability students' steady drop may be due to reduced achievement motivation brought on by a sense of invincibility, which the high status of the program combined with nonexistent exit criteria helped reinforce (Myer, 1972; cited in Hess, 1978). Rising high-ability student self-esteem scores at the same time achievement NCE scores were dropping provide some support to this hypothesis. The relative stability of regular student self-esteem scores in the fifth grade was accompanied by stable achievement scores.

Similar results have been reported before. Findley and Bryan (1975) showed that percentile losses were reported in approximately one-third of high-ability classes. Borg (1966) found no significant gains in 96 of the 144 classes monitored but also found that 15 of the 19 significant differences which favored ability grouping at the elementary school level failed to increase or persist after the first year.

The losses in NCEs by high-ability students represent a failure to learn as much as expected and are by no means obvious to school administrators. Most of these

Figure 2.—Self-Esteem Scores by Group and Performance Level



students score high, in the 80th and 90th percentile. The drop from the 90th to the 86th NCE unit (from 97th to 96th percentile) between grades 4 and 5 still represents a gain of 1.1 years from 7.2 to 8.3 in grade equivalents. On the other hand, upper regular students, who maintained their standing at the 57th NCE (64th percentile) gained 1.0 year per year of instruction. The use of NCEs, which is rarely seen in the literature, offers the opportunity to observe this phenomenon that cannot be adequately demonstrated with grade equivalents or percentiles. Most grouping studies, which were conducted in the 1950s and 1960s (NEA, 1968) do not show the measurement sophistication necessary to detect possible achievement decline of high-scoring students.

Students nearest the cutoff score were the ones most influenced by grouping. Very high- and very low-scoring students showed less achievement or self-concept change in response to changes in classmates and programs. In short, the only winners of ability grouping appeared to be the lower students in grouped classes who were given an educational and social opportunity to be with the best. Theories relating performance changes to reference groups or individual feedback (Bandura, 1982; Carver & Scheier, 1982; Festinger, 1954; Jackson, 1959; Richer, 1976) would predict such an outcome.

These results offer little support for complete ability grouping although they do not show this practice to be as deleterious as has been reported elsewhere. It is possible that upper-level high-ability students would have shown larger score decreases without it, but gifted-student education might be a more appropriate avenue to educate these students. The decision of the school district to replace the program with more flexible alternatives has proved to be wise.

REFERENCES

- Abadzi, H. (1984) Ability grouping effects on academic achievement and self-esteem in a southwestern school district. *Journal of Educational Research*, 77, 287-292.
- Bandura, A. (1982) Self-efficacy mechanism in human agency. *American Psychologist*, 41, 963-980.
- Borg, W. R. (1966). *Grouping in the public schools* (2nd ed.). Madison, WI: Dembar Educational Research Services, Inc.
- Carver, S. C., & Scheier, M. F. (1982) Control Theory: A useful conceptual framework for personality, social, clinical, and health psychology. *Psychological Bulletin*, 92, 111-135.
- Coopersmith, S. (1981) *Self-esteem inventories*. Palo Alto, CA: Consulting Psychologists Press.
- Esposito, D. (1971) *Homogeneous and heterogeneous grouping: Principal findings and implications of a re-research of the literature*. New York: Columbia University (ERIC Document Reproduction Service ED 056 150).
- Festinger, L. (1954) A theory of social comparison processes. *Human Relationships*, 2, 117-143.
- Findley, W. G., & Bryan, M. M. (1975) *The pros and cons of ability grouping*. Fastback No. 66. Bloomington, IN: Phi Delta Kappa Foundation.

- Hess, F. C. (Ed.) (1978). Issues in education: A documented look at seven important topics. *Educational Document Reproduction Services*, (ED 158 391).
- Jackson, J. M. (1959) Reference group processes in a formal organization. *Sociometry*, 22, 307-327.
- Morrison, C. M. (1976) *Ability grouping and mixed ability grouping in secondary schools*. Educational Issues Review No. 1, Scottish Council for Research in Education, Edinburgh, Scotland (ERIC Document Reproduction Service ED 142 535).
- Myer, K. R. (1972). *The Self Concept of Students in IPI*. Report on Educational Research. (Cited by Hess, 1978, see above).
- National Education Association (1968). *Ability grouping*. Washington, DC: Research summary 1968-35, Research Division, National Education Association.
- Reid, M. (1977) Mixed feelings. *Times educational supplement*, 3236, 15-17 (Abstract).
- Richer, S. (1976) Reference group theory and ability grouping: A convergence of sociological theory and educational research. *Sociology of Education*, 49, 65-71.
- Tallmadge, G. K. (1976). *Interpreting NCE's*. Washington, DC: ESEA Title I Evaluation Reporting System, Technical Paper No. 2, U.S. Office of Education.
- Wilson, R. M., & Ribovich, J. K. (1973). Ability grouping? Stop and reconsider. *Reading World*, 13, 84-91.
- Wilson, B. J., & Schmits, D. W. (1978) What's new in ability grouping. *Phi Delta Kappan*, 59, 535-536.

Depardon/magnum.

**IN SOME COUNTRIES,
ECONOMY
MEANS MAKING
ONE BOWL OF RICE
LAST THREE DAYS.**



**YOU ARE THE HOPE
OF THE HUNGRY.**

Interfaith Hunger Appeal, P.O. Box 1000
FDR Station, New York, NY 10150

PLEASE GIVE.

A public service of Catholic Relief Services,
Church World Service, the American Jewish
Joint Distribution Committee, Inc.,
this magazine, and the Advertising Council.

