



The National Center on
Education in the Inner Cities

The CEIC
REVIEW

A catalyst for merging research, policy, and practice.

Volume 10 • Number 5 • May 2001

Can Unlike Students Learn Together?

Research and Recommendations on Grade Retention, Tracking, and Grouping

Arthur Reynolds, University of Wisconsin-Madison; Margaret C. Wang, Distinguished Professor, Founder and Director, Temple University Center for Research in Human Development and Education; and Herb Walberg, University of Illinois at Chicago

With school districts' increased dedication to raising academic standards and abolishing social promotion, tremendous pressure has been placed on teachers and students to raise standardized test scores. While this may appear admirable from afar, its practical and real-life implications are not often as glowing. In fact, the push toward higher standards often leads to tracking, ability grouping, and grade retention—all of which have inherent problems.

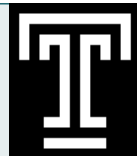
Tracking, grouping, and retention are widely practiced in the United States and in many other countries, and they are founded on both theory and research. Tracking, most often practiced in secondary schools, groups students into courses or sequences of courses of various levels of difficulty suited to their levels of achievement. Ability grouping, most often practiced in primary schools, assigns students within classrooms to homogeneous groups of like ability. Grade retention requires students who

have not attained achievement standards to repeat one or more grades. All three practices are based on the belief that children of like abilities or levels of achievement can learn together more efficiently than can heterogeneous students.

Other theories and research suggest that these practices may be inefficient and unwise. Some argue, for example, that students retained in grade may suffer declining self-concept which may deter their progress so that they are less likely to catch up with grade level standards. This is due, in part, to the fact that, by itself, grade retention does not address the causes of academic failure. Others counter that, to the contrary, such students would eventually fall further behind and drop out whether or not they were retained. To “socially promote” ill-prepared students would depreciate the value of the high school diplomas of those who meet rigorous standards. Similarly, some argue that it is more

efficient to teach subjects such as mathematics when students share similar abilities. For example, it would seem difficult for consumer mathematics and calculus to be learned efficiently in one group. Still, it may be argued that faster-learning students may benefit from helping slower-learning students. Schools might also provide more classroom time and intensified instructional services to at-risk students for remediation or to prevent them from falling behind in the first place.

The articles included in this issue of the CEIC *Review* will summarize some of the most recent theories and research emerging from the analysis of tracking, ability grouping, and grade retention. The synopses are of papers that were commissioned for a National Invitational Conference on Can Unlike Students Learn Together? Grade Retention, Tracking, and Grouping, sponsored by the Laboratory for Student Success and the National Center on Education in the Inner



The National Center on Education in the Inner Cities is a unit in the Temple University Center for Research in Human Development and Education, an interdisciplinary center devoted to fostering healthy developmental and educational success of children and families in this nation's urban communities. Inquiries about the work of the Center should be sent to Information Services, CRHDE, Temple University, 1301 Cecil B. Moore Avenue, Philadelphia, PA 19122-6091. Copyright © 2001

Cities at Temple University Center for Research in Human Development and Education, held on October 18–19, 2000 in Alexandria, Virginia. The papers discuss research findings in light of implications for policy, programs, and practices.

The conference organizers invited education leaders and scholars known for their differing views. Also represented were teachers, principals, superintendents, and state and federal officials. In addition to addressing the key issues framed by the commissioned papers, conference participants devoted much time to small work groups. The work groups discussed what is known from research on grade retention, tracking, and grouping, the impact these policies have on student achievement, and alternatives that can be implemented at the schoolwide and classroom level. In addition, the work groups offered next-step recommendations for helping unlike students learn together.

The recommendations fell into two categories. The conferees did not reach complete consensus on all issues and interpretations of findings, but agreed that the major issues and views had been expressed in the papers and the work groups.

Retention

While there is no magical cure for the ails of retention, alternatives must be examined before it's too late—that is, before a student is about to be retained. By studying the experiences of successful students and making findings available to practitioners, researchers can help teachers focus on using teaching strategies

that have been proven successful. The following recommendations could also be helpful.

- Encourage preschool enrollment in order to reduce retention rates.
- Require full-day kindergarten.
- Provide remediation that is proportional to children's academic needs without regard to whether they are retained.
- Develop a strong advisor network that will allow faculty to get to know the students.
- Maximize peer relationships through cooperative learning and tutoring.
- Shift to interest-based learning where high school students are exposed to career-based or project-based education instead of the lecture and test-taking practices now used.
- Extend the academic calendar either to year-round schooling or longer school days.
- Focus on retaining motivated and qualified teachers.
- Hold teachers to expectations of higher levels of curriculum and instruction.

Researchers' and practitioners' voices aren't the only ones that should be heard. Parents must also become more involved in helping their children avoid retention. Some ways to boost parent involvement are:

- Develop "tip sheets" that have helpful hints on how parents can get more involved in their child's education.
- Develop parent education and outreach programs.
- Don't wait until students are at risk of failing; begin communication with parents at an early stage.

Grouping and Tracking

Why does neither retention, grouping, nor tracking enhance the academic progress of most children? Unfortunately, in many schools, grouping and tracking have led to stagnant and generalized courses designed to meet minimum curriculum standards. In order for true progress to be made, the intent, purpose, and design of grouped classes must be examined and a high level of integrity maintained. The following recommendations deserve further consideration.

- Consider multi-age classrooms as a way to enrich children's learning and development.
- Prioritize collaborative efforts among schools, employers, and higher education in supporting academic excellence.
- Have goal conferences with students. Integrate students' self-assessments into decisions on their grouping.
- Provide stronger teacher and principal preparation coursework that will address diversity in learning rates and styles.
- Keep grouping flexible.
- Grouping should include high expectations, rigorous curriculum, and equitable access to high-quality instruction.
- Promote cultural awareness that will help teachers meet the diverse needs of their students.
- Promote public awareness. Educate the community on the best ways to group students.
- Hold administrators, teachers, parents, and students accountable. All must work together to achieve the optimum level of student success. ☘

Dropout in Relation to Grade Retention

An Accounting from the Beginning School Study

Karl L. Alexander, Johns Hopkins University; Doris R. Entwisle, Johns Hopkins University; Susan L. Dauber, The Spencer Foundation; and Nader Kabbani, U.S. Department of Agriculture

This paper examines the relationship between retention in the primary grades and high school dropout from the perspective of the Beginning School Study (BSS), a panel of Baltimore school children who began first grade in the fall of 1982 in 20 city public schools. It extends earlier work on effects of grade retention, which investigated consequences for children's academic performance and socio-emotional development through the middle-school years.

When children are not keeping up, is it better to hold them back or move them ahead? That is the question addressed in earlier work. For answers, the experience of first-, second-, and third-grade repeaters, and, as a group, children held back in grades four through seven were examined. Their academic progress and attitudes were monitored from the fall of first grade, before anyone had been held back, to the end of seventh grade (in the case of repeaters) or eighth grade (in the case of children never retained).

Retention rates in the BSS panel are quite high: almost 17% of the cohort was held back at the end of first grade, and over 5 years (ordinarily the end of elementary school) the figure stood at 40% retained at least once, with many double retentions in the mix. The study monitored these children's academic and socio-emotional standing as their school careers were being launched in 1982, and BSS fieldwork has continued an intensive schedule in following years.

The analyses reported in the authors' 1994 work spanned the first 8 years of the group's schooling; the present paper extends that time frame to the end of high school and to a rather different outcome: high school dropout. The earlier study generated considerable interest, and some controversy. Retention's effects were assessed in a host of ways and, though the results were complex, it was concluded that repeaters in most instances were doing better in elementary school after retention than they had been doing before, and that these advances generally held up for a number of years (although in diminishing measure). The experience certainly did not set them back academically (as reflected in achievement test scores and report card marks). Nor was there evidence of great stigma attaching to grade retention. Instead, in most of the comparisons, repeating a grade was associated with improved attitudes toward self and school. These findings contradicted the results of most similar contemporary studies. However, despite the benefits of retention on school achievement and self-esteem, retained students are more likely to drop out of school. In fact, repeating a grade in the BSS increases dropout risk, and later the risk of noncompletion, from three- to eight-fold.

The standards for judging an educational intervention often are left implicit, as though the appropriate criteria were self-evident. In fact, the issue is far

from straightforward—is improvement sufficient, for instance, even if that improvement does not bring poor performing students up to desired levels? One principle seems fundamental: an intervention intended to help should “do no harm,” and there can be no doubt that elevated dropout risk of the magnitude seen in these results qualifies as “harm.” This challenges earlier conclusions on the merits of grade retention, and an accounting of the dropout results in light of those earlier findings is offered.

Age-Grading and the Dropout Dynamic

Children held back in the upper grades and multiple repeaters are especially prone to leave school without degrees, but single repeaters are also at elevated risk. The 1994 study concluded that double repeaters and first-grade repeaters were helped least by repeating a grade, so for them to have elevated levels of dropping out and non-completion is not surprising. But single repeaters who were held back in second grade also drop out in numbers greater than expected, and in at least one comparison so do third-grade repeaters. If repeating a grade in elementary school boosts children's school performance and shores up their self-regard, as exhibited in the BSS, why would it later increase dropout risk?

The fact that this risk is especially pronounced among repeaters held back in grades 4–7, as we find, is significant. When

these children were held back, they were not as academically far behind their promoted classmates as were children held back earlier. If retention were simply a proxy for relevant academic difficulties, then repeating first or second grade, and not grades 4–7, would pose the greatest problems later, but that is not the case.

If not academics, then what? The social side of schooling seems a likely candidate. Grade retention takes children off the prescribed timetable of grade progressions in a rigidly age-graded system. This makes them conspicuous and complicates their social integration. Being “off-time” in school can cause problems at any age, but conditions peculiar to adolescence, the onset of puberty, and the impending transition to middle school very likely heightens them.

The early adolescent years (typically age 12–14) are a time of heightened self-consciousness, when “fitting in” is paramount, but “fitting in” is not easy for late repeaters. The separation from their friends is still fresh when the time comes to change levels of school, and the disruption of peer groups they suffer is two-fold—their age-peers move on to middle school while they are left behind with younger classmates whom they may view as lower on the age/status hierarchy. Since repeating is less common in the upper elementary years than in first and second grade, there are relatively few age-peers available in late repeaters’ classes to help ease their adjustment.

Repeaters’ academic standing began to slide when they moved from elementary to middle school. Reflecting transition shock, their

marks and test scores began to trail off at that point, and although they usually remained ahead of where they originated, there was little room for them to absorb additional setbacks.

Thus, repeaters’ situation in middle school was precarious, and even greater challenges awaited them at the transition into ninth grade. Any school transition is hard, but the transition to high school is especially difficult. Relative to middle schools, high schools are larger, more bureaucratic, impersonal, and academically demanding. Under such circumstances, even high-achieving, well-integrated students often experience difficulty. And what of repeaters? Their academic and social standing are low, which leaves them especially vulnerable. Consider this one “symptom”: in their ninth year of school, future dropouts in the BSS averaged 46.8 absences compared with an average of 13.5 absences among nondropouts. With 47 *recorded* absences, these students were missing about one day out of every four, which was interpreted as a signal that the dropout process already had begun.

Some Thoughts on Policy and Practice

The new evidence presented here showing that grade retention elevates dropout risk certainly reinforces the conviction that retaining children ought to be a last resort. But as before, it is still believed that repeating a year may be appropriate when extra time is needed to consolidate skills and master material missed the first time through.

Still, for most children under most circumstances, traditional

retention (i.e., grade repetition without supplemental services) ought to be rare. But candidates for retention typically are far behind academically and often exhibit serious behavior problems. Absent an effective intervention, many of these children are on a path that will lead to dropping out whether they are held back or not. Ignoring the problem (i.e., simply moving them ahead to the next grade level) and hoping for the best certainly is a formula for failure. Children who are far behind and struggling don’t suddenly spurt ahead, even though a spurt is what is required for them to catch up.

The first priority should be to keep children from reaching the point where they are retention candidates in the first place. Many poor and minority children start school already behind, but it is known that high-quality preschool programs can enhance school readiness. More of those programs are needed, and more disadvantaged children need to have access to them. Likewise, there is a need for high-quality, full-day kindergarten and supplemental services to help preserve the gains realized as a result of those early interventions.

Children learn at different rates. Yet all are expected to be “ready” for first grade at age six; they are expected to move in lockstep annually thereafter from one grade to the next; and within the year, they are expected to master the curriculum in roughly the same time frame: nine months, fall to spring. The current calendar-driven model of schooling sets a severe pace; children who aren’t caught up when the teacher

(Dropout, continued on p. 12)

Grade Retention and School Dropout

Another Look at the Evidence

Judy Temple, Northern Illinois University; Arthur Reynolds, University of Wisconsin-Madison; and Suh-Ruu Ou, University of Wisconsin-Madison

Despite the current emphasis on grade retention as an educational policy designed to help low-achieving students, the majority of empirical studies suggest that grade retention does not benefit most of the students it is designed to help. Years after being retained, students have significantly lower achievement than similar students who were not retained. Many retained students never catch up to their promoted same-age peers with similarly low test scores. Whatever performance advantage retained students have over their younger, same-grade peers is short-lived, as they typically fall behind these students after one or two years.

Even more striking is the strong positive correlation between grade retention and dropping out of high school. Several longitudinal studies indicate that, relative to low-achieving students who are promoted to the next grade, retained students are significantly more likely to drop out of school. After accounting for socioeconomic status and prior performance, dropout rates for retained students often exceed comparable promoted students by 50% or more.

The Present Study

Using data in the Chicago Longitudinal Study (CLS), this paper addresses the following questions:

1. After children's growth rates in achievement prior to retention (and other factors) are taken into account, is grade retention associated with

significantly lower levels of school achievement and higher rates of school dropout?

2. Among children who are retained during the early school years, is participation in a comprehensive instructional intervention associated with improved school achievement and a lower likelihood of dropping out of school? Does this participation lead to better performance than promotion with remediation?

The Chicago Longitudinal Study

The Chicago Longitudinal Study is an ongoing investigation of 1,539 low-income, minority (93% African American) children born in 1980 who attended kindergarten programs in 25 Chicago public schools in 1985–1986. Most children attended the Title I Child-Parent Center (CPC) Program, a comprehensive pre-school and school-age preventive intervention for children from high-poverty neighborhoods. Children in the CLS completed the elementary grades (eighth or ninth grade) in 1995 before the new retention policy in Chicago was enacted in 1996.

The study sample for this paper includes 1,267 students who enrolled in the Chicago public schools for at least six years (from kindergarten to ninth grade) and whose school dropout status was known by age 20. Children who have left the study or cannot be located are similar to those that remain in the sample on measures of kindergarten achievement and socioeconomic status.

Since 1986, the CLS has collected data from multiple sources. Information on grade retention and high school completion as of January 2000 were obtained from school records. School records provided descriptive information on children including gender, race, and name of the school in which the student is enrolled at the end of each school year. Standardized test scores in reading and math were obtained annually from the beginning of kindergarten (1985) through ninth grade (1995). Teacher and parent surveys were used to obtain information on classroom adjustment, parent involvement, and family background.

Two measures of educational attainment were used in analyzing the effects of retention. Data were collected from school records, surveys, and interviews from youth and, if necessary, their parents. "School dropout" measured whether youth left their formal education or diploma-granting high school prior to graduation for any reason other than death or school transfer. Students who graduated from high school or were active in high school were defined as nondropouts. "High school completion" measured whether youth completed their secondary education with an official diploma or were awarded a GED. All others, including those who remained in high school as of January 2000, were coded as noncompleters. School records from kindergarten to eighth grade measured grade retention.

Characteristics of Retained and Promoted Students

Of the 1,267 youth for whom school dropout status was known, 360 (or 28.4%) were retained at least once from kindergarten to eighth grade. Importantly, the retained students had higher dropout rates and lower completion rates than promoted students. The two groups of students are also different in other ways. Students who participated in the Child-Parent Center program were less likely to be retained, and students who were retained had lower achievement levels both before retention occurred and years later (as of age 14).

Relatively few kindergartners repeated a grade ($n = 12$), while the largest group of students was retained in first grade ($n = 134$), followed by third grade ($n = 71$), second grade ($n = 76$), and fourth grade ($n = 48$). Fewer numbers of students were retained in fifth to eighth grades.

Predictors of Grade Retention

Before investigating the association between grade retention and high school completion or dropout, a comprehensive set of predictors of retention was examined, including child and family background, early adjustment indicators (kindergarten and first-grade academic performance and achievement), and intervening school experiences (e.g., school mobility and special education placement). In order of magnitude, the following factors *increased* the odds of being retained: low family income (2.22; children eligible for a subsidized lunch had twice the risk of retention than those not eligible); sex of child (2.04; boys had twice the risk of

retention), and number of school moves from ages 10 to 14 (1.28).

The following factors *decreased* the odds of being retained: overage at kindergarten entry (0.17), number of years of average or better parent involvement in school (0.76), reading and math achievement in first grade (0.97), grade in reading in first grade (0.67), and math achievement in kindergarten (0.99). Findings that the number of school moves increases the risk of retention and parent involvement in school decreases the risk are relatively new, and especially significant. Variables such as race/ethnicity, parent education, years of CPC intervention, residence in a high-poverty school attendance area, and special education placement were not associated with retention.

Discussion

Findings indicate that grade retention—no matter when it occurs—is associated with significantly lower levels of school achievement and higher rates of school dropout. The students who were retained fell further behind their similarly low-achieving former classmates as early as kindergarten and first grade. By the end of their eighth-grade year, retained students were 1 to 2 years behind these former classmates. Retained students had a rate of school dropout that was 25% higher than that of promoted students (controlling for preretention achievement growth and other factors).

Does grade retention harm students, or are the large estimated adverse effects of grade retention due at least in part to the difficulty in controlling for

observed and unobserved differences between retained and promoted students that may be correlated with later educational attainment? The main strength of this study was the inclusion of a variety of preretention control variables such as achievement at different times that take account of such differences. Results indicated that, although there were substantial differences between the unadjusted and adjusted models, both indicated a significant link between grade retention and school dropout rates as well as lower rates of school completion.

RETENTION PLUS REMEDIATION

The finding that students who were retained in the first three grades did not benefit academically from 1 to 3 years of participation in the Child-Parent Center program suggests that retention plus remediation strategies may not prevent the typical achievement declines that have been shown for simple grade retention without remediation. Indeed, the CPC follow-on intervention is more comprehensive and longer-lasting than most remedial services that retained students receive under many current retention practices in schools. Moreover, comparable students who were promoted (instead of retained) and then participated in intervention for 1 to 3 years had substantial performance advantages over retained students who participated in the intervention.

THE MISSING LINK: PREVENTION OF LEARNING DIFFICULTIES

Grade retention is a response to academic problems. Little
(**Retention**, continued on p. 21)

Meta-Analysis of the Effects of Grade Retention 1990-1999

A Basis for Moving Beyond Grade Retention and Social Promotion

Shane Jimerson, University of California-Santa Barbara

“Flunking,” “retained,” “being held back,” and “grade retention” refer to the practice of requiring a student who has been in a given grade for a full school year to remain at the same grade level for a subsequent year. There is an abundance of research and scholarly analysis examining the efficacy of grade retention. Research published between 1900 and 1989 produced mixed results regarding the efficacy of early grade retention on ameliorating children’s socio-emotional and achievement needs. Concerns regarding the quality of many studies of grade retention have been presented in several reviews and reiterated in recent publications. These methodological concerns include: (a) data collected 30–40 years ago may be outdated; (b) characteristics of comparison groups are rarely delineated; (c) comparing pre- and post-test scores of retained students rather than employing a comparison group may pose problems; (d) most studies do not consider socio-emotional outcomes; (e) remedial services during the repeated year are rarely documented, and (f) most studies do not examine the long-term outcomes associated with early grade retention. These methodological considerations limit unequivocal conclusions from any single study; however, the confluence of results clearly warrants further consideration. This study provides a meta-analysis of empirical studies

published between 1990–1999 examining the efficacy of grade retention.

Methodology Used in Present Study

This project began with a systematic search of the literature to identify studies of grade retention published between 1990 and 1999. Descriptors used to search reference databases included grade retention, grade repetition, nonpromotion, grade failure, flunked, failed, retained, and other related synonyms. Computerized reference databases searched included the Psychological Information Abstracts (PsychINFO) and the Education Research Information Center (ERIC). Results of these searches yielded over 400 references between 1990–1999. In addition, other studies were identified through a review of references in each publication obtained, resulting in nearly 450 references for consideration.

The following selection criteria were used to reduce the bibliography to a core set of research appropriate for this review. To be included in this review: (a) the research must have been presented in a professional publication (e.g., journal article or book); (b) the results must have addressed the efficacy of grade retention (i.e., achievement, socio-emotional, or other); (c) the study must have included an identifiable comparison group of promoted students; and (d) the research

must have been published during the past decade (i.e., 1990–1999). Based on the above selection criteria, 19 articles were included in this review.

Procedures for Summary and Analysis

The plan for summary and analysis of the 19 articles was to provide the following information: (a) variables used for matching the comparison group and retained students (i.e., IQ, academic achievement, socio-emotional and behavioral adjustment, socio-economic status (SES), and gender); (b) specification of the age/grade at which retention and the measurement of outcome variables occurred; (c) designation of the location of the sample population; (d) a review of analyses comparing retained students to a matched group (i.e., academic achievement and socio-emotional and behavioral adjustment); and (e) the overall conclusion of the author(s) regarding the efficacy of grade retention.

SUMMARY OF EACH STUDY

Each study was examined to identify the variables used for matching and the grade level at which the outcomes were studied. Most studies included only students retained during kindergarten, 1st, 2nd, and 3rd grades, however, a few included students retained kindergarten through eighth grade. Population samples for these studies are distributed across the nation.

STATISTICAL META-ANALYSIS

Statistical meta-analysis is based on the concept of effect size (ES). Computation of the effect size is a statistical procedure that allows researchers to systematically pool the results across studies, to examine the benefit or harm of an educational intervention. Meta-analysis procedures result in a measure of the difference between the two groups expressed in quantitative units that are comparable across studies. Each effect size is standardized with respect to the comparison group standard deviation; thus, it is possible to combine the results from different measures at different grade levels. A negative effect size suggests that an intervention (retention) had a negative effect relative to the comparison group of promoted students.

Consistent with past meta-analyses of grade retention, the effect size was defined as the difference between the mean of the retained group, X_r , and the mean of the comparison (promoted) group, X_p , divided by the standard deviation of the comparison group, S_p ($ES = (X_r - X_p)/S_p$). Group means adjusted for past differences were used when available and calculated when possible. In studies where the necessary group means and standard deviations were not included in the publication, the authors were contacted to provide the necessary data. For a few analyses, the effect sizes were estimated by working backwards from the reported significance tests.

Many of the results examined in the meta-analysis

fell into two categories:

(1) academic achievement and (2) socio-emotional/behavioral adjustment. Academic achievement analyses included language, arts, reading, mathematics, and grade point average. Socio-emotional/behavioral adjustment analyses included social (e.g., peer competence), emotional (e.g., internalizing problems), and behavioral (e.g., externalizing problems). Analyses also included self-concept, general academic adjustment, and attendance.

Because some studies yielded one effect size and others yielded as many as 25, additional analyses were performed to discern whether any single study had produced substantial distortions in the effect sizes. For each study, all individual effect sizes were summed and averaged. These means were used to recalculate the effect sizes for each of the outcomes. This procedure gives each study equal weight in determining the overall result. Effect sizes weighted by study were not found to differ significantly from reported effect sizes weighted by the number of effects; thus, they do not appear in the results.

Brief Overview of Findings

Most studies published during the past decade utilized a combination of IQ, academic achievement, socio-emotional adjustment, SES, and gender to match groups or control analyses between the comparison group and the retained students. Of the 19 studies included, 15 examined outcomes through grade seven; only five included outcomes

during eighth grade and beyond. Overall, results of the meta-analyses yielded average effect sizes indicating that the retained groups were .30 standard deviation units below the matched comparison groups. The average effect size for socio-emotional/behavioral adjustment (-.19) and academic achievement (-.40) favored the matched comparison group over the retained group of students. The results indicate that the greatest differences between groups were evident on measures of attendance, reading, mathematics, language, and emotional adjustment (-.65, -.56, -.49, -.40, and -.25, respectively). In regards to the authors' conclusions pertaining to the efficacy of grade retention as an intervention, of the 19 studies comparing retained students with a matched control group, the authors of 15 studies (79%) concluded that grade retention is ineffective as an intervention for academic achievement and socio-emotional adjustment.

In Sum

This meta-analysis includes studies published between 1990 and 1999 provides additional information regarding the effectiveness of grade retention. The results of research published during the past are very similar to findings reported throughout the remainder of the century. In particular, these studies fail to demonstrate that grade retention provides greater benefits to students with academic or adjustment difficulties than does promotion to the next grade. Thus, it seems practical to move beyond the question "to retain or (**Analysis**, continued on p. 21)

Can Unlike Children Learn Together?

A Question that Goes to the Heart of Democratic Public Schooling

Jeannie Oakes, UCLA and Martin Lipton, UCLA

The question “Can unlike children learn together?” goes to the very heart of democratic public schooling. This paper argues that for most of the 20th century, schools have constructed multiple categories of “unlikeness” or unlike ability, and that these categories were created or soon appropriated to mean “children who cannot learn together.” Important evidence collected throughout the century, but most especially in the past twenty years, reveals that school categories favoring children’s likeness, rather than their “unlikeness” promise to improve educational fairness and the country’s educational quality.

Ability grouping has been bolstered by the argument that equal opportunity in a democracy requires schools to provide each student access to the kind of knowledge and skills that best suit his or her abilities and likely adult lives. To make the argument more palatable in a culture that, rhetorically at least, values classless and colorblind policies, educators and policymakers have reified categorical differences among people. So, in contemporary schools, there are “gifted” students, “average” students, “Title I” students, “learning disabled” students, and so on, in order to justify the different access and opportunities students receive. Assessment and evaluation technology permits schools to categorize, compare, rank, and assign value to students’ abilities and achievements in relationship to

one another (as well as to students in other schools, states, and countries—past and present).

Deep Seated Myths and Prejudices and the Internal Organization of Schools

Homogeneous grouping began in earnest early in the 20th century. It matched the prevailing IQ conception of intelligence, behavioral theories of learning, a transmission and training model of teaching, and the factory model of school organization. It fit with schools’ role in maintaining a social and economic order in which those with power and privilege routinely pass on their advantages to their children. Homogeneous grouping embodied a belief that permeated schooling during the 20th century—that we understand most about students when we look at their differences, and the more differences that can be identified, the better our understanding and teaching.

Homogeneous grouping provided policymakers and educators a way to “solve” an array of problems attributed to the growing diversity of students. New immigrants needed to learn English and American ways. Factories needed trained workers. Urban youth needed supervision. And schools needed to continue their traditional role of providing high-status knowledge to prepare some students for the professions. Policymakers defined equal educational opportunity as giving

all students the chance to prepare for largely predetermined and certainly different adult lives. Concurrently, two phenomena shaped a uniquely American definition of democratic schooling: (1) universal schooling would give all students some access to knowledge; (2) IQ could justify differentiated access to knowledge as a hallmark of democratic fairness.

IQ and Testing

While most current grouping practices don’t rely on IQ—at least exclusively—the early dependence upon it set a pattern that continues today. Standardized achievement tests, strikingly similar to IQ tests, play an important role in dividing students into ability groups and qualifying students for compensatory education programs; standardized language proficiency tests determine which class “level” is appropriate for limited English students. In conjunction with other measures, IQ remains central in the identification of gifted and cognitively disabled students.

The Press for Universal Education

Over the course of the 20th century, compulsory education laws and the necessity of a high-school diploma drew more and more students to school—even those previously considered uneducable. States and local school systems developed an array of special programs for students who, in earlier times,

simply would not have been in school. By the 1960s, the federal government had turned to special categorical programs as its principal way to guarantee education for all American students. The Elementary and Secondary Education Act (ESEA) provided categorical funding for “educationally deprived” students. *Lau et. al. v. Nichols et. al.* was brought on behalf of Chinese students in San Francisco and led to legislation requiring that all schools provide special assistance to their students whose native language is not English. The Individuals with Disabilities Education Act (IDEA) provided funds to classify students with physical and neurological problems and provide these students with special education programs when it was believed that they could not be accommodated in regular programs. Advocates for “gifted” students increasingly used the “bell curve” logic to argue that the gifted and the cognitively disabled are like a pair of bookends, and that those at the high end of the curve also required special support because they are as different from “normal” students as the disabled.

Educators responded in culturally predictable ways. They identified students who were “different,” diagnosed their differences as scientifically as possible, and assigned them to a category. They then grouped students for instruction with others in the same category and tailored curriculum and teaching to what each group “needs” and what the culture expects. So, today, educators routinely assign

“normal” students to “regular” classes at different levels (e.g., high, average, slow). They place the others in “special” programs for learning disabled, behavioral problems, gifted, limited English, poverty-related academic deficiencies, and more. Within homogenous groups, teachers assume students can move lock step through lessons and that all class members will profit from the same instruction on the same content at the same pace. Lurking just beneath the surface of these highly rationalized practices, however, are the illusion of homogeneity, the social construction of classifications, the prevailing biases of race and social class, and self-fulfilling prophecies of opportunities and outcomes.

Socially and Politically Constructed Categories

The considerable student differences *within* supposedly homogenous classes are obvious and well documented. And yet, for most people, the characteristics and categories by which students are sorted remain more salient than the “exceptions” that impugn those categories. Many educational constructs, including those used to classify students, began as narrowly defined, highly specialized, technical terms or measures. However, as they make their way from research to professional journals and teacher preparation programs to popular media to the everyday talk of policymakers and the public, they lose their narrow definitions and specialized uses. What may have begun as specific technical concepts or as informal notions such as “at

risk,” “gifted,” “high ability,” “college prep,” “attention deficit,” “hyperactive,” “handicapped,” etc. are quickly reified and become a deeply embedded feature of students’ identities in their own and others’ minds.

Race and Social Class Bias

African American, Latino, and low-income students are consistently overrepresented in low-ability, remedial, and special education classes and programs. They are less likely to participate in “gifted” programs. This is not surprising, given that grouping practices grew from the once-accepted practice of preparing students of different racial, ethnic and social-class backgrounds for their separate (and unequal) places in society.

In part, placement patterns reflect differences in minority and White students’ learning opportunities that affect their preparation and achievements. But they also reflect the fact that US schools use White, largely middle-class standards of culture and language styles to screen for academic ability and talent. Teachers and school psychologists sometimes mistake the language and dialect differences of Hispanic and Black students for poor language skills, conceptual misunderstandings, or even poor attitudes. An additional hazard for students of color is that schools often confuse cultural differences with cognitive disabilities, particularly retardation.

Researchers have noted for the past 25 years that students with identical IQs but different race and social class have been (Unlike, continued on p. 21)

Race-ethnicity, Social Background, and Grade Retention

An Analysis of the Last Thirty Years

Robert M. Hauser; Devah I. Pager; and Solon J. Simmons, University of Wisconsin-Madison

Despite the visible popularity of policies to end social promotion, little is known about the prevalence of grade retention in American schools or about the effects of race-ethnicity and other social and economic background characteristics on retention. This paper reports analyses of race-ethnic differences in age-grade retardation, or enrollment below the modal grade level for a child's age using data from the October Current Population Surveys (CPS) from 1972 to 1998. Age-grade retardation is employed here as a convenient measure of grade retention (and no broader meaning is either intended or implied); at older ages, school dropout is treated as a component of age-grade retardation. The analysis focuses on dependent children at selected ages from 6 to 17. These ages span the period between normative entry to grade school and the later years of high school. Typical developmental patterns of retention and of differentials in retention can be observed by looking at several ages. By combining data from 27 annual surveys, trends in retention practices across three decades are identified.

From 1972 to 1998, the October CPS data files include between 57,500 and 63,500 cases at each age. The data are drawn from a specially prepared file that attaches characteristics of households and of householders to demographic characteristics and enrollment data for school-age youth and ensures uniformity in the measures from year to year. For each youth in the sample, sex, race-ethnicity, enrollment status, grade level, region of

residence, and metropolitan location can be determined. We have linked several social and economic characteristics of the household and householders to each child or youth's record: family income, number of children in the household, single-parent household, education of household head and of spouse of head, head or spouse without an occupation, occupational status of head and spouse of head, and housing tenure.

Retention in the Primary and Secondary Grades

National rates of age-grade retardation were examined by age, sex, and race-ethnicity for three-year age groups at ages 6 to 17 from 1971 to 1998. Age-grade retardation increased in every age group (6 to 8, 9 to 11, 12 to 14, and 15 to 17) from cohorts of the early 1970s through those of the middle to late 1980s. Age-grade retardation increased at ages 15 to 17 after the mid-1970s despite a slow decline in its early school dropout component throughout the period. That is, grade retention increased while dropout decreased. Peak rates occurred earlier at older than at younger ages, suggesting that policy changes occurred in specific calendar years, rather than consistently throughout the life of successive birth cohorts. Among cohorts entering school after 1970, the percentage enrolled below the modal grade level was never less than 10% at ages 6 to 8, and it exceeded 20% for cohorts of the late 1980s. Increased ages at entry to the first grade are probably due, in part, to changes in legal ages at school entry as well as to retention

in kindergarten and the primary grades. The trendlines suggest that age-grade retardation has declined slightly for cohorts entering school after the mid-1980s, but rates have not approached the much lower levels of the early 1970s. Overall, a large share of each birth cohort now experiences grade retention during elementary school. By the end of high school, over 35% of students aged 15 to 17 in the last cohort for which complete data exist are below the modal grade for their age.

Social Differences in Retention

While there are similarities in the age pattern of grade retardation among major population groups—boys and girls and majority and minority groups—there are also substantial differences in rates of grade retardation among them, many of which develop well after school entry. The gender differential gradually increases with age from 5 percentage points at ages 6 to 8 to 10 percentage points at ages 15 to 17. That is, boys are initially more likely than girls to be placed below the modal grade for their age, and they fall further behind girls as they pass through childhood and adolescence.

The differentiation of age-grade relationships by race and ethnicity is even more striking than that by gender. Here, unlike the case of gender differentiation, the rates of age-grade retardation are very similar among Whites, Blacks, and Hispanics at ages 6 to 8. However, by ages 9 to 11, the percentages enrolled below modal grade levels are typically 5 to 10 percentage points higher among Blacks or

Hispanics than among Whites. The differentials continue to grow with age and, at ages 15 to 17, rates of grade retardation range from 40 to 50% among Blacks and Hispanics, while they have drifted up more gradually from 25% to 35% among Whites.

Race-ethnicity, Geography, and Social Background

These patterns in the gross distributions for different social groups are quite stark. In order to analyze the causes and correlates of age–grade retardation in more detail, and to separate gross differences from net effects, a logistic regression analyses has been carried out on enrollment below modal grade level vs. enrollment at or above modal grade level for ages 6, 9, 12, 15, and 17. At these ages, the modal October grade levels are 1, 4, 7, 10, and 12. In these models, the net influence of geography and social background on trends and differentials in school retention can be explored.

At age 6, many of the effects of social and economic background characteristics are small. Mainly, this reflects the lack of social differentiation at school entry. One strong and expected effect is that of gender: The odds of boys' enrollment below the first grade, other things being equal, are 40% higher than those of girls. At each successive age, social and geographic differentials become more pronounced: gross race–ethnic differentials become larger, the effects of socioeconomic background variables increase, central cities become notably more likely to have overage students than suburbs, and regional differences between the South and all other regions become sharper (with

students in the South significantly more likely to be below modal grade for age).

There is a geographic pattern to the ordering of cities: southern cities have the highest rates of age–grade retardation, while northern and western cities have the lowest rates. Also, there is increasing differentiation between central cities and their suburbs with increases in age. By age 17, rates of age–grade retardation are roughly 20% higher in central cities than in suburbs, controlling for social background characteristics.

Perhaps most striking in these results are the net effects of social background vis a vis race–ethnic differentials. Once the full set of social background and geographic characteristics have been controlled, the major differences among race–ethnic groups disappear. The only exception pertains to 17 year olds. Although most of the very large race–ethnic differential at age 17 is explained by the other variables in the model, there remain modestly larger odds of age–grade retardation among minorities.

Conclusion

With or without these details, one main finding is strong and clear: during the period from 1972 to 1998, social background, along with geographic location, accounted for almost all of the large race–ethnic differentials in age–grade retardation. Although the odds of falling behind are about twice as great in minority groups as among Whites, the race–ethnic differentials are small after social background and geographic location are controlled. At present, there is little evidence of direct race–ethnic discrimination in progress through the elementary and secondary

grades. However, given the large and ubiquitous race–ethnic differentials in achievement test scores, the recent movement toward high stakes testing for promotion could magnify race–ethnic differentials in retention.☞

(Dropout, continued from p. 4)
is obliged to move to the next lesson plan fall behind, and if they are far behind at year's end, then what? Should these children be moved ahead knowing they're not ready; or should they be held back knowing that most won't be helped enough for them to keep up later? Either way, they are trapped in the same structure and many will simply slip farther back.

The challenge is to build more flexibility into the system without the stigma and other problems that come with being "off-time" for one's age. Most school systems haven't been especially imaginative in addressing the needs of overage students, and some of the more popular approaches risk making matters worse rather than better. So-called alternative schools for overage, pregnant, or parenting students often suffer an "image" problem and, with typically only one or two in the area, there may be logistical problems also. But beyond that, it is asking a great deal of someone shouldering heavy work or parenting responsibilities, as many repeaters do, to commit to the traditional school schedule, and even then, he or she still will be in the company of a student body preoccupied with the traditional concerns of adolescence—hardly a congenial fit. Current arrangements segregate and marginalize these youth. To break

(Dropout, continued on p. 21)

Race Differences in Ability Group Effects on Achievement

Moving beyond the Myths

Maureen Hallinan, University of Notre Dame

Most middle and secondary schools group students by ability. Ability grouping is believed to be an efficient and effective way to instruct a large population of students. The method's efficiency stems from the fact that it provides a fairly straightforward basis on which to assign students to classes. It is effective because it permits teachers to gear instruction to the ability level of their students and to utilize pedagogical techniques appropriate to the students' level of understanding.

Despite its pedagogical advantages, ability grouping has many critics. A major criticism is based on the belief that students learn less in low-ability groups than in higher ability groups. Critics claim that ability grouping channels unequal learning opportunities to students. Empirical research provides support for this belief by showing that students assigned to higher ability groups make greater gains in achievement than those assigned to lower ability groups, controlling for student ability.

A further criticism of ability grouping is based on a fear that the practice discriminates against minority students. Empirical data show that minority students are disproportionately assigned to low-ability groups that have fewer learning opportunities. Thus, the practice of ability grouping is believed to be discriminatory.

Given the somewhat widespread concern that racial or ethnic biases influence the

assignment of students to ability groups, several empirical studies have examined this assignment process. These studies identify several factors that influence ability group placement, including standardized test scores, grades, previous course history, teacher and counselor recommendations, parental choice, and student choice. In addition, organizational factors—such as the availability of teachers, the size of classrooms, the master schedule, and school resources—affect placement decisions. Virtually no evidence has been found that race or ethnicity affects ability group placement. Indeed, some studies imply the opposite: that at the elementary- and middle-school levels, principals and teachers tend to expand the size of higher ability groups to ensure racial and ethnic diversity in these groups.

Despite the failure of critics to find evidence of racial or ethnic bias in the assignment of students to ability groups, the belief that ability grouping disadvantages minority students continues to influence the debate about the equity of this practice. In an effort to shed new light on this debate, this paper takes a different approach. The research examines whether race affects the amount of change in a student's achievement when the student is moved to a higher or lower ability group.

Research Approaches

Previous research examining race differences in ability group

effects on learning investigated two aspects of the practice: the assignment of students to ability groups, and achievement differences in student achievement across ability groups. The approach to the study of race and ability grouping taken in this paper differs from these previous strategies. Based on the author's previous finding that virtually all students make achievement gains when moved to a higher ability group, the analysis examines whether Black and White students make equivalent gains in a higher group. The analysis will determine whether Black and White students respond in a similar manner to a more demanding academic environment. If the results reveal race differences in achievement gains in higher ability groups, then the study might suggest that ability grouping disadvantages minority students in a previously unrecognized way.

Methodology and Sample

The empirical analysis uses survey data obtained from students in five secondary schools in an urban school district in the Midwest. Two cohorts of students were followed from ninth through eleventh grade. Background information on these students, as well as their previous test scores, grades, and ability group assignments from eighth grade, were obtained. Since Asian, Hispanic, Native American, and other non-White, non-Black

students comprised less than five percent of the sample, these non-Black students are classified as White for the analysis. Twenty-seven percent of the sample was Black and 23% qualified for free lunch, a measure of low socioeconomic status. Special education students, those for whom English was a second language, and students who were not taking English or Mathematics in ninth grade were excluded from the sample.

VARIABLES

The dependent variable for the analysis is a student's percentile scores on a standardized test in English and Mathematics. The test was administered annually, on a statewide basis, to third-, sixth-, eighth-, ninth-, and eleventh-grade students in the school district. A national sample was employed as the reference group. The students in the sample took the test in the spring of eighth and ninth grades.

Independent variables for these analyses include measures of student background: gender, race, age (minus an appropriate integer for that year), number of days absent first semester, and free lunch status. The exogenous variables included eighth-grade ability group, eighth-grade grades, and eighth-grade standardized test scores in English and Mathematics.

The middle schools in the district had three ability group levels in eighth-grade English: Basic, Regular, and Honors; and four levels in eighth-grade Mathematics: Basic, Regular, Honors, and Advanced. In ninth

grade, an Advanced English group and a Very Basic Mathematics group were added. Ability group membership for eighth and ninth grades were obtained from school records. Student percentile test scores in eighth-grade in English and Mathematics also were obtained from school records, as were student grades in these subjects. Grades were converted into the usual four-point scale with A = 4.0, B = 3.0, C = 2.0, D = 1.0, and F = 0.0.

Results

The results show a strong race difference in the change in English and Mathematics achievement when a student is moved to a higher or lower ability group. White students show greater gains in achievement when moved to a higher group and greater losses in achievement when moved to a lower group than Blacks. These results suggest that Black students do not benefit as much as White students when they are placed in a more challenging learning environment.

Three possible causes for the race differences in achievement gains and losses have been suggested: learning differences, differences in teacher expectations and peer influences, and ability differences. With respect to the first explanation, the analyses show no significant race differences in how Black and White students learn when assigned to the same ability group. The second explanation cannot be tested directly. However, teacher expectations and peer influences influence student motivation,

which is likely correlated with eighth-grade test scores and grades which are controlled for in the analysis. Since the results show no race difference in the effects of previous achievement factors, the data suggest that teacher and peer influences do not account for race differences in ninth-grade predicted achievement gains.

The third explanation is that it is not race per se that accounts for the differential effect of movement to a higher or lower ability group, but rather ability. The analysis reveals that the pattern of gains and losses accruing from assignment to higher or lower ability groups is similar whether the students are divided by race or by prior ability and achievement. However, the patterns are more pronounced and consistent for the ability comparison. These results indicate that prior ability, not race, explains the differences in achievement gains and losses.

Conclusions

The criticism that Blacks and other minorities have been discriminated against in the assignment of students to ability groups has been a major factor driving the effort to detrack schools. Educators and community members have noted that minority students are disproportionately assigned to the lower ability groups compared to their White peers. If the distribution of Black achievement is indistinguishable from that of Whites, then the fairness of assigning relatively more Blacks than Whites to lower ability groups is called into question. (**Race**, continued on p. 22)

Classroom Organization and Instructional Quality

An Examination of Tracking and De-tracking

Adam Gamoran, University of Wisconsin-Madison

When teachers divide students into separate classes or groups on the basis of prior performance, they do so because they think students are best served by receiving instruction targeted to their particular levels of accomplishment up to that point. Consider the case of first-grade reading: Some children enter school without knowing the alphabet, others are familiar with the letter sounds, and still others are already strong readers. To accommodate these differences, teachers typically divide students into reading groups. Another example would be ninth-grade mathematics: Some students enter high school with eighth-grade algebra under their belts, others have yet to master arithmetic, and many others are in between. In response, high schools commonly divide students for ninth-grade mathematics into general math, pre-algebra, algebra, and geometry classes. To most teachers, these divisions make sense. They make it possible to think about instruction as organized in a clear sequence, to find each student's place in the sequence using criteria they consider objective, and to provide instruction intended to move each student along the instructional hierarchy. In short, ability grouping seems like a neutral device for matching instruction to students' needs.

Despite this sensible logic, there are three reasons why ability grouping cannot be viewed as neutral. First, it leads

to divisions that go beyond academic differences. Because of inequalities outside of schools, which are perpetuated as students move up the ladder of grade levels, when teachers divide students on the basis of academic performance, they tend to separate students who differ from one another by race, ethnicity, and social class. Students from disadvantaged backgrounds tend to score lower on tests for a host of reasons, many of which are unrelated to schooling, and therefore the division of students on the basis of academic performance results in social as well as academic segregation. Second, when teachers create classes that are relatively homogeneous in student performance, they eliminate much of the diversity that might foster rich and productive conversations in classrooms. Although grouping students by performance level may make it possible to sharpen the delivery of instruction to meet students' levels of skills and knowledge, that sharpening may be double-edged, as it eliminates the very differences that some teachers build upon in their instruction. Third, although teachers may intend to provide instruction of equal quality at all levels, in practice that rarely occurs. Instead, compared to their peers in higher-ranked classes, students in lower-ranked classes and groups encounter instruction of lower quality. Consequently, instead of helping low-achieving students catch up,

ability grouping tends to result in widening achievement gaps over time.

Although the problems of grouping and tracking can clearly be identified, eliminating these practices is not easy, because detracking is also associated with problems of instructional quality, and successful de-tracking is rare. Consequently, after we examine the challenges to instruction associated with dividing students by performance level, it will be equally important to consider the instructional challenges associated with mixed-ability grouping.

Tracking and Instruction

Quantitative studies, ranging from elementary to high school, support the contention that instructional differences across groups and tracks contribute to achievement differences. At the high school level, students in college-preparatory programs enroll in more academic courses, and particularly more advanced courses in mathematics and science, and this contributes to their achievement advantages in those subjects. In elementary school, students in higher-ranked reading groups cover more new words and read more stories over the course of a school year, making the reading gap between high- and low-ranked groups wider at the end of the year than it was at the beginning of the year. Two students who start the school year at similar reading levels, but who are assigned to different reading groups, end up

with different reading achievement at the end of the year depending on whether they were assigned to a higher or a lower group. Thus, although providing different instruction to different groups seems like it would help low-achieving students catch up, usually that is not what happens; instead they fall further behind.

CAN TRACKING HELP INSTEAD OF HARM LOW ACHIEVERS?

Not all uses of grouping and tracking are damaging to the prospects of low-achieving students. Catholic high schools, for example, produce less inequality between tracks than do public high schools. This occurs because Catholic schools place more academic demands on students who are not enrolled in the college-preparatory program than do public high schools. Case studies also suggest that a school climate of effort and caring common in Catholic schools enhances teacher and student motivation in both low- and high-track classes. Of course, Catholic schools have the advantage of being able to select their students, and low-achieving students who attend Catholic schools may be more responsive to academic demands than those who attend public schools.

Public school programs in New York and California that aimed to improve the quality of high school mathematics instruction for low-achieving, low-income youth also achieved some success. Students were still sorted into separate classes, but teachers provided instruction that bridged the gap between elementary and college-

preparatory mathematics. More rigorous course content contributed to achievement benefits over the general math classes that were being replaced.

Although it is not what usually happens, it may be possible to group students in a way that promotes equity instead of inequity. Maintaining high standards seems a key to success; providing a rigorous curriculum, communicating expectations, teaching with passion, and avoiding a system in which less-experienced teachers are relegated to lower-level classes all played important roles in these rare success stories.

Detracking and Instruction

Even the most successful uses of grouping still encounter the problem of separating students of different social backgrounds. Social and economic inequalities outside schools contribute to substantial differences in test scores inside school. The result when educators divide students by achievement level is classes that differ by social background. For this reason, many educators would prefer to avoid the practice altogether rather than trying to use grouping more effectively.

However, detracking offers its own set of challenges. While tracking often results in poor instruction for low achievers, it also tends to sustain high-quality instruction for high achievers. Thus, efforts to detrack seem to confront the classic tension between excellence and equity. Can this tension be surmounted? Ultimately, it may come down to a question of values: Is it worth

sacrificing some opportunities for the highest achievers for the sake of more equitable opportunities for all students?

Research can still contribute much to resolving this dilemma by showing what the tradeoffs are and what it might take to provide equitable opportunities without sacrificing high standards of excellence. Can the same high-quality instruction that is now typical in high-track classes be provided in mixed-ability classes?

Resistance to detracking among teachers seems strongest in subjects such as mathematics and foreign language where they perceive the curriculum as rigidly sequential, so that students must master one topic before they proceed to the next. It is difficult to know whether these perceptions are inherent in the subject matter, or if they reflect ingrained beliefs that might be successfully challenged if teachers could be shown that high-quality instruction in mixed-ability classes is possible.

Conclusions

Both practitioners and researchers can respond to research findings about classroom organization and instructional quality. The first reaction from practitioners may be to strengthen their inclinations to reduce grouping and tracking, because these practices are associated with unequal classroom instruction and unequal achievement. The research is clear that some forms of tracking should be eliminated because better alternatives are available. For example, the
(**Classroom**, continued on p. 23)

Tracking, De-tracking, and Skill Grouping

Conclusions from Experimental, Ethnographic, and Regression Studies

James Kulik, University of Michigan

Reformers used to encourage school systems to develop comprehensive schools with an academic track for college-bound students, a vocational track for students headed for jobs, and a general track for students with less definite goals. But views are different today. Many reformers view tracking as one of education's major problems, and they advocate the complete de-tracking of American schools. This paper presents a summary of research findings relevant to the de-tracking debate. Three types of studies have been reviewed: experimental studies, ethnographic studies, and regression analyses.

Experimental Studies

Experimental studies are currently the only dependable guide to the effects of grouping on children. They show that effects depend on both the type of student and the type of grouping that is involved. Different types of programs have different effects on different students. For example, higher aptitude students benefit academically from ability grouping. The academic benefits are positive but usually small when the grouping is done as a part of a broader program for students of all abilities. For example, in XYZ classes where students are divided by ability but taught with the same materials and according to the same curriculum, the test scores of higher ability students are

raised by about 0.1 standard deviations. Within-class and cross-grade programs, which entail curricular adjustment, boost test scores of higher aptitude students by about 0.2 to 0.3 standard deviations.

Benefits on higher aptitude students are usually largest in special accelerated and enriched classes. The largest gains are usually associated with acceleration. Classes in which talented children cover four grades in three years, for example, usually boost achievement levels a good deal. Test scores of children accelerated in this fashion are about one year higher on a grade-equivalent scale than they would be if the children were not accelerated. Enriched classes, in which students have a varied educational experience, boost student achievement by more moderate amounts. The average gain on a grade-equivalent scale is 4 months in a typical program.

Grouping programs usually have smaller effects on middle and lower aptitude learners. XYZ classes, for example, have virtually no effect on the achievement of such students. Test scores of middle and lower aptitude students learning in XYZ classes are indistinguishable from those of similar students in mixed-ability classes. Cross-grade and within-class programs, however, usually raise test scores of middle and lower aptitude pupils by between 0.2 and 0.3 standard deviations. The clear adjustment of curriculum to

pupil ability in within-class and cross-grade programs may be the key to their effectiveness.

Experimental studies fail to support the charge that students in the lower tracks suffer irreparable damage to their self-esteem. Students in the high groups drop a little in self-esteem; the self-esteem of students in low groups actually increases in ability-grouped classes. The finding is inconsistent with the labeling or stigma theory, which predicts a drop in self-esteem for the lower-status groups. It is consistent, however, with predictions of the social comparison theory, which states that people make self-evaluations by comparing themselves to those around them. The theory predicts that slow learners will feel more adequate in a slow-learning group and that fast learners will feel less special in a fast-learning group.

Ethnographic Studies

Although some ethnographic studies include quantitative data, most provide only qualitative observations. Ethnographers try to uncover the subjective meaning of events and patterns of life in schools through rational analysis of their observations.

Ethnographers have reached four main conclusions:

- Instruction is conceptually simplified and proceeds more slowly in lower tracks.
- More experienced teachers seem to be disproportionately assigned to the higher tracks.

- Teachers' views of high-track students are more positive than their views of low-track students.
- Most of a student's friends are found in the same track.

While ethnographers have reported that the curriculum is debased, teachers are inexperienced, and instruction is poor in lower track classes, careful scrutiny of the ethnographic evidence provides little support for such interpretations. When ethnographers have quantified their observations, for example, differences between instruction in upper and lower track classes actually appear to be small. What is more important, the interpretation of the differences is unclear. The reported differences between upper and lower track classes may simply indicate that teachers try to adjust the pace of their instruction to the preparation of their students.

Thus, the true yield from these ethnographic studies is rather slim. They show that the amounts of time on-task are different in upper- and lower-track classrooms, and they also suggest that there are differences in teachers, in teacher reactions to students, and in instructional emphasis. But ethnographic studies do not show what lies behind these differences. Differences in instruction for fast and slower students may be appropriate adjustments, or they may reflect real differences in instructional quality in different curricular tracks.

Regression Studies

Regression analyses show that the achievement gap

between students in upper and lower tracks is due mostly to student self-selection. A second, less important factor that may contribute to the achievement gap is the different number of advanced courses in core subjects taken by students in collegiate and noncollegiate tracks. A third factor may be the difference in the way that the same courses are taught for collegiate and noncollegiate students. Regression analyses do not provide conclusive evidence on the second and third factors, however. The controls for self-selection in these analyses are not adequate, and so conclusions from regression analyses are tentative at best.

The main goal for most regression analysts of tracking data has been to determine whether academic ability or socioeconomic status plays a more important role in track placement. Although estimates of the importance of ability, socioeconomic status, and other influences on track placement differ somewhat from one study to the next, the pattern of results is fairly consistent. Four points emerge from the studies reviewed:

- Personal preference seems to be the most important determinant of curricular track.
- After personal preference, the strongest determinant of curriculum placement is academic ability.
- Social class plays a less significant role in high school curriculum placement, except insofar as it influences test scores. All studies indicate that the direct effect of

ability is larger than is the direct effect of socioeconomic class, but the importance of socioeconomic status varies across studies.

- Race and gender play a smaller role in track placement. Blacks have a higher probability of ending up in the college preparatory track than do Whites of equivalent aptitude and socioeconomic status.

Reviewers have criticized the studies that produced these findings on methodological grounds. Most of the studies, for example, compared achievement of students in academic and nonacademic programs. The observed differences in aptitude of students in these programs are so great that attempts to equate the group statistically may be futile.

Although it is true that regression methods can produce misleading results when academic and nonacademic students are being compared, they are more reliable when vocational and general students are being compared because vocational and general students are similar in important characteristics that influence school outcomes. Comparisons of these groups have seldom received much attention in research on curricular tracking, however.

De-tracking

The effects of de-tracking would vary according to the type of grouping program that was eliminated. If typical XYZ classes were eliminated from all schools, the achievement level of
(**Tracking**, continued on p. 23)

Understanding Research on the Consequences of Retention

An Overview of the Research

Lorrie Shepard, University of Colorado at Boulder

Current no-social-promotion policies reflect an urgent desire to improve the quality of education in America. Policymakers are dismayed by examples of students' low performance such as fast-food restaurant servers who can't make change, the poor showing of U.S. seniors in the Third International Math and Science Survey, and complaints from business leaders about the inadequate skills of entry-level workers—and they attribute this poor performance to low standards and the willingness of educators to promote students to the next grade whether or not they have mastered requisite skills.

Like the majority of educators and lay citizens, policymakers are convinced that by ending social promotion they can improve student learning. But what policymakers may not realize is that the lowest scoring students were most likely retained, some more than once.

It is important to understand that retention rates and social promotion rates are only loosely coupled. Therefore, it is possible for both rates to be quite high. In fact, if as comparative studies show retention does not improve achievement, it is likely that many retained students will subsequently become social promotion "statistics" in the years following retention.

Retention and Social Promotion Rates

Retention statistics are not collected nationally but can be

inferred from census data which shows that grade retention is prevalent in American schools. For the most recent cohort of 15–17 year olds, for example, 36% were either below the modal grade level or had left school.

Data documenting high rates of retention are surprising. If significant numbers of poor performing students are being retained, why does the rhetoric of social promotion imply that so many low achievers are being passed through the system? The key to understanding this apparent contradiction is to understand the difference between *annual* and *cumulative* percentages. The percentage cited above is a cumulative rate, meaning that it accumulates for a given group of students across all of their K–12 years of schooling. By the time a cohort of same-age students has reached high school age, 36% are no longer in the appropriate grade. Some have dropped out without having been retained and some started school late, but most (somewhere between 20–25%) repeated at least one grade.

In contrast, annual retention rates are the percentage of students retained in a grade in a given year out of the total number of students in that grade. Annual rates vary from an average of 1.4 students retained in grade in Indiana to 9.8% of students retained each year in Mississippi. By following the same group of students from the

time they enter school, it is possible to see how small annual retention rates add up to a large cumulative rate.

Given the serious consequences of grade retention for dropping out of school, discussed next, many school systems have policies against double retentions or at least require that a second retention not occur within the same level of schooling—i.e., within the primary, intermediate, or middle school grades. Following this same reasoning, it is the cumulative rate that is most relevant for policy analysis, because it reflects the proportion of students affected by retention at sometime during their school career. The annual data, however, help to illustrate why there can be a high percentage of social promotion *decisions* at the same time that such a high proportion of *students* are retained. In principle, promotion–retention decisions are made 13 times in a student's career (14 times) if they repeat a grade. Thus, it is very likely that a retained student will also be a social promotion statistic in years following retention unless retention ensures a permanent improvement in achievement.

Grade Retention and School Dropout

A large number of studies have documented the link between retention and dropping out. Given that a third variable, poor academic performance, predicts both retention and dropping out, the most rigorous

studies use statistical adjustments to control for prior achievement and attendance as well as background variables such as socioeconomic status, sex, and ethnicity.

Across all studies, in both the retention and dropout literatures, retention has been shown to substantially increase the rate of school leaving beyond what would be expected based on poor achievement alone. Direct effects of grade retention on increased dropout rates have also been documented in evaluation studies of earlier no-social-promotion interventions such as the New York City Gates Program and Chicago's crackdown of the 1980s.

The conclusion that repeating a grade increases a student's risk of dropping out is not controversial among researchers. However, the implications of these findings for educational policy depend on the viewpoint of policymakers and the underlying purpose of no-social-promotion policies. If grade retention is intended as an educational intervention aimed primarily at raising the achievement of retained students, then dropping out must be considered a serious side effect of such medicine. However, some policymakers appear to believe that an increased dropout rate is a necessary (or at least a tolerable) corollary of raising standards.

Effects of Grade Retention on Achievement

Major reviews of research on retention have consistently concluded that there is no reliable body of evidence

showing that grade retention is more beneficial than grade promotion for students with serious academic or adjustment difficulties. For example, in 47 studies that measured academic outcomes, repeating a grade had a negative effect on the achievement of retained students compared to promoted students matched initially for equally low achievement.

Recent, large-scale studies have seemingly found some positive effects from retention, although there is still no consistent evidence that retention improves achievement.

- In Baltimore, dramatic gains in the repeat year itself disappeared when students went on to new material.
- In Chicago, the no social promotion policy improved achievement for at-risk students who were threatened by retention but were not retained. For retained students, the finding was the same as in previous studies. There was no improvement either in achievement or an acceleration of school leaving.
- In Texas, much touted gains for retained students may or may not be valid once researchers make adjustments for differences in starting points for the retained and control groups and for generally rising TAAS scores across years.
- In a national sample, the view of retention effects as positive or negative depended on the comparison used, with retained students appearing to have closed the achievement gap when

compared to students in their new grade (not in comparison to students in their original cohort).

Models for Evaluating and Weighing Evidence on Retention

Grade retention is intended to cure (or at least to improve) poor achievement. If retention were evaluated by the Federal Drug Administration, would it be judged to be a safe and effective treatment? The FDA approval process asks two questions:

- Do the results of well-controlled studies provide substantial evidence of effectiveness?
- Do the results show the product is safe—which means that the benefits of the drug appear to outweigh its risks?

By FDA standards, grade retention would not be approved for use. At best, controlled studies show that retention does not harm achievement. But, retention has not been proven effective in improving achievement in subsequent grades. And, it has serious negative side effects, namely students' poorer attitudes toward school and substantially increased risk for dropping out of school.

Given their track record, other treatments for poor achievement have a greater chance of success. After-school programs, tutoring, summer school, and one-on-one reading instruction are more effective in raising achievement than repeating a grade, as shown by large positive results in their research literatures. ☘

(Dropout, continued from p. 12) down these barriers requires somehow relaxing the overly tight link between “age” and “grade.” Doing so would likely improve the graduation prospects of children who are a year or two behind, and it certainly would give educators more options for addressing their needs. Under this accounting, the problem isn’t so much grade retention as it is the structure within which grade retention is embedded, a structure that makes deviants of otherwise perfectly normal children. ❧

(Retention, continued from p. 6) attempt is made to address the underlying conditions such as low motivation, poverty, poor nutrition, or inadequate instruction that cause underachievement. It would be surprising if retention or limited retention-plus policies substantially altered children’s achievement. Underachieving children require educational experiences that affect their rates of early learning.

Contrast this reactive approach to intervention with prevention. Instead of waiting until the early signs of academic failure are evident, proactive education support would seek to promote the skills and attitudes needed for mastery of the grade-level curriculum before learning difficulties are observed. Prevention programs do this by addressing the underlying causes that give rise to underachievement such as building language and literacy skills before formal reading instruction, instilling pride in achievement, enhancing motivation to learn, and promoting family–school partnerships to help reinforce learning at home. Not surprisingly, programs that succeed in these areas are

associated with higher levels of school achievement and lower rates of grade retention.

The importance of prevention is easily lost in an era of school accountability and high-stakes testing, which highlight children’s learning difficulties. Given the consistent evidence that retention is not an effective strategy for improving children’s school success and growing evidence that retention plus remediation strategies do little to enhance children’s achievement, the alternatives to retention appear to deserve much higher funding priority than they currently receive. Among these alternatives are universal access to high-quality preschool education, full-day kindergarten programs, reduced class sizes in the early grades, family–school partnerships that provide family resource centers in schools, and school restructuring programs. Investments in preschool education have shown among the most positive long-term effects on the school success of children at risk. One of the most consistent findings in the 35 years of research is that participation in preschool programs for low-income, at-risk children reduces the need for grade retention in the elementary grades. Only increased funding for such programs can help break the cycle of school failure that many children face. ❧

(Analysis, continued from p. 8) not to retain?” as we enter the new millennium. Available evidence suggests that neither social promotion nor grade retention will solve our nation’s educational ills nor facilitate the academic success of children.

Instead, attention must be directed toward empirically supported prevention and remedial programs. It is suggested that educational professionals, scholars, and politicians commit to implementing and investigating specific prevention and remedial intervention strategies designed to facilitate educational achievement and socio-emotional adjustment of children at risk of school failure. It is time to move beyond the rhetoric regarding retention and social promotion; we should seriously consider the results of empirical research examining the efficacy of grade retention. Educational research provides valuable insight regarding the effectiveness of various prevention and academic intervention programs, these studies warrant further consideration as we attempt to enhance the educational outcomes of at-risk children. Considering the results of research from the past decade, grade retention fails to demonstrate effectiveness and would not be considered to be an empirically supported intervention. ❧

(Unlike, continued from p. 10) classified and treated very differently in special education placements. By the late 1970s, the misidentification problem triggered both federal and state court decisions requiring that potentially disabled students receive due process. In a far-reaching decision, the California courts ruled in *Larry P. v. Wilson Riles* (1979) that schools could no longer use intelligence **(Unlike**, continued on p. 22)

(Unlike, continued from p. 21) tests to identify minority students as mentally retarded. However, substantial problems remain and new ones emerge, including recent evidence that African American boys are disproportionately identified as having Attention Deficit Hyperactivity Disorder (ADHD).

Self-fulfilling Prophecies

Placement in a low class becomes a self-fulfilling prophecy of low expectations, fewer opportunities, and poor academic performance. Poor performance begins the cycle anew, giving additional justification to schools to reduce expectations and opportunities. Extensive research makes clear that, in every aspect of what makes for a quality education, kids in lower tracks typically get less than those in higher tracks and gifted programs.

Finally, grouping practices help shape students' identities, status, and expectations for themselves. Both students and adults mistake labels such as "gifted," "honor student," "average," "remedial," "learning disabled," and "mild mental retardation" for certification of overall ability or worth. Everyone without the "gifted" label has the *de facto* label of "not gifted." The resource classroom is a low-status place and students who go there are low-status students. The result of all this is that most students have needlessly low self-concepts and schools have low expectations. Few students or teachers can defy those identities and expectations.

We Have Much to Do

Since the late 1980s, policymakers, educators, and advocacy groups have responded to problems with homogenous grouping by recommending that schools dismantle it. These recommendations reflect growing support for heterogeneous grouping as necessary to ensure that all students have access to high-quality curriculum, teachers, and learning experiences. For example, early analyses of the disappointing performance of U.S. students on the Third International Mathematics and Science Study (TIMSS) support mounting concerns that the low scores stem, in part, from the tracking of most American students in less academically demanding math and science classes. Increasingly, educators and policymakers are developing an awareness that schools cannot teach or achieve social justice unless they eliminate grouping practices. A number of school desegregation cases have cited the practice as a source of continuing racial discrimination. However, this goal will not be accomplished quickly, and policy reports will simply gather dust unless enlightened educators understand and act to change the norms and political relations these grouping practices embody. There is a long, hard road ahead. ☞

In the next
CEIC REVIEW . . .

*Social-Emotional Learning and
School Success*

(Race, continued from p. 14)

However, empirical evidence shows that Blacks lag behind Whites in achievement. When ability or prior achievement is controlled, the effect of being Black on ability group assignment virtually disappears. These results remove one of the major criticisms of ability grouping. Nevertheless, discrimination against Black students is still possible within ability groups. The instructional processes and learning conditions that characterize ability groups might be more supportive of White students than of Blacks. Several studies show that, within ability groups of any level, Black students generally score lower than White students.

The analyses presented in this paper shed further light on whether ability grouping discriminates against Black students. Based on previous research showing that all students benefit from assignment in a higher ability group, the study examined whether the advantages of higher group placement benefit Black and White students equally. The analysis showed that, on average, White students make greater gains in achievement than Black students when assigned to a higher ability group. This finding could be interpreted as a contextual effect on Black learning. If conditions in higher ability groups were more conducive to White achievement gains than Black gains, then Blacks would be disadvantaged relative to Whites even in the more demanding learning environments.

To address this issue, the achievement gains of Black and

White students at both ends of the ability distribution in each ability group were examined. The results show that students in the lower quarter of any ability distribution achieved less growth in achievement than their peers in the top three fourths of the distribution. The pattern of these results was more pronounced than when Black and White achievement growth were compared. Both Black and White students are found in the lower and upper parts of the ability distribution in each ability group, but Blacks are somewhat more likely than Whites to be found in the lower part of the distribution. Thus, the results indicate that it is ability, not race, which is governing the differences in achievement growth.

While arguments can be made to detrack schools, racial bias should not be among them. Of greater validity is the argument that the practice of ability grouping, as it is presently practiced, discriminates against slow learners. These students, regardless of race, are offered fewer learning opportunities in lower ability groups. If the educational resources and positive learning climate evident in most higher ability groups were recreated in the lower groups, a major concern about the equity of ability grouping would disappear. ❧

(Classroom, continued from p. 16) research points towards elimination of general math classes in high school. These courses simply repeat the arithmetic curriculum of elementary and junior high school and block access to college-bound curricula and college enrollment.

At the elementary level, research also indicates that the practice of rigid tracking of students in elementary school for the entire school day on the basis of a sole criterion should cease.

Practitioners can consider whether to maintain less extreme versions of tracking, or to eliminate all divisions among students. The research is inconclusive as to which alternative is better, and practitioners must consider their own unique circumstances in deciding which approach best fits their school. One view holds that, as has occurred in some Catholic schools and in some restructured public schools, some divisions among students for particular subjects are appropriate as long as teachers hold students at all levels to high standards of accomplishment. Another view holds that all such divisions should be eliminated. To adopt this approach, it will be necessary to develop curricula and pedagogies that are suited to mixed-ability classes.

The lack of conclusive evidence on alternatives to traditional tracking structures also shows where researchers must direct their attention. It is most essential to examine a broader range of schools engaged in responding to the tracking problem, so that we can move beyond “existence proofs” to a more generalizable conclusion about the advantages and disadvantages of each policy choice. ❧

(Tracking, continued from p. 18) the country’s brightest students might fall by a trivial amount, but the effects would not be noticeable on most other

students. If the grouping programs that were eliminated were ones that actually adjusted methods and materials to student aptitude, the damage to student achievement would be greater, and the effects would be felt more broadly. Both higher and lower aptitude students would suffer academically from such de-tracking. But the damage would be truly profound if, in the name of de-tracking, schools eliminated enriched and accelerated classes for their brightest learners. The achievement level of such students would fall dramatically if they were required to move at the common pace. No one can be certain that there would be a way to repair the harm that would be done. ❧

The CEIC
REVIEW
Lydia Hoag
Editor

This publication is supported in part by the Laboratory for Student Success at Temple University Center for Research in Human Development and Education (CRHDE).

The opinions expressed do not necessarily reflect the position of the supporting agencies, and no official endorsement should be inferred.

For more information, write CEIC, Temple University Center for Research in Human Development and Education, 1301 Cecil B. Moore Avenue, Philadelphia, PA 19122-6091 or call (215) 204-3000.

Selected CEIC publications and abstracts of articles, including back issues of *The CEIC Review*, are available at <http://www.temple.edu/LSS/ceicrevlist.htm>