DOCUMENT RESUME

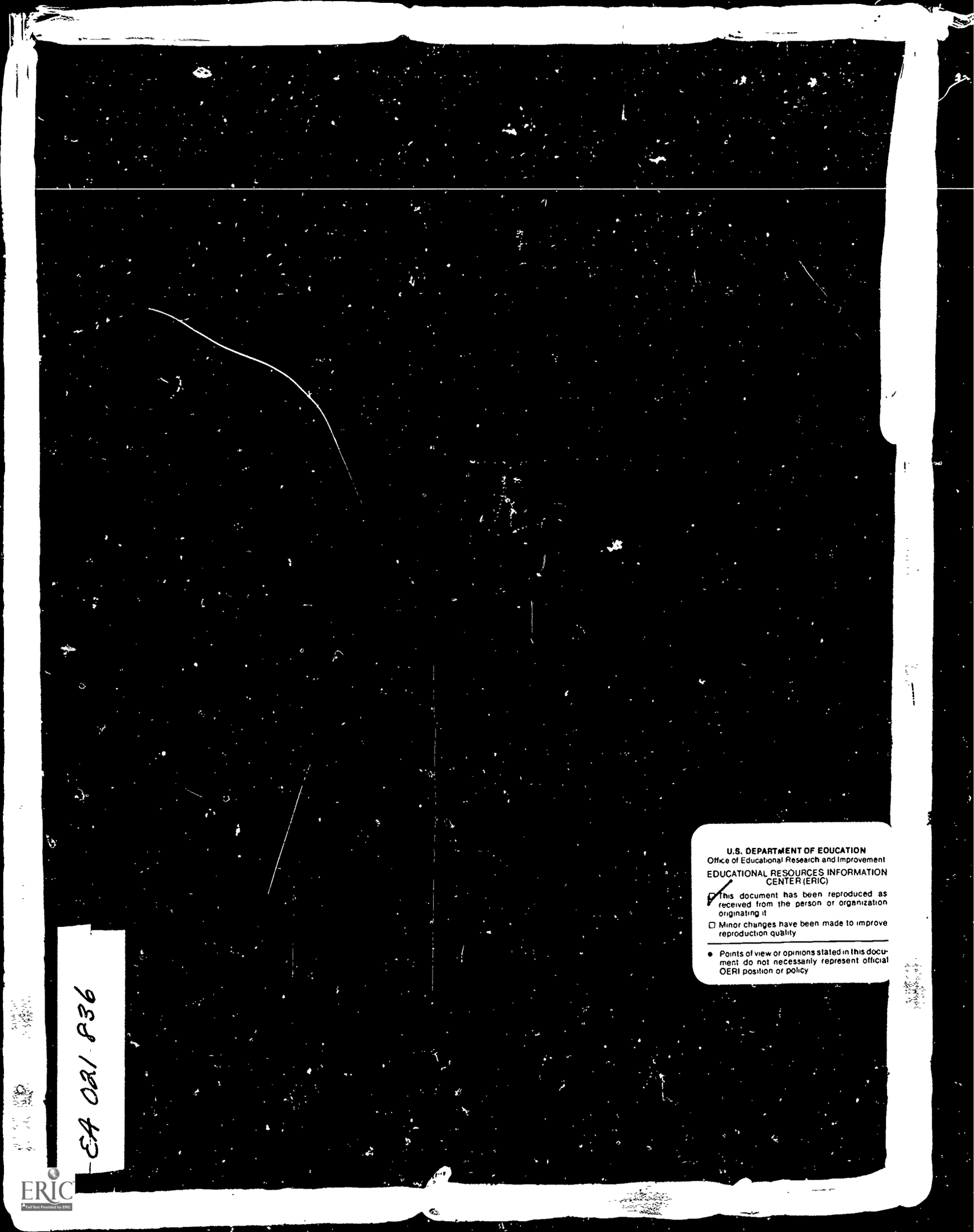
ED 322 565 EA 021 836

AUTHOR Slavin, Robert E.
TITLE Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis.
INSTITUTION National Center on Effective Secondary Schools, Madison, WI.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 90
CONTRACT G-008690007
NOTE 37p.
AVAILABLE FROM Document Service, National Center on Effective Secondary Schools, University of Wisconsin-Madison, Wisconsin Center for Education Research, 1025 W. Johnson Street, Madison, WI 53706 ($7.00 prepaid).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability Grouping; *Academic Achievement; *Effect Size; Literature Reviews; Meta Analysis; Secondary Education; *Secondary School Students; *Student Evaluation; Student Placement; *Track System (Education)

ABSTRACT

A "best evidence" review synthesis, which incorporates features of meta-analytic and traditional literature reviews, is used in this review of studies on the effects of ability grouping on secondary school students' achievement. The focus was on 29 studies that compared between-class ability grouping to heterogeneous placements. Effect sizes were used to characterize study results. Findings indicate that comprehensive between-class ability grouping plans, different forms of ability grouping, and ability grouping by subject (except in social studies) had no effect on student achievement. The finding of zero effects of grouping for all ability levels contradicts earlier conclusions that demonstrated benefits of ability grouping for high-level students and detriments for low-level students. Explanations for this discrepancy are discussed. An implication is that policy decisions about ability grouping must be based on criteria other than effect on academic achievement. A recommendation is made for reduction of between class ability grouping practices and consideration of cooperative learning methods. An extensive bibliography and statistical tables are included. (LMI)



ACHIEVEMENT EFFECTS OF ABILITY GROUPING IN SECONDARY SCHOOLS: A BEST-EVIDENCE SYNTHESIS

Robert E. Slavin
The Johns Hopkins University

This paper was prepared for the National Center on Effective Secondary Schools, supported by the U.S. Department of Education, Office of Educational Research and Improvement (Grant No. G-008690007) and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the supporting agencies. I would like to thank Fred Newmann, Jeannie Oakes, Adam Gamoran, Harry Passow, and Alan Kerckhoff for their comments on an earlier draft.

ACHIEVEMENT EFFECTS OF ABILITY GROUPING IN SECONDARY SCHOOLS: A BEST-EVIDENCE SYNTHESIS

EXECUTIVE SUMMARY

This article reviews research on the effects of ability grouping on the achievement of secondary school students. The principal focus of the review was on studies which compared between-class ability grouping to heterogeneous placements. Six randomized experiments, 9 matched experiments, and 14 correlational studies made this comparison. Across twenty studies from which effect sizes would be computed, the median effect size was essentially zero (ES = -.02), and no differences were found in the remaining nine studies. Effect sizes were also near zero for high achievers (ES = +.01), average achievers (ES = -.08), and low achievers (ES = -.02). While most studies involved grades 7-9, senior high school studies did not produce results different from those involving junior high schools. Effects were similar in all subjects except social studies, where heterogeneous placement was usually superior to ability grouping. A very small set of studies of forms of grouping other than typical between-class plans (e.g., within-class grouping, flexible grouping, Joplin Plan) failed to find positive effects of these methods.

The finding of zero effects of grouping for all ability levels contradicts earlier findings from studies comparing students in high, average, and low ability groups which had suggested that ability grouping was beneficial to students in high groups and detrimental to those in low groups. Several explanations are advanced to account for this discrepancy.

The report concludes with a recommendation that, in the absence of any evidence of instructional effectiveness, secondary schools should reduce their use of between-class ability grouping.

ACHIEVEMENT EFFECTS OF ABILITY GROUPING IN SECONDARY SCHOOLS: A BEST-EVIDENCE SYNTHESIS

For more than seventy years, ability grouping has been one of the most controversial issues in education. Its effects, particularly on student achievement, have been extensively studied over that time period, and many reviews of the literature have been written. In recent years, a comprehensive review of the achievement effects of ability grouping in elementary schools was published by Slavin (1987), but only brief meta-analyses by Kulik and Kulik (1982, 1987) have reviewed the evidence on ability grouping and heterogeneous placement in secondary schools.

The purpose of this paper is to present a comprehensive review of all research published in English which evaluated the effects of ability grouping on student achievement in secondary schools. "Secondary schools" are defined here as middle, junior, or senior high schools in the U.S., or similarly configured secondary schools in other countries. Secondary schools can include grades as low as five, but they usually begin with sixth or seventh grades. Ability grouping is defined as any school or classroom organization plan which is intended to reduce the heterogeneity of instructional groups; between-class ability grouping reduces the heterogeneity of each class for a given subject and within-class ability grouping reduces the heterogeneity of groups within the class (e.g., reading groups).

Unlike the situation in elementary schools, ability grouping in secondary schools is overwhelmingly between-class grouping (McPartland, Coldiron, & Braddock, 1987). Several closely related forms of ability grouping are used. Sometimes students are assigned to a track within which all courses are taken, based on some combination of composite achievement, IQ, and teacher judgments. For example, senior high school students are often assigned to academic, general, and vocational tracks; middle/ junior high school students are often assigned to advanced, basic, and remedial tracks (in either case, the number of tracks and the names used to describe them vary widely). This type of grouping plan is generally called tracking in the U.S. or streaming in Europe. It is an example of what Slavin (1987) called "ability-grouped class assignment." In addition to assignment to higher and lower sections of the same courses, tracking in senior high schools usually also involves different courses or course requirements. For example, a student in the academic track may have to take more years of mathematics than a student in the general track, or may take French III rather than metal shop.

A particular form of tracking often seen in middle/junior high schools is block scheduling, where students spend all or most of the day with one homogeneous group of students. Some schools rank-order students from top to bottom and assign them to, say, 7-1, 7-2, 7-3, and so on. Many senior high schools allow students to choose their track or to choose the level they wish to take in each subject, but in plans of this kind counselors tend to steer students into the level of classes to which they would have been assigned if the school were not allowing students a choice (Rosenbaum, 1978).

Another form of ability grouping common in secondary schools involves assigning students to ability-grouped classes for all academic subjects, but allows for the possibility that students will be placed in a high-ranking group for one subject and a low-ranking group for another. In practice, scheduling constraints often make this type of grouping similar to plans in which all courses are

taken within the same track. In some cases schools group by ability for some subjects and not for others; for example, students may be in ability-grouped math and English classes but in heterogeneous social studies and science classes. Ability grouping usually involves higher and lower sections of the same course, but sometimes consists of assignment to completely different courses, as when ninth graders are assigned either to Algebra I or to general math. When high achievers are assigned to markedly different courses usually offered to older students (as when seventh graders take algebra), this is called acceleration. More commonly, high achievers may be assigned to "honors" or "advanced placement" sections of a given course, while low achievers may be assigned to special "remedial" sections.

While between-class ability grouping is by far the most common type of ability grouping in secondary schools, forms of within-class grouping are also occasionally seen. These are plans in which students are assigned to homogeneous instructional groups within their classes. Within-class ability grouping, such as use of reading or math groups, is the most common form of grouping at the elementary level (McPartland et al., 1987). Complex plans, such as plans that involve grouping across grade lines, flexible grouping for particular topics, and part-time grouping, are also occasionally seen in secondary schools. In general, a wider range of grouping plans are used in middle/ junior high schools than in senior high schools.

Arguments for and against ability grouping have been essentially similar for seventy years. For example, Turney (1931), summarizing writings of the 1920s, listed the following advantages and disadvantages:

Advantages (according to Turney, 1931)

1. It permits pupils to make progress commensurate with their abilities.
2. It makes possible an adaption of the technique of instruction to the needs of the group.
3. It reduces failures.
4. It helps to maintain interest and incentive, because bright students are not bored by the participation of the dull.
5. Slower pupils participate more when not eclipsed by those much brighter.
6. It makes teaching easier.
7. It makes possible individual instruction to small slow groups.

Disadvantages (According to Turney, 1931).

1. Slow pupils need the presence of the able students to stimulate them and encourage them.
2. A stigma is attached to low sections, operating to discourage the pupils in these sections.

3. Teachers are unable, or do not have time, to differentiate the work for different levels of ability.

4. Teachers object to the slower groups.

A research symposium, school board meeting, or PTA meeting on the topic of ability grouping in 1990 is likely to bring up much the same arguments on both sides, with two important additions: the argument that ability grouping discriminates against minority and lower-class students (e.g., Braddock, 1989; Rosenbaum, 1976), and the argument that the low tracks receive a lower pace and lower quality of instruction than do students in the higher tracks (e.g, Gamoran, 1989; Oakes, 1985).

In essence, the argument in favor of ability grouping is that grouping will allow teachers to adapt instruction to the needs of a diverse student body, with an opportunity to provide more difficult material to high achievers and more support to low achievers. For high achievers, the challenge and stimulation of other high achievers is felt to be beneficial (see Feldhusen, 1989). Arguments opposed to ability grouping focus primarily on the perceived damage to low achievers, who experience a slower pace and lower quality of instruction; teachers who are less experienced or able and who do not want to teach low-track classes; low expectations for performance; and few positive behavioral models (e.g., Gamoran, 1989; Oakes, 1985; Persell, 1977; Rosenbaum, 1980). Because of the demoralization, low expectations, and poor behavioral models, students in the low tracks are felt to be more prone to delinquency, absenteeism, dropping out, and other social problems (Crespo & Michelna, 1981; Wiatrowski, Hansell, Massey, & Wilson, 1982). With few college-bound peers, students in the low tracks are found to be less likely to attend college than other students (Gamoran, 1987). Ability grouping is perceived to perpetuate social class and racial inequities because lower-class and minority students are disproportionally represented in the lower tracks. Ability grouping is often considered to be a major factor in the development of elite and underclass groups in society (Persell, 1977; Rosenbaum, 1980). Perhaps most importantly, tracking is felt to work against egalitarian, democratic ideals by sorting students into categories from which escape is difficult or impossible.

There are important differences between the pro-grouping and anti-grouping positions that go beyond the arguments themselves. Arguments in favor of ability grouping focus on *effectiveness*, saying in effect that as distasteful as grouping may be, it so enhances the learning of students (particularly but not only high achievers) that its use is necessary. In contrast, arguments opposed to grouping focus at least as much on *equity* as on effectiveness, on democratic values as much as on outcomes. In one sense, then, the burden of proof is on those who favor grouping, for if grouping is not found to be clearly more effective than heterogeneous placement, none of the pro-grouping arguments apply. The same is not true of anti-grouping arguments, which provide a rationale for abolishing grouping that would be plausible even if grouping were found to have no adverse effect on achievement.

Research on the achievement effects of ability grouping has taken two broad forms. One type of research compares the achievement gains of students who were in one or another form of grouping to those of students in ungrouped, heterogeneous placements. Another type of research compares the achievement gains made by students in high-ability groups to those made by students in the low groups.

Reviews of the grouping vs. non-grouping literature have consistently found ability grouping to have little or no impact on student achievement overall in elementary and secondary schools (e.g., Borg, 1965; Esposito, 1973; Findley & Bryan, 1971; Good & Marshall, 1984; Heathers, 1969; Kulik & Kulik, 1982). Based primarily on his own empirical research, Borg (1965) claimed that ability grouping had a slight positive effect on the achievement of high achievers and a slight negative effect on low achievers, but Kulik and Kulik (1987) found no such trend.

In contrast, researchers who have compared gains made by students in different tracks have generally concluded that when ability level, socioeconomic status, and other factors are controlled, high-track assignment accelerates achievement while low-track assignment significantly reduces achievement (Alexander, Cook, & McDill, 1978; Dar & Resh, 1986; Gamoran & Berends, 1987; Gamoran & Mare, 1989; Oakes, 1982; Persell, 1977; Sorensen & Hallinan, 1986). In fact, many researchers and theorists in the sociological tradition maintain that tracking is a principal engine of social inequality in society and that it causes or greatly magnifies differences along lines of class and ethnicity (e.g., Braddock, 1990; Jones, Erickson, & Crowell, 1972; Schafer & Olexa, 1971; Vanfossen, Jones, & Spade, 1987).

One area of research has investigated the quality of instruction offered to students in high- and low-ability groups, usually concluding that low-ability group classes receive a quality of instruction that is significantly lower than that received by students in high-track classes (e.g., Evertson, 1982; Gamoran, 1989; Oakes, 1985; Trimble & Sinclair, 1987). However, it is difficult to compare "quality of instruction" in high- and low-track classes. For example, teachers typically cover less material in a low-track class (e.g., Oakes, 1985). Is this an indication of poor quality of instruction or an appropriate pace of instruction? Students in low-track classes are more off-task than those in high-track classes (e.g., Evertson, 1982). Is this due to the poor behavioral models and low expectations in the low-track classes, or would low achievers be more off-task than high achievers in any grouping arrangement? However, evidence that low-track classes are often taught by less experienced or less qualified teachers or that they manifest other indicators of lower-quality instruction could justify the conclusion that regardless of measurable effects on learning, students in the lower tracks do not receive equal treatment.

In addition to synthesizing research on overall effects of ability grouping on the achievement of high-average- and low- achieving secondary students, this review will attempt to reconcile research comparing achievement gains in different tracks with research comparing grouped and ungrouped settings.

Review Methods

This article uses a review procedure called "best-evidence synthesis" (Slavin, 1986), which incorporates the best features of meta-analytic and traditional reviews. Best-evidence syntheses specify clear, well-justified methodological and substantive criteria for inclusion of studies in the main review and describe individual studies and critical research issues in the depth typical of good-quality narrative reviews. However, whenever possible, effect sizes are used to characterize study outcomes, as in meta-analyses (Glass, McGaw, & Smith, 1981). Systematic literature search procedures, also characteristic of meta-analysis, are similarly applied in best-evidence syntheses.

Criteria for Study Inclusion

The studies on which this review is based had to meet a set of a priori criteria with respect to relevance to the topic and methodological adequacy. First, all studies had to involve comprehensive ability grouping plans, which incorporated most or all students in the school. This excludes studies of special programs for the gifted or other high achievers as well as studies of special education, remedial programs, or other special programs for low achievers. Studies of within-class ability grouping are included, but studies of such grouping-related programs as individualized instruction, mastery learning, cooperative learning, and continuous-progress groupings are excluded.

Studies had to be available in English, but otherwise no restrictions were placed on study location or year of publication. Every attempt was made to locate dissertations and other unpublished documents in addition to the published literature.

Methodological requirements for inclusion. Criteria for inclusion of studies in the main review were essentially identical to those used in an earlier review of elementary ability grouping (Slavin, 1987). These were as follows:

1. Ability-grouped classes were compared to heterogeneously-grouped classes. This requirement excluded a few studies that correlated "degree of heterogeneity" with achievement gain (e.g., Millman & Johnson, 1964; Wilcox, 1963). Studies that compared achievement gains for students in different tracks (e.g., Alexander, Cook, & McDill, 1978) were excluded from the main review but are discussed in a separate section.

2. Achievement data from standardized or teacher-made tests were presented. This excluded many anecdotal reports and studies which used grades as the dependent measure. Teacher-made tests, used in a very small number of studies, were accepted only if there was evidence that they were designed to assess objectives taught in all classes.

3. Initial comparability of samples was established by use of random assignment or matching of students or classes. When individual students in intact schools or classes were matched, evidence had to be presented that the intact groups were comparable.

4. Ability grouping had to be in place for at least a semester.

5. At least three ability-grouped and three control classes were involved.

The criteria outlined above excluded very few studies comparing comprehensive ability grouping plans to heterogeneous placements. Every study located which satisfied criteria 1, 2, and 3 also satisfied criteria 4 and 5. Excluding studies of special programs for high achievers (e.g., Atkinson & O'Connor, 1963), all but two of the studies included in meta-analyses by Kulik and Kulik (1982, 1987) were also included in the present review. The exception was a study by Adamson (1971) which had substantial IQ differences favoring the ability-grouped school, and one by Wilcox (1963) which compared more and less heterogeneously tracked classes.

One major category of studies included in the present review but excluded by the Kuliks are those which did not present data from which effect scores could be computed (e.g., Borg, 1965; Ferri,

1971; Lovell, 1960; Postlethwaite & Denton, 1978). These studies are discussed in terms of the direction and statistical significance of their findings.

Literature Search Procedures

The studies included in this review were located in an extensive search. Principal sources included the Education Resources Information Center (ERIC), Dissertation Abstracts, and citations made in other reviews, meta-analyses, and primary sources. Every attempt was made to obtain a complete set of published and unpublished studies that met the criteria outline above.

Computation of Effect Sizes

Effect sizes were generally computed as the difference between the experimental and control means divided by the control group's standard deviation (Glass et al., 1981). In the ability grouping literature, the heterogeneous group is almost always considered the control group, and this convention is followed in the present article; positive effect sizes are ones that favored ability grouping, while negative effect sizes indicated higher means in the heterogeneous groups. The standard deviation of the heterogeneous group is also preferred as the denominator because of the possibility that ability grouping may alter the distribution of scores. However, when means or standard deviations were omitted in studies that otherwise met the inclusion criteria, effect sizes were estimated when possible from t's, F's, exact p values, sums of squares in factorial designs, or other information, following procedures described by Glass et al. (1981).

Several of the studies included in this review presented data comparing gain scores without reporting actual pre- or posttest means. Standard deviations of gain scores are typically lower than those of raw scores (to the degree that pre-post correlations exceed + 0.5), so effect sizes computed on gain scores are often inflated. If pre-post correlations are known, effect sizes from all scores can be transformed to the scale of posttest values. However, because none of the studies using gain scores also provided pre-post correlations, a pre-post correlation of + 0.8 was assumed (following Slavin, 1987). Using a formula from Glass et al. (1981), this correlation produces a multiplier of 0.632, which was used to deflate effect size estimates from gain score data. The purpose of this and other procedures was to attempt to put all effect size estimates in the same metric, the unadjusted standard deviation of the heterogeneous classes. However, because this multiplier is only a rough approximation, effect sizes from studies using gain scores should be interpreted with even more caution than that which is warranted for effect sizes in general.

Another deviation from usual meta-analytic procedure used in the present review involved adjustments of posttest scores for any pretest differences. This was done either by subtracting pretest means from posttests (if the same tests were used), by converting pre- and posttest means to z-scores and then subtracting (if different tests were used), or by using covariance-adjusted scores. However, even when such adjustments were made affecting the numerator of the effect size formula, the denominator remained the unadjusted posttest standard deviation.

One effect size is reported for each study (see Bangert-Drowns, 1986). When multiple subsamples, subjects, or tests were used, medians were computed across the data points. For example, if four measures were used with three subgroups (e.g., high, middle, and low achievers), the effect size for the study as a whole would be the median of the twelve (4 x 3) resulting effect sizes. Whenever

possible, findings were also broken down by achievement level (high, average, low), and separate effect sizes were also computed for each major subject.

In pooling findings across studies, medians rather than means were used, principally to avoid giving too much weight to outliers. However, any measure of central tendency in a meta- analysis or best-evidence synthesis should be interpreted in light of the quality and consistency of the studies from which it was derived, not as a finding in its own right.

Research on Ability Grouping in Secondary Schools

A total of 29 studies of tracking or streaming in secondary schools met the inclusion criteria listed earlier. The studies, their major characteristics, and their findings are listed in Table 1.

Table 1 Here

The studies listed in Table 1 are organized in three categories according to their research designs. Six studies used random assignment of students to ability-grouped or heterogeneous classes. Nine studies took groups of students, matched them individually on IQ, composite achievement, and other measures, and then assigned one of each matched pair of students to an ability-grouped class, one to a heterogeneous class. The quality of these randomized or matched experimental designs is very high, and the findings of the 15 studies using such designs must be given special weight. The remaining 14 studies investigated existing schools or classrooms which used or did not use ability grouping, and then either selected matched groups of students from within each type of school or used analyses of covariance or other statistical procedures to equate the groups. The difficulty inherent in such designs is that any differences between schools that are systematically related to ability grouping would be confounded with the practice of ability grouping per se. For example, a secondary school that used heterogeneous grouping might have a staff, principal, or community more concerned about equity, affective development, or other goals than would a "matched" school that used ability grouping. However, several of the correlational studies used very large samples and longitudinal designs, and these provide important additional information not obtainable from the typically smaller and shorter experimental studies.

Within each category studies are listed in descending order of sample size. All other things being equal, therefore, studies near the top of Table 1 should be considered as better evidence of the effects of ability-grouping than studies near the end of the Table. However, the nature and quality of the studies are discussed in more detail in the following sections.

Overall Findings

Across the 29 studies listed in Table 1, the effects of ability grouping on student achievement are essentially zero.

The median effect size for the 20 studies from which effect sizes could be estimated was -.02, and none of the nine additional studies found statistically significant effects. Counting the studies with nonsignificant differences as though they had effect sizes of .00, the median effect size for all 29 studies would be .00. Results from the 15 randomized and matched experimental studies were not much different; the median effect size was -.06 for the 13 studies from which effect sizes could be estimated. In nine of these thirteen studies (including all five of the randomized studies) results favored the heterogeneous groups, but these effects are mostly very small.

There are few consistent patterns in the study findings. Most of the studies involved grades 7-9, with ninth graders sometimes in junior high schools and sometimes in senior high schools. No apparent trend is discernable within this range. Above the ninth grade the evidence is too sparse for firm conclusions. Lovell (1960) found that high achievers performed significantly better in ability-grouped English classes, but there were no effects in biology or algebra and no effects for average or low achievers. In a four-year study of students in grades 9-12, Borg (1965) found significant positive effects of ability grouping for average and low achievers in math, but found no differences in science or for high achievers. Cohorts followed from grades 7-10 and 8-11 showed no significant differences on any measure for any ability level. On the other hand, Thompson (1974), in a study of eleventh-grade social studies, found the largest effects favoring heterogeneous grouping of all studies located (ES = -.48), while Kline (1964), in another four-year study of students in grades 9-12, found no differences.

Twelve of the 29 studies tracked students for all subjects according to one composite ability or achievement measure. The remaining seventeen studies grouped on the basis of performance in one or more specific subjects. However, there were no differences in the outcomes of these different forms of ability grouping. In addition, there were no consistent patterns in terms of the number of ability groups to which students were assigned (the great majority of studies used three). Study duration had no apparent impact on outcome. Studies which used adjusted gain scores produced the same effects as other studies, and the use of the adjustment of gain scores described above made no difference in outcomes.

There was no discernible pattern of findings with respect to different subjects, with one possible exception. Studies by Marascuilo and McSweeney (1972), Thompson (1974), and Fowlkes(1931) found relatively strong effects favoring heterogeneous grouping in social studies, and three additional studies by Peterson (1966), Martin (1927), and Postlethwaite and Denton (1978) found no differences or slight effects in the same direction. This is not enough evidence to conclusively point to a positive effect of heterogeneous grouping in social studies, but it is important to note that all three of the randomized or matched experimental studies found differences in this direction.

There were no consistent effects according to study location. All four of the British studies found no differences between streamed and unstreamed classes. A large, longitudinal Swedish study by Svensson (1962), not shown in Table 1 because it lacked adequate evidence of initial equality, also found no differences between streamed and unstreamed classes. Urban, suburban, and rural schools had similar outcomes. The one study which involved large numbers of minority students, a randomized experiment in a New York City high school by Ford (1974), found no differences between ability-grouped and heterogeneous math classes.

Studies conducted before 1950 were no more likely than more recent studies to find achievement differences. On this topic, it is interesting to note that experimental-control studies of ability grouping have not been done in recent years. The only study of the 1980s, by Kerckhoff (1986), was done by a sociologist who focused his attention on differences between students in different streams. This study is described in more detail later on. Otherwise, the most recent experimental-control comparisons were done in the early 1970s.

Differential Effects According to Achievement Levels

One of the most important questions about ability grouping in secondary schools concerns the degree to which it affects students at different achievement levels differently. As noted earlier, many researchers and reviewers, particularly those working in the sociological tradition, have emphasized the *relative* impact of grouping for different groups of students far more than the average effect for all students.

Twenty-one of the 29 studies presented in Table 1 presented data on the effects of ability grouping on students of different ability levels. Most studies broke their samples into three categories (high, average, and low achievers), but some used two or four categories.

Across the 15 studies from which effect sizes could be computed, the median effect size was +.01 for high achievers, -.08 for average achievers, and -.02 for low achievers. Effects of this size are indistinguishable from zero, and if all the nonsignificant differences found in studies from which effect sizes could not be computed are counted as effect sizes of .00, the median effect size for each level of student becomes .00. In addition, only one of seven studies from which effect sizes could not be computed (Lovell, 1960) found significantly positive effects of ability grouping for high achievers, and none of these studies found significant effects in either direction for average and low achievers. The randomized and matched experimental studies provided slightly more support for the idea that ability grouping has a differential effect; the median effects sizes for high, average, and low achievers were +.05, -.10, and -.06, respectively. It is interesting to note that the study by Borg (1965), which is often cited to support the differential effect of ability grouping on students of different ability levels, in fact provides very weak support for this phenomenon. Across two measures given to members of four-year cohorts which principally included secondary years, significant effects favoring ability grouping were found for high achievers in one out of eight comparisons, for average achievers in three out of eight, and for low achievers in one out of eight. Only in a cohort that went from grades 4 to 7 were there significant effects favoring heterogeneous grouping for low achievers.

It might be expected that differential effects of track placement would build over time, and that longitudinal studies would show more of a differential impact than one-year studies. The one multi-year randomized study, by Marascuilo and McSweeney (1972), did find that over a two-year period, students in the top social studies classes gained slightly more than similar students in heterogeneous classes (ES=+.14), while middle (ES=-.37), and low (ES=-.43) groups gained significantly less than their ungrouped counterparts. However, across seven multi-year correlational studies of up to five years' duration, not one found a clear pattern of differential effects.

A few studies provided additional information on differential impacts of ability grouping by investigating effects of grouping on high or low achievers only. For example, Torgelson (1963)

randomly assigned low-achieving students in grades 7-9 to homogeneous or heterogeneous classes. Across several performance measures, the median effect size was +.13 (non-significantly favoring ability grouping). Similarly, Borg and Prpich (1966) randomly assigned low-achieving tenth graders to ability-grouped or heterogeneous English classes, and found that there were no differences in one cohort. In a second cohort, differences favoring ability grouping on a writing measure were found, but there were no differences on eight other measures.

Studies of ability grouping of high achievers are difficult to distinguish from studies of special programs for the gifted. Well-designed studies of programs for the gifted generally find few effects of separate programs for high achievers unless the programs include acceleration (exposure to material usually taught at a higher grade level) (Fox, 1979; Kulik & Kulik, 1984). That is, grouping per se has little effect on the achievement of high achievers. An outstanding study that illustrates this is a dissertation by Mikkelson (1962), who randomly assigned high-achieving seventh and eighth graders to ability-grouped or heterogeneous math classes. The seventh-grade homogeneous classes were given enrichment, but the eighth graders were accelerated, skipping to ninth-grade algebra. No effects were found for the seventh graders. The accelerated eighth graders of course did substantially better than similar students who were not accelerated on an algebra test, and they did no worse on a test of eighth grade math.

Taken together, research comparing ability-grouped to heterogeneous placements provides little support for the proposition that high achievers gain from grouping while low-achievers lose. However, there is an important limitation to this conclusion. In most of the studies which compared tracked to untracked grouping plans (including all of the randomized and matched experimental studies), tracked students took different levels of the same courses (e.g., high, average, or low sections of Algebra 1). Yet much of the practical impact of tracking, particularly at the senior high school level, is on determining the nature and number of courses taken in a given area. The experimental studies do not compare students in Algebra 1 to those in Math 9, or students who take four years of math to those who take two. The conclusions drawn in this section are limited, therefore, to the effects of between-class grouping *within the same courses,* and should not be read as indicating a lack of differential effects of tracking as it affects course selection and course requirements.

Other Forms of Ability Grouping

The studies discussed above and summarized in Table 1 evaluated the most common forms of ability grouping in secondary schools; full-time, between-class ability grouping for one or more subjects. However, a few studies have evaluated other grouping plans.

The most widely used form of grouping in elementary schools, within-class ability grouping, has also been evaluated in a few studies involving middle and junior high schools. Campbell (1965) compared the use of three math groups within the class to heterogeneous assignment in two Kansas City junior high schools. There were no differences between the two programs in achievement. Harrah (1956) compared five types of within-class grouping in grades 7-9 in West Virginia, and found ability grouping to be no more successful than other grouping methods. Note that these findings conflict with those of studies of within-class ability grouping in mathematics in the upper elementary grades, which tended to support the use of math groups (Slavin, 1987).

Vakos (1969) evaluated the use of a combination of heterogeneous and homogeneous instruction in eleventh-grade social studies classes in Minneapolis. Students were grouped by ability two days each week, but heterogeneously grouped the other three days. No achievement differences were found. Zweibelson, Bahnmuller, & Lyman (1965) evaluated a similar mixed approach to teaching ninth-grade social studies in New Rochelle, New York, and also found no achievement differences. Chiotti (1961) compared a flexible plan for grouping junior high school students across grade lines for mathematics to both ability-grouped and heterogeneous grouping plans, and again found no differences in achievement. A cross-grade grouping arrangement similar to the Joplin Plan (Slavin, 1987) was compared to within-class grouping in reading by Chismar (1971) in grades 4-8. Significantly positive effects of this program were found in grades 4 and 7 but not 5,6, and 8.

Reconciling Track/No Track and High Track/Low Track Studies

As noted earlier in this review, two very different traditions of research have dominated research on ability grouping. One involves comparisons of ability-grouped to heterogeneous placements. The other involves comparisons of the progress made by students in different ability groups or tracks. While there has been little experimental research comparing ability-grouped to heterogeneous placements since the early 1970s, research comparing the achievement of students in different tracks largely began in the 1970s and continues to the present.

The findings of high track/low track studies of ability grouping conflict with those emphasized in this review, in that they generally find that even after controlling for IQ, socioeconomic status, pretests, and other measures, students in high tracks gain significantly more in achievement than do students in low tracks, especially in mathematics (see Gamoran & Berends, 1987, for a review). How can these findings be reconciled with those of the experimental studies?

One important difference between experimental and correlational studies of ability grouping is that, as mentioned earlier, correlational studies (especially at the senior high school level) often include not only the effects of being in a high, average, or low class, but also the effects of differential course-taking. Students in academic tracks may score better than those in general or vocational classes because they take more courses or more advanced courses. The experimental studies comparing grouped and ungrouped classes are all studies of grouping per se, holding course-taking and other factors constant, while the correlational studies examine tracking as it is in practice, where track placement implies differences in course requirements, course-taking patterns, and so on. Also, experimental track vs. no track studies are rare beyond the ninth grade, while most correlational studies comparing students in high vs. low tracks involve senior high schools. The lack of track vs. no track studies at the senior high school level is hardly surprising given the nearly universal use of some form of tracking at that level. However, tracking usually has a different meaning in senior than in junior high school. While junior high school tracking mostly involves different levels of courses (e.g. high English vs. low English), senior high tracking is more likely to involve completely different patterns of cousework (e.g., metal shop vs. French III). Also, the problem of dropouts becomes serious in senior high school; a study of twelfth graders unavoidably excludes the students who may have suffered most from being in the low track and left school (see Gamoran, 1987). This could reduce observed differences between high- and low-track students.

There is limited evidence, however, that differences in course-taking or grade level account for the different conclusions of the track/no track and high track/low track studies. Four-year longitudinal studies in U.S. senior high schools by Kline (1964) and Borg (1965) found no differential effects of track placement for high, average, and low achievers (as compared to similar students in untracked placements). Presumably, course-taking patterns in these senior high school studies varied by track. A correlational study by Alexander and Cook (1982) found that while taking more courses in senior high school did increase achievement (net of background factors), different course-taking patterns in different tracks did not account for track differences in achievement. Gamoran (1987) found that track effects on math and science achievement were explained in part by the fact that students in the academic tracks take more math and science courses and, in particular, more *advanced* courses in these areas. However, no such patterns were seen on reading, vocabulary, writing, or civics achievement measures. Gamoran notes the difficulty of disentangling track and course-taking, which are highly correlated in math and science (and of course both track and course-taking are strongly correlated with ability, socioeconomic status, and other factors). It is certainly logical to expect correlational studies of senior high school tracking to find different effects of different track placements because of different course-taking patterns, but because of confounding of tracking, course-taking, and student background factors, this is difficult to determine conclusively.

Another likely explanation for different findings of track/no track and high track/low track studies has to do with the difficulty of statistically controlling for large differences. Students in higher tracks tend to achieve at much higher levels than those in lower tracks (both before and after taking secondary courses), and statistically controlling for these differences is probably not enough to completely remove the influence of ability or prior performance on later achievement. Further, students in higher tracks are also likely to be higher in such attributes as motivation, internal locus of control, academic self-esteem, and effort; factors which are not likely to be controlled in correlational studies, measures of prior ability and achievement.

To understand the difficulty of controlling for large initial differences between students, imagine an experiment in which a new instructional method was to be evaluated. The experimenter selects a group of students who have high test scores and high IQ scores, and are nominated by their teachers as being hard-working, motivated, and college material. This group becomes the experimental group, and the remaining students serve as the control group. To control for the differences between the groups, prior composite achievement and socioeconomic status are used as covariates or control variables.

In such an experiment, no one would doubt that regardless of the true effectiveness of the innovative treatment, the experimental group would score far better than the control group, even controlling for prior achievement and socioeconomic status. No journal, or dissertation committee would accept such a study. Yet this "experiment" is essentially what is being done when researchers compare students in different tracks. When there are significant pretest differences, use of statistical controls through analysis of covariance or regression are considered inadequate to equate the groups. Most often, the statistical controls will undercontrol for true differences (Lord, 1960; Reichardt, 1979). Yet high- and low-track students usually differ in pretests or IQ by one to two standard deviations, an enormous systematic difference for which no statistical procedure can adequately control.

The only study which compared both tracked to untracked schools and high-track to low-track students was a five-year longitudinal study by Kerckhoff (1986) in Britain. This study illustrates the problem of controlling for large differences. For example, in mathematics, boys in the high track of 3-group ability-grouping programs gained about 11 z-score points from a test given at age 11 to one given at age 16, while students in a remedial track gained 18 z-score points. Yet the regression coefficient comparing the high track to ungrouped students was + 2.34, indicating performance about 42% of a standard deviation above "predicted" performance. In contrast, the remedial-track boys had a regression coefficient (in comparison to ungrouped students) of -.72, indicating performance about 13% of a standard deviation below "predicted" performance, despite the fact that the remedial students actually gained more than the top-track students. The reason for this is that the remedial students started out (at age 11) scoring 1.64 standard deviations below the ungrouped students, while top-track students started out 1.02 standard deviations above the ungrouped students, a total difference between top-track and remedial students at 2.66. No regression or analysis of covariance can adequately control for such large pretest differences. Because of unreliability in the measures and less than perfect within-group correlations of pre- and posttests, "predicted" scores based on pretests and other covariates will (other things being equal) be too low for high achievers and too high for low achievers.

Another factor that can contribute to overestimates of the effects of curriculum track on achievement in studies lacking heterogeneous comparison groups is fan spread. Put simply, high achievers usually gain more per year than do low achievers, so over time the gap between high and low achievers grows. This increasing gap cannot be unambiguously ascribed to ability grouping or other school practices, as it occurs under virtually all circumstances. A student who is performing at the 16th percentile in the sixth grade and is still at the 16th percentile in twelfth-grade will be further "behind" the twelfth-grade mean in grade equivalents, for example (Coleman & Karweit, 1972).

An additional factor that can contribute to spurious findings indicating a benefit of being in the high track is that factors other than test scores influence placement decisions. For example, a study by Balow (1964) found that on math tests not used for group placement, there was enormous overlap between students in supposedly homogeneous seventh-grade math classes. More than 72% of the students scored between the lowest score in the top group and the highest score in the bottom group. Among these students in the "area of overlap," students who were in the top group gained the most in math achievement over the course of the year, while those in the low group gained the least.

On its surface this study provides support to the "self- fulfilling prophecy" argument. Yet consider what is going on. Imagine two students with identical scores, one assigned to the high group and one to the low group. Why were they so assigned? Random error is a possibility, but all the systematic possibilities weigh in the direction of higher performance for the student assigned to the high group. Since teacher judgment was involved, teachers may have accurate knowledge to enable them to predict who will do well and who will not. The actual assignments were done on different tests than those used in the Balow study; it is likely that students who scored low on Balow's pretests but were put in the high groups scored high on the test used for placement, and then regressed to a higher mean on Balow's posttest.

What this discussion is meant to convey is not that different tracks do or do not have a differential impact on student achievement, but that comparisons of students in existing tracks cannot tell us one way or another. To learn about the differential impacts of track placement, there are two types of research that might be done. One would be to randomly assign students at the margin to different tracks, something that has never been done. The other is to compare similar students randomly assigned to ability-grouped or ungrouped systems. This has been done several times, and, as noted earlier in this review, there is no clear trend indicating that students in high-track classes learn any more than high-achieving students in heterogeneous classes, or that students in low-track classes learn any less than low-achieving students in heterogeneous classes.

Why is Ability Grouping Ineffective?

The evidence summarized in Table 1 and discussed in this review is generally consistent with the conclusions of earlier reviews comparing homogeneous and heterogeneous grouping (e.g., Kulik & Kulik, 1982, 1987; Noland, 1985), but runs counter to two quite different kinds of "common sense." On one hand, it is surprising to find that assignment to the low ability group is not detrimental to student learning. A substantial literature has indicated the low quality of instruction in low groups (e.g., Evertson, 1982; Gamoran, 1989; Oakes, 1985), and a related body of research has documented the negative impact of ability grouping on the motivations and self-esteems of students assigned to low groups (e.g., Cottle, 1974; Schafer & Olexa, 1971; Trimble & Sinclair, 1987). How can the effect of ability grouping on low-achieving students be zero, as this review concludes?

On the other hand, another kind of "common sense" would argue that, at least in certain subjects, ability grouping is imperative in secondary schools. How could an eighth-grade math teacher teach a class composed of students who are fully ready for algebra and students who are still not firm in subtraction and multiplication? How does an English teacher teach literature and writing to a class in which reading levels range from third to twelfth grade? Yet study after study, including randomized experiments of a quality rarely seen in educational research, finds no positive effect of ability grouping in any subject or at any grade level, even for the high achievers most widely assumed to benefit from grouping.

The present review cannot provide definitive answers to these questions. However, it is worthwhile to speculate on them.

One possibility is that the standardized tests used in virtually all the studies discussed in this review are too insensitive to pick up effects of grouping. This seems particularly plausible in looking at tests of reading, because reading has not generally been taught as such in secondary schools. However, standardized tests of mathematics do have a great deal of face validity and curricular relevance, and these show no more consistent a pattern of outcomes. Marasculio & McSweeney (1972) used both teacher-made and standardized measures of social studies achievement and found similar results with each.

Another possibility is that it simply does not matter whom students sit next to in a secondary class. Secondary teachers use a very narrow range of teaching methods, overwhelmingly using some form of lecture/discussion (Goodlad, 1983). In this setting, the direct impact of students on one another may be minimal. If this is so, then any impacts of ability grouping on students would have to be mediated by teacher characteristics or behaviors or by student perceptions and motivations.

Studies contrasting teaching behaviors in high- and low-track classes usually find that the low tracks have a slower pace of instruction and lower time on-task (e.g, Evertson, 1982; Oakes, 1982). Yet, as noted earlier, the meaning and impact of these differences are not self-evident. It may be that a slower pace of instruction is appropriate with lower-achieving students, or that pace is relatively unimportant because a higher pace with lower mastery is essentially equivalent to a lower pace with higher mastery. Higher time on-task should certainly be related to higher achievement (Brophy & Good, 1986), but the comparisons of time on task between high and low tracks are misleading. What would be important to compare is time on task *for low achievers* in homogeneous and heterogeneous classes, because low achievers may simply be off-task more than high achievers regardless of their class placement. In this regard, it is important to note that Evertson, Sanford, & Emmer (1981) found time on-task to be lower in extremely heterogeneous junior high school classes than in less heterogeneous ones because teachers had difficulty managing the more heterogeneous classes.

The lesson to be drawn from research on ability grouping may be that unless teaching methods are systematically changed, school organization has little impact on student achievement. This conclusion would be consistent with the equally puzzling finding that substantial reductions in class size have little impact on achievement (Slavin, 1989); if teachers continue to use some form of lecture/discussion/seatwork/quiz, then it may matter very little in the aggregate which students the teachers are facing or how many of them there are. In contrast, forms of ability grouping which were found to make a difference in the upper elementary grades, the Joplin Plan (cross-grade grouping in reading to allow for whole-class instruction) and within-class grouping in mathematics (Slavin, 1987) both significantly change time allocations and instructional activities within the classroom.

Alternatives to Ability Grouping

If the effects of ability grouping on student achievement are zero, then there is little reason to maintain the practice. As noted earlier in this article, arguments in favor of ability grouping depend on assumptions about the effectiveness of grouping, at least for high achievers. In the absence of any evidence of effectiveness, these arguments cannot be sustained.

Yet there is also no evidence that simply moving away from traditional ability grouping practices will in itself enhance student achievement, and there are legitimate concerns expressed by teachers and others about the practical difficulties of teaching extremely heterogeneous classes at the secondary level. How can schools moving away from traditional ability grouping use this opportunity to contribute to student achievement?

One alternative to ability grouping often proposed (e.g, Oakes, 1985) is use of cooperative learning methods, which involve students working in small, heterogeneous learning groups. Research on cooperative learning consistently finds positive effects of these methods if they incorporate two major elements: group goals and individual accountability (Slavin, 1990). That is, the cooperating groups must be rewarded or recognized based on the sum or average of individual learning performances. Cooperative learning methods of this kind have been successfully used at all grade levels, but there is less research on them in grades 10-12 than in grades 2-9 (see Newmann & Thompson, 1987). Cooperative learning methods have also had consistently positive impacts on

such outcomes as self-esteem, race relations, acceptance of mainstreamed academically handicapped students, and ability to work cooperatively (Slavin, 1990).

One category of cooperative learning methods may be particularly useful in middle schools moving toward heterogeneous class assignments. These are Cooperative Integrated Reading and Composition (Stevens, Madden, Slavin, & Farnish, 1987) and Team Assisted Individualization - Mathematics (Slavin, Madden, & Leavey, 1984; Slavin & Karweit, 1985). Both of these methods are designed to accommodate a wide range of student performance levels in one classroom, using both homogeneous and heterogeneous within-class groupings. These programs have been successfully researched in grades 3-6, but are often used up to the eighth grade level.

Other alternatives to between-class ability grouping have also been found to be successful in the upper elementary grades (see Slavin, 1987) and could probably be effective in middle schools as well. These include within-class ability grouping in mathematics (e.g., teaching two or three math groups within a heterogeneous class), and the Joplin Plan in reading. The Joplin Plan involves regrouping students for reading across grade levels but according to reading level, so that no within-class reading groups are necessary. However, while these alternatives to between-class grouping are promising because of their success in the upper elementary grades, the few studies of within-class ability grouping at the junior high school level have not found this practice to be effective (Campbell, 1965; Harrah, 1956) and the one middle school study of the Joplin Plan found only inconsistent positive effects (Chismar, 1971).

For descriptions of secondary schools implementing alternatives to traditional ability grouping, see Slavin, Braddock, Hall, & Petza, 1989.

Limitations of This Review

It is important to note several limitations of the present review. Perhaps the most important is that in none of the studies reviewed here were there systematic observations made of teaching and learning. Observational studies and outcome studies have proceeded on parallel tracks; it would be important to be able to relate evidence of outcomes to changes in teacher behaviors or classroom characteristics. Another limitation, mentioned earlier, is that almost all studies reviewed here used standardized tests of unknown relationship to what was actually taught. A third limitation is the age of most of the studies reviewed. It is possible that schools, students, or ability grouping have changed enough since the 1960s or 1970s to make conclusions from these and older studies tenuous.

As noted earlier, the results reported in this review mainly concern the effects of grouping per se, with little regard for the effects of tracking on such factors as course-taking. Effects of tracking on differential course-taking are most important in senior high schools. There is a need for additional research comparing tracked to untracked situations at the senior high school level, particularly research designed to disentangle the effects of tracking from those of differential course-tracking.

In addition, it would add greatly to the understanding of ability grouping in secondary schools to have evaluations or even descriptions of a wider range of alternatives to traditional ability grouping. The few studies of within-class grouping, cross-grade grouping, and flexible grouping plans are not

nearly adequate to explore alternatives. Cooperative learning, often proposed as an alternative to ability grouping, has frequently been found to increase student achievement in ability grouped as well as ungrouped secondary classes (Slavin, 1990; Newmann & Thompson, 1987), but no study has yet compared cooperative learning in heterogeneous classes to traditional instruction in homogeneous ones. Descriptions of creative alternatives to ability grouping currently exist only at the anecdotal level (Slavin, Braddock, Hall, & Petza, 1989).

Conclusions

While there are limitations to the scope of this review and to the studies on which it is based, there are several conclusions that can be advanced with some confidence. These are as follows:

1. Comprehensive between-class ability grouping plans have little or no effect on the achievement of secondary students. This conclusion is most strongly supported in grades 7-9, but the more limited evidence that does exist from studies in grades 10-12 also fails to support any effect of ability grouping.

2. Different forms of ability grouping are equally ineffective.

3. Ability grouping is equally ineffective in all subjects, except that there may be a negative effect of ability grouping in social studies.

4. Assigning students to different levels of the same course has no consistent positive or negative effects on students of high, average, or low ability.

For the narrow but extremely important purpose of determining the impact of ability grouping on standardized achievement measures, the studies reviewed here are exemplary. Six of them randomly assigned individual students to ability-grouped or heterogeneous classes, and nine more individually matched students and then assigned them to one or the other grouping plan. Many of the studies followed students for two or more years. If there were any true effect of ability grouping on student achievement, this set of studies would surely have detected it.

For practitioners, the findings summarized above mean that decisions about whether or not to group by ability must be made on bases other than likely impacts on achievement. Given the antidemocratic, antiegalitarian nature of ability grouping, the burden of proof should be on those who would group rather than those who favor heterogeneous grouping, and in the absence of evidence that grouping is beneficial, it is hard to justify continuation of the practice. The possibility that students in the low groups are at risk for delinquency, dropping out, and other social problems (e.g., Rosenbaum, 1980) should also weigh against the use of ability grouping. Yet schools and districts moving toward heterogeneous grouping have little basis for expecting that abolishing ability grouping will in itself significantly accelerate student achievement unless they also undertake changes in curriculum or instruction likely to improve actual teaching.

There is much research still to be done to understand the effects of ability grouping in secondary schools on student achievement. Studies of grouping at grades 10-12, studies of a broader range of alternatives to grouping, and studies relating observations to outcomes of grouping are areas of particular need. Enough research has been done comparing tracked to heterogeneous classes and

achievement in high, middle, and low tracks, at least up through the ninth grade. It is time to move beyond these simple comparisons to consider more fully how secondary schools can adapt instruction to the needs of a heterogeneous student body.

REFERENCES

Adamson, D.P. (1971). Differentiated multi-track grouping vs. uni-track educational grouping in mathematics. Unpublished doctoral dissertation, Brigham Young University.

Alexander, K.L., Cook, M.A., & McDill, E.L. (1978). Curriculum tracking and educational stratification. American Sociological Review, 43, 47-66.

Alexander, K.L., & Cook, M.A. (1982). Curricula and coursework: A surprise ending to a familiar story. American Sociological Review, 47, 626-640.

Atkinson, J.W., & O'Connor, P. (1963). Effects of ability grouping in schools related to individual differences in achievement - related motivation. Final Report, Cooperative Research Project No. OE-2-10-024, U.S. Dept. of H.E.W.

Bachman, A.M. (1968). Factors related to the achievement of junior high school students in mathematics. Unpublished doctoral dissertation, University of Oregon.

Bailey, H.P. (1968). A study of the effectiveness of ability grouping on success in first-year algebra. Dissertation Abstracts, 28, 3061 A. (University Microfilms No. 68-1249).

Balow, I. (1964). The effects of "homogeneous" grouping in seventh grade arithmetic. Arithmetic Teacher, 12, 186-191.

Bangert-Drowns, R.L. (1986). Review of developments in meta- analytic method. Psychological Bulletin, 99, 388-399.

Barton, D.P. (1964). An evaluation of ability grouping in ninth grade English. Dissertation Abstracts, 25, 1731 (University Microfilms No. 64-9939).

Bicak, L.J. (1962). Achievement in eighth grade science by heterogeneous and homogeneous classes. Unpublished doctoral dissertation, University of Minnesota.

Billett, R.O. (1928). A controlled experiment to determine the advantages of homogeneous grouping. Educational Research Bulletin, 7, 139.

Borg, W. (1965). Ability grouping in the public schools. The Journal of Experimental Education, 34, 1-97.

Borg, W. & Prpich, T. (1966). Grouping of slow learning high school pupils.

Braddock, J.H. (1990). Tracking: Implications for student race-ethnic subgroups (Tech. Rep. No. 1). Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students.

Breidenstine, A.G. (1936). The educational achievement of pupils in differentiated and undifferentiated groups. Journal of Experimental Education, 5, 91-135.

Brophy, J.E., & Good, T.C. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), Handbook of Research on Teaching (3rd ed.) (pp.328-375). New York: Macmillan.

Campbell, A.L. (1965). A comparison of the effectiveness of two methods of class organization for the teaching of arithmetic in junior high school. Dissertation Abstracts International, 26, 813-814. (University Microfilms, No. 65-6726).

Chiotti, J.F. (1961). A progress comparison of ninth grade students in mathematics from three school districts in the State of Washington with varied methods of grouping. Unpublished doctoral dissertation, University of Northern Colorado, Greeley, CO.

Chismar, M.H. (1971). A study of the effectiveness of cross-level grouping of middle school under-achievers for reading instruction. Unpublished doctoral dissertation, Kent State University, Kent, OH.

Cochran, J. (1961). Grouping students in junior high school. Educational Leadership, 18, 414-419.

Coleman, J.S., & Karweit, N.L. (1972). Information systems and performance measures in schools. Englewood Cliffs, NJ: Educational Technology Publications.

Cottle, T.J. (1974). What tracking did to Ollie Taylor. Social Policy, 5, 21-24.

Crespo, M., & Michelna, J. (1981). Streaming, absenteeism, and dropping out. Canadian Journal of Education, 6, 40-55

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. American Educational Research Journal, 23, 357-374.

Drews, E.M. (1963). Student abilities, grouping patterns, and classroom interaction. Cooperative Research Project No. 608, U.S. Department of Health, Education, and Welfare.

Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. Review of Educational Research, 43, 163-179.

Evertson, C.M. (1982). Differences in instructional activities in higher- and lower-achieving junior high English and math classes. Elementary School Journal, 82, 329-350.

Evertson, C., Sanford, J., & Emmer, E., (1981). Effects of class heterogeneity in junior high school. American Educational Research Journal, 18, 219-232.

Feldhusen, J.F. (1989). Synthesis of research on gifted youth. Educational Leadership, 46,(6), 6-11.

Ferri, E. (1971). Streaming: Two years later. London: National Foundation for Educational Research in England and Wales.

Fick, W.W. (1963). The effectiveness of ability grouping in seventh grade core classes. Dissertation Abstracts, 23, 2753.

Findley, W.G., & Bryan, M. (1971). Ability grouping: 1970 Status, impact, and alternatives. Athens, GA: Center for Educational Improvement, University of Georgia. (ERIC Document Reproduction Service No. ED 060 595).

Fogelman, K., Essen, J. & Tibbenham, A. (1978). Ability-grouping in secondary schools and attainment. Educational Studies, 4, 201-212.

Ford, S. (1974). Grouping in mathematics: Effects on achievement and learning environment. Unpublished doctoral dissertation, Yeshiva University, New York, NY.

Fowlkes, J.G. (1931). Homogeneous or heterogeneous grouping -- which? The Nation's Schools, 8, 74-78.

Fox, L.H. (1979). Programs for the gifted and talented: An overview. In A.H. Passow (Ed.), The gifted and talented: Their education and development (pp. 104-126). Chicago: University of Chicago Press.

Gamoran, A. (1987). The stratification of high school learning opportunities. Sociology of Education, 60, 135-155.

Gamoran, A. (1989). Measuring curriculum differentiation. American Journal of Education, 97, 129-143.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. Review of Educational Research, 57, 415-435.

Gamoran, A., & Mare, R.D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? American Journal of Sociology, 94, 1146-1183.

Glass, G., McGaw, B., & Smith, M.L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Good, T., & Marshall, S. (1984). Do students learn more in heterogeneous or homogeneous groups? In P. Peterson, L.C. Wilkinson, & M. Hallinan (Eds.), The social context of instruction: Group organization and group processes (pp. 15-38). New York: Academic Press.

Goodlad, J.I. (1983). A place called school. New York: McGraw-Hill.

Harrah, D.D. (1956). A study of the effectiveness of five kinds of grouping in the classroom. Dissertation Abstracts, 17, 715. (University Microfilms No. 56-1114).

Heathers, G. (1969). Grouping. In R. Ebel (Ed.). Encyclopedia of educational research. (4th ed.). New York: Macmillan.

Holy, T.C., & Sutton, D.H. (1930). Ability grouping in the ninth grade. Educational Research Bulletin, 9, 419-422.

Jones, J.D., Erickson, E.L., & Crowell, R. (1972). Increasing the gap between whites and Blacks: Tracking as a contribution source. Education and Urban Society, 4, 339-349.

Kerckhoff, A.C. (1986). Effects of ability grouping in British secondary schools. American Sociological Review, 51, 842-858.

Kline, R.E. (1964). A longitudinal study of the effectiveness of the track plan in the secondary schools of a metropolitan community. Dissertation Abstracts, 25, 324. (University Microfilms No. 64-4257).

Kulik, C.L., & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. American Educational Research Journal, 19, 415-428.

Kulik, C., & Kulik, J. (1984). Effects of accelerated instruction on students. Review of Educational Research, 54, 409-425.

Kulik, J.A., & Kulik, C.L. (1987). Effects of ability grouping on student achievement. Equity and Excellence, 23, 22-30.

Lord, F.M. (1960). Large-sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 55, 307-321.

Lovell, J.T. (1960). The Bay High School experiment. Educational Leadership, 17, 383-387.

Martin, W.H. (1927). The results of homogeneous grouping in the junior high school. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Martin, W.B. (1959). Effects of ability grouping on junior high school achievement. Dissertation Abstracts, 20, 2810. (University Microfilms No. 59-1108).

Marascuilo, L.A., & McSweeney, M. (1972). Tracking and minority student attitudes and performance. Urban Education, 6, 303-319.

McPartland, J.M., Coldiron, J.R., Braddock, J.H. (1987). School structures and classroom practices in elementary, middle, and secondary schools. Baltimore, MD: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

Mikkelson, J.E. (1962). An experimental study of selective grouping and acceleration in junior high school mathematics. Unpublished doctoral dissertation, University of Minnesota.

Millman, J., & Johnson, M. (1964). Relation of section variance to achievement gains in English and mathematics in grades 7 and 8. American Educational Research Journal, 1, 47-51, 54.

Newbold, D. (1977). Ability-grouping, the Banbury enquiry. Slough, England: National Foundation for Educational Research in England and Wales.

Newmann, F.M., & Thompson, J. (1987). Effects of cooperative learning on achievement in secondary schools: A summary of research. Madison, WI: University of Wisconsin, National Center on Effective Secondary Schools.

Noland, T.K. (1985). The effects of ability grouping: A meta-analysis of research findings. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Oakes, J. (1982). The reproduction of inequity: The content of secondary school tracking. Urban Review, 14, 107-120.

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

Persell, C.H. (1977). Education and inequality: A theoretical and empirical synthesis. New York: Free Press.

Peterson, R.L. (1966). An experimental study of the effects of ability grouping in grades 7 and 8. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Platz, E.F. (1965). The effectiveness of ability grouping in general science classes. Dissertation Abstracts, 26, 1459-1460A. (University Microfilms No. 65-6914).

Postlethwaite, K., & Denton, C. (1978). Streams for the future? Slough, England: National Foundation for Educational Research.

Purdom, T.L. (1929). The value of homogeneous grouping. Baltimore, MD: Warwick & York.

Reichardt, C.S. (1979). The statistical analysis of data from nonequivalent group designs. In T.C. Cook & D.T. Campbell, Quasi-experimentation: Design and analysis issues for field settings (pp. 147-205). Chicago: Rand McNally.

Rosenbaum, J.E. (1976). Making inequality: The hidden curriculum of high school tracking. New York: Wiley.

Rosenbaum, J.E. (1978). The structure of opportunity in school. Social Forces, 57, 236-256.

Rosenblum, J.E. (1980). Social implications of educational grouping. Review of Research in Education, 8, 361-401.

Schafer, W. & Olexa, C. (1971). Tracking and opportunity: The locking-out process and beyond. Scranton, PA: Chandler.

Slavin, R.E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher, 15(9), 5-11.

Slavin, R.E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. Review of Educational Research, 57, 347-350.

Slavin, R.E. (1989). Achievement effects of substantial reductions in class size. In R.E. Slavin (Ed.), School and classroom organization (pp. 247-257). Hillsdale, NJ: Erlbaum.

Slavin, R.E. (1990). Cooperative learning: Theory, research, and practice. Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R.E., Braddock, J.H., Hall, C., & Petza, R.J. (1989). Alternatives to ability grouping. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R.E., Karweit, N.L. (1985). Effects of whole-class, ability grouped, and individualized instruction on mathematics achievement. American Educational Research Journal, 22, 351-367.

Slavin, R.E., Madden, N.A., & Leavey, M.B. (1984). Effects of team assisted individualization on mathematics achievement of academically handicapped and non-handicapped students. Journal of Educational Psychology, 76, 813-819.

Sorensen, A.B., & Hallinan, M.T. (1986). Effects of ability grouping on growth in academic achievement. American Educational Research Journal, 23, 543-557.

Stevens, R.J., Madden, N.A., Slavin, R.E., & Farnish, A.M. (1987). Cooperative integrated reading and composition: Two field experiments. Reading Research Quarterly, 22, 433-454.

Stoakes, D.W. (1964). An educational experiment with the homogeneous grouping of mentally advanced and slow learning students in the junior high school. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Svensson, N.E. (1962). Ability grouping and scholastic achievement. Stockholm: Almqvist & Wirsell.

Thompson, G.W. (1974). The effects of ability grouping upon achievement in eleventh grade American history. Journal of Experimental Education, 42, 76-79.

Torgelson, J.W. (1963). A comparison of homogeneous and heterogeneous grouping for below-average junior high school students. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Trimble, K.D., & Sinclair, R.L. (1987). On the wrong track: Ability grouping and the threat to equity. Equity and Excellence, 23, 15-21.

Turney, A.H. (1931). The status of ability grouping. Educational Administration and Supervision, 17, 21-42, 110-127.

Vakos, H.N. (1969). The effect of part-time grouping on achievement in social studies. Dissertation Abstracts, 30, 2271. (University Microfilms No. 69-20, 066).

Vanfossen, B.E., Jones, J.D., & Spade, J.Z. (1987). Curriculum tracking and status maintenance. Sociology of Education, 60, 104-122.

Wiatrowski, M., Hansell, S., Massey, C.R., & Wilson, D.L. (1982). Curriculum tracking and delinquency. American Sociological Review, 47, 151-160.

Wilcox, J. (1963). A search for the multiple effects of grouping upon the growth and behavior of junior high school pupils. Dissertation Abstracts, 24, 205.

Willcutt, R.E. (1969). Ability grouping by content topics in junior high school mathematics. Journal of Educational Research, 63, 152-156.

Zweibelson, I., Bahnmuller, M., Lyman, L. (1965). Team teaching and flexible grouping in the junior high-school social studies. Journal of Experimental Education, 34, 20-32.

Table 1
Studies of Secondary Tracking

Article	Grades	Location	Sample Size	Duration	Design	Effect Sizes By Achievement		Effect Sizes By Subject		Total
Randomized Experimental Studies										
Marascuilo & McSweeney, 1972	8-9	Berkeley, CA	603 Students	2 yrs.	Students randomly assigned to 3-group AG or hetero social studies classes. Compared students on teacher-made and standardized tests.	HI AV LO	+.14 -.37 -.43	Social Studies	-.22	-.22
Drews, 1963	9	Lansing, MI	4 Schools 432 Students	1 yr.	Students randomly assigned to 3-group AG or hetero English classes compared on standardized tests.	HI AV LO	-.16 +.01 -.01	Reading Language	-.11 .00	-.05
Fick, 1963	7	Olathe, KS	1 School 168 Students	1 yr.	Students randomly assigned to 3-group AG or hetero "core" classes. Both classes taught by same teacher. Iowa Tests of Basic Skills used as posttests.	HI AV LO	+.01 .00 -.04	Reading Language	-.01 .00	-.01
Peterson, 1966	7-8	Chisholm, MN	1 School 152 Students	1 yr.	Students randomly assigned to AG or hetero classes. AG based on composite ach., grades. Compared on standardized tests.	HI AV LO	+.05 -.44 -.06	Reading Language Math Social Studies	+.02 -.01 -.25 -.07	-.04
Ford, 1974	9	New York, NY (Low SES minority)	80 Students	1 sem.	Students randomly assigned to 2-group AG or hetero math classes. Same teachers taught both types of classes. Students compared on Metropolitan Achievement Test.			Math	(0)	(0)
Bicak, 1962	8	Minneapolis, MN (lab school)	1 School 75 Students	1 sem.	Students randomly assigned to 2-group AG or hetero science classes at university lab school.	HI LO	-.39 -.10	Science	-.25	-.25

Key: AG = Ability Grouping
Hetero = Heterogeneous Assignment
HI = High Achieving Students
AV = Average Achieving Students
LO = Low Achieving Students

Article	Grades	Location	Sample Size	Duration	Design	Effect Sizes By Achievement	Effect Sizes By Subject	Total
Matched Experimental Studies								
Lovell, 1960	10	Panama City, FL	500 Students	1 yr.	Matched students assigned to 5-group AG or hetero English, biology, and algebra classes.	HI (+) AV (0) LO (0)	Language (+) Math (0) Biology (0)	(0)
Billett, 1928	9	Painesville, OH	408 Students	1 yr.	In three successive years, matched students assigned to 3-group AG or hetero English classes. Compared gains on standardized tests.	HI -.11 AV +.03 LO +18	English +.04	+.04
Platz, 1965	9	?	298 Students	1 sem.	Matched students assigned to 3-group AG or hetero science classes. Students compared on standardized science test.	HI +.24 AV -.10 LO +.22	Science +.22	+.22
Bailey, 1968	9	St. Louis, MO	255 Students	1 yr.	Matched students assigned to 2-group AG or hetero algebra classes. Same teachers taught both types of classes. Students compared on gains on standardized algebra measure.	HI +.18 LO -.24	Math -.03	-.03
Thompson, 1974	11	Suburban VA	240 Students	1 yr.	Compared students in 2 schools, one AG in social studies, one hetero. Students matched. Compared gain scores on teacher-made tests.	HI-.50 HI AV-.47 LO AV-.43 LO-.54	Social Studies -.48	-.48
Barton, 1964	9	Rural UT	204 Students	1 yr.	Matched students assigned to 4-group AG or hetero English classes, compared on California Achievement Test gains.	HI+.22 HI AV-.03 LO AV-.13 LO-.20	Reading +.06 Language -.13	-.04
Wilcutt, 1969	7	Bloomington, IN (lab school)	156 Students	1 yr.	Matched students assigned to 4-group flexible AG in math or to hetero. Grouping changed 8 times in the year.		Math -.15	-.15
Holy & Sutton, 1930	9	Marion, OH	148 Students	1 sem.	Matched students assigned to AG, hetero algebra classes. Same teacher taught all classes.		Math +.28	+.28
Martin, 1927	7	New Haven, CT	83 Students	1 yr.	Matched students assigned to 3-group AG or hetero.	HI +.12 AV -.06 LO +.23	Reading +.17 Math +.13 Language +.03 Social Studies .00 Science -.04	+.10

Article	Grades	Location	Sample Size	Duration	Design	Effect Sizes By Achievement	Effect Sizes By Subject	Total
Correlational Studies								
Kerckhoff, 1986	5-10	Britain	8,500 Students	5 yrs.	Longitudinal study of students throughout Britain who attended streamed or unstreamed secondary schools.		Reading +.02 Math +.03	+.03
Fogelman, Essen, & Tibbenham, 1978	6-10	Britain	5,923 Students	4 yrs.	Retrospective study compared students who had been in streamed, partially streamed, or heterogeneous schools throughout secondary school, controlling for grade 5 general ability.		Reading +.02 Math +.03	+.03
Borg, 1965	6-9 7-10 8-11 9-12	Utah	2,934 Students	4 yrs.	Longitudinal study of students in districts using AG compared to students in neighboring district using heterogeneous grouping, controlling for pretests.	HI (0) AV (0) LO (0)	Math (0) Science (0)	(0)
Ferri, 1971	5-6	Britain	28 Schools 1,716 Students	2 yrs.	Streamed and non-streamed schools matched on 7+ (grade 2) reading, followed 4 years in junior school, 2 years in secondary.	HI (0) AV (0) LO (0)	Math (0) English (0)	(0)
Breidenstine, 1936	7-9	Soudersburg, PA	11 Schools 860 Students	1 yr.	Compared students in 4 AG, 7 hetero schools matched on IQ.		Composite Achievement	.19
Purdom, 1929	?	?	700 Students	1 sem.	Matched students in AG, hetero English and algebra classes compared in achievement.	HI -.02 AV -.08 LO +.07	English -.02 Algebra .00	-.01
Postlethwaite & Denton, 1978; Newbold, 1977	5-7	Britain	1 School 450 Students	2 yrs.	Students within one secondary school assigned to streamed or unstreamed halls. Achievement assessed on national examinations.	HI (0) AV (0) LO (0)	Math (0) English (0) Social Studies (0) French (0)	(0)
Bachman, 1968	7	Portland, OR	15 Schools 23 Teachers 404 Students	1 yr.	Math classes in schools using AG compared to hetero classes, controlling for IQ.		Math (0)	(0)
Kline, 1964	9-12	St. Louis, MO	4 Schools	4 yrs.	Retrospective study of successive cohorts of students, one in 3- or 4-group AG, one hetero, in 4 schools. Compared on standardized tests after 4 years of AG or hetero placement.	V.HI -.02 HI +.08 AV .00 LO -.02	Reading -.05 Language +.07 Math +.01	+.01

Article	Grades	Location	Sample Size	Duration	Design	Effect Sizes By Achievement		By Subject		Total
Stoakes, 1964	7	Cedar Rapids, IA	3 Schools	1 yr.	Matched mentally advanced and slow-learning students compared in schools using AG or hetero assignment. Compared on standardized tests.	HI LO	(0) (0)	Reading English Math	(0) (0) (0)	(0)
Martin, 1959	6-8	Nashville, TN	3 Schools	2 yrs.	Retrospective study compared gains on Stanford Achievement Tests for 2 AG and 1 hetero school from grades 6-8.	HI AV LO	(0) (0) (0)	Reading Language Math	(0) (0) (0)	(0)
Chiotti, 1961	9	Issaquah, WA	3 Schools	1 yr.	Matched students in 3-group AG and hetero schools compared in math achievement.	HI AV LO	+.14 +.06 +.35	Math	+.18	+.18
Fowlkes, 1931	7	Glendale, CA	2 Schools	1 sem.	Students in school using 3-group AG based on IQ matched with students in hetero school. Compared gains on Stanford Achievement Tests.	HI AV LO	-.45 -.18 -.05	Reading Language Math Social Studies	-.04 -.17 -.17 -.21	-.20
Cochrane, 1961	8	Kalamazoo, MI	1 School	1 yr.	Compared students grouped separately for English, math, to previous year (hetero) students matched in IQ, age, sex, ach.	HI AV LO	(0) (0) (0)	Math English	(0) (0)	(0)