

| AOTHOR | Glass, Gene V.; Smith, Mary Lee |
| :---: | :---: |
| IItLe | Msta-Analysis of Research on the Relationship of Class-Size and Achievement. The Class Size and |
|  | Instruction Project. |
| INSTITOTION | Par Hest lab. for Educational Research and |
|  | Development, San Francisco, Calif. |
| SPONS AGENCY | National Inst. of Education (DHEW), Washington, f |
| sub Date | Sep 78 首 |
| CONTRACT | C80 88 |
| NOTE | 81p.; appendix may not reproduce clearly |
| AVAILABLE FROM | Far West Laboratory for Educational Research and |
|  | Development, 1855 Folsom Street, San Francisco, |
|  | California 94103 (\$5.50; prepaid orders only) |
| EdRS PRICE | MFO 1/PCO4. Plus Postage. |
| DESCRIPTORS | *Academic Achievement; Achievement Rating; Cnarts: |
|  | *Class Size; *Comparative Analysis; Educational |
|  | Research; Elementary Secondary Education; Literature |
| $\because$ | Reviews; *Statistical Analysis; Statistical Data; |
| - | Statistical Studies: Tables (Data) |

## ABSTRACT

The first in a series of reports by the Far dest Laboratory for Educational Research and Development, this report demonstrates the positive relationship between reduced class sizze and pupil achievement. The researchers collected about 80 studies that yielded over 700 comparisons of the achievement of smaller and larger classes. The results showed that as class size increases, achievement decreases. For example, the difference in being taught in a class of 20 versus a class of 40 shows an advantage of 6 percentile ranks. The relationship between class size and achievement is slightly stronger at the secondary level, but it does not differ appreciably across different school subjects, levels of pupil IQ, or several other demographic features of classrooms. The report suggests that schools cannot afford the consequences of maintaining large classes all the time and must find ways to finance smaller classes for some pupils or for all pupils for part of the school dar. (Author/LD)

[^0]
## META-ANALYSIS OF RESEARCH ON THE RELATIONSHIP OF CLASS-SIZE AND ACHIEVEMENT <br> Gene V Glass <br> Mary Lee Smith

"PERMISSION TO "REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Far hest Lub. for.
Edveational R 4 D
to The educational resources INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The project presented or reported, herein was performed pursuant to a grant from the National Institute of Education, Deportment of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, the Department of Health, Education, and Welfare, or the Far West Laboratory for Educational Research and Development, and no official endorsement by these agencies should be inferred.

# META-ANALYSIS OF RESEARCH ON THE RELATIONSHIP OF CLASS-SIZE AND ACHIEVEMENT 

Gene V Glass<br>Mary Lee Smith<br>Laboratory of Educational Research University of Colorado

September 1978

## ACKNOWLEDGMENTS

This report was produced under a subcontract (C8088) from the Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94103. The subcontract was made possible by grants from the National Institute of Education to the Far West Lab.

Our work was undertaken at the request of Dr. Leonard S. Cahen and Dr. Nikola N. Filby of the Far West Lab, whose consultation early in the project and sustained interest are here gratefuily acknowledged. At the University of Colorado, we enjoyed the able statistical consultation of Gregory A. Camilli, the freedom of the Latoratory of Educational Research to drop other things and work on the current enthusiasm, and the patient tolerance of colleagues Kenneth Hopkins and Lorrie Shepard while we were preoccupied.

Viki Bergquist was forced to watch as her extraordinarily competent and ceïeritous typing was mutilated by last-minute changes.

## PREFACE

This is the first in a series of reports to be published by the Class Size and Instruction Project, of the Far West Laboratory. A second meta-analysis, also under the direction of Drs. Gene V Glass and Mary Lee Smith, will be focused on the relationship of class size and classroom processes, teacher satisfaction, and pupil affect. It is scheduled for publication in early 1979. In the spring of 1979, a group of policy-makers will be commissioned to react to the metaanalyses. Information on obtaining these documents as they become available plus other publications emanating from the Class Size and Instruction Project may be obtained by contacting me, Dr. Leonard S. Cahen, at the address below.

Drs. Glass and Smith have demonstrated that reduced class size and pupil achievement are indeed associated. Their search has uncovered many studies that have not been examined in earlier investigations of class size. The class size issue begs in vain for a simple answer to the complex question, "What is the ideal class size?" The research synthesis reported here does demonstrate the trend: very small achievement advantages are expected when small reductions are made in class size in the $20-30$ pupil range and large advantages when class size is reduced below 20. The reader must wrestle with value judgments. Are the advantages worth the cost? In a country that prides itself on quality education for all, the answer might be straightforward: schools cannot afford the consequences of maintaining large classes all the time, and ways must be found to finance smaller classes, at least for some pupils or for all pupils for part of the school day.

I wish to thank my colleague, Dr. Nikola N. Filby, for her major contributions to the Project and for her consultation on the meta-analysis report. Drs. David C. Berliner and Richard M. Jaeger provided excellent critical reviews of an earlier version of the manuscript. I also wish to thank Joseph Vaughan and Virginia Koehler of the National Institute of Education for their continued support and counsel.

Dr. Leonard S. Cahen
Principal Investigator
Class Size and Instruction Project
Far West Laboratory
1855 Folsom Street
San Francisco, CA 94103
20 September 1978

SUMMARY

Research on the relationship between school class-size and academic achievement is old, huge and widely believed to be inconclusive. Previous reviews of the evidence "have been overly selective and insufficently quantitative. Timid qualifications were offered where boid generalizations were possible. In the summer of 1978, the New York Times gave front-page coverage to a study published by Educational Research Service, Inc. (Porwell, 1978). This organization is funded jointly by the American Association of School Administrators, the Council of Chief State School Officers, and several other professional administration groups. The "Porwell:Report" staggered visibly under the weight of the research data and eventually arrived at the following conclusions sad for teachers to behold:

Research findings on class size to this point.document repeatedly that the relationship between pupil achievement and class size is highly somplex.

There is general consensus that the research findings on the effects of class size on pupil achievement across all grades are contradictory and inconclusive.

Existing research findings do not support the contention that smaller classes will of themselves result in greater academic achievement gains for pupils.
(Porwel1; 1978, pp. 68-69)

The research reported herein contradicts the conclusions of the Porwell Report. Indeed, it established clearly that reduced class-size can be expecied to produce increased academic achievement. In pursuing this conclusion, we
discovered many of the reasons why previous research reviewers lost their way in the forest of data and failed to find a defensible generalization.

We collected nearly 80 studies of the relationship between class-size and achievement. These siudies yielded over 700 comparisons of the achievement of smaller and larger classes; these comparisons rest on data accumulated from nearly 900,000 pupils of all ages and aptitudes studying in all manner of school subjects. Using complex methods of regression analysis, the 700 comparisons were integrated into a single curve showing the relationship between class-size and achievement in general. This curve revealed a definite inverse relationship between class-size and pupil learning. Similar curves were derived for a variety of circuinstances hypothesized to alter the relationship between achievement and class-size. Virtually none, of these special circumstances altered the basic relationship; not grade level, nor subject taught, nor ability of pupi`s. Only one factor substantially affected the curve, viz., whether the original sidy controlled adequately (in the experimental sense) for initial differences among pupils and teachers in smaller and larger classes. The nearly 100 comparisons of achịevement from the well-controlled studies this form the basis of our conclusion about how class-size is related to academic achievement. The most accurate representation of this relationship is a curve derived from the 100 comparisons froin well-controlled studies. This curve appears in the Figure below. As classsize increases, achievement decreases. A pupil, who would score at about the 83rd percentile on a national test when taught individually, wouid score at about the 50th percentile when taught in a class of 40 pupils. The difference in being taught in a class of 20 versus a class of 40 is an advantage of 6 percentile ranks. The major benefits from reduced class-size are obtained as size is reduced below 20 pupils.


Figure 1. Relationship between achievement and class-size. (Data integrated across approximately) 10 l comparisons from studies exercising good experimental control.)

There is no point in recording the obvious about class-size: that teachers worry about it more than nearly anjthing else, that administrators want to increase it, that it is economicaliy imporiant, and the like. The problem with class-size is the research. It is unclear. It has variously been read as supporting larger classes, supporting smaller classes, and supporting nothing but the need for better research. Review after review of the topic has dissolved into cyrical despair or epistemological confusion. The notion is wide-spread among educators and researchers that class-size tears no relationship to achievement. It is a dead issue in the minds of most instructional researchers. To return to the class-size literature in search of defensible interpretations and conclusions strikes many as fruitless. The endeavor is surrounded by a faint aroma of Chippendale, which it resembles in other respects: unwieldy and antique.

One could document the confusion, in previous revieivs of research on the class-size ant achievement relationship. It would be simple to quote reviewer $x$ claiming that large classes are better, reviewer $Y$ to the effect that small classes are bettei;, and reviewer $Z$ that neither is better. But to do so would only embarrass others and add nothing to one's appreciation of the complexity of the research. The problems with previcus reviews of the class-size literature are several: (1) literature searches were haphazard and often overly selective; dissertations were avoided, as a rule, and few reviewers sought out large archives of pertinent data; (2) reviews were typically narrative anc discursive; the multiplicity of findings cannot be absorbed without quantitative methods of
reviewing; (3) reviowers that attempted quantitative integration of findings made 'several mistakes: (a) they used crude classifications of class-sizes; (b) they took "statistical significance" of differences far too seriously; and (c) they lacked sufficiently sophisticated techniques of integrating results.

In the research reported here, an attempt was made to correc: these shortcomings and determine if the huge research literature on class-size and achievement really was hopelessly confusing or if its message was merely buried in myriad results waiting to be coaxed out with more advanced methods of research integration.

## The Literature Search

- The search for class-size studies was carried out in three places: (1) document retrieval and abstracting resources; (2) previous reviews of the classsize literature; and (3) the bibliographies of studies once found. The ERIC system and Dissertation Abstracts were searched completely on the key words "size," "class size," and "tutoring." The dissertation literature was covered as far back as 1900 , and the fugititive educational résearch literature was covered from the mid-1960s to 1978. Of the many huncireds of doctoral dissertations scanned in Dissertation Abstracts, about thirty microfilm copies were purchased. About a dozen of these dissertations were incorporated; the remainder dealt with non-achievement and process variables that will be covered in subsequent work. The journal literature on class-size was located in the traditional way; one or two current reviews of the research were found -- the Ryan and Greenfield (1975) review and the comprehensive review by Lafieur, Sumner and Witton (1974) --

Were particularly comprehensive and helpful -.. un arcicles cited were located, and the articles ced in these articles were located in turn.

Approximately 300 documents were obtained and read. One hundred-fifty of them were found to contain no usable data, i.e., no data whatsoever were reported on the comparison of small- and large-class achievement. About 70 studies examined the relationship of class-size to non-achievement outcomes and classroom process variables. Approximately 80 studies on the class-size and achievement relationship were included in this analysis.

It is difficult to estimate what portion of the existing literature was captured by this search. Even though the corpus of 80 studies exceeds by 50 percent the most extensive reviews published to date -- and these reviews are narrative and inconciusive -- it is conceivable that less than half of all studies that exist on the topic were found. Some studies (credited to school districts) could not be located even after several phone calls and letters. Other studies were surely missed because of odd or nondescript titles. The dissertation search was conducted on key words such as "size," "class-size," anci "tutoring;" but thile words must appear in titles to be registered in the index to Dissertation Abstracts. (Fortunately, the ERIC system uses key words based on the contents of a paper and not titles alone.) Several studies found in the journal literature by branching off existing bibliographies had neither "size" nor "class-size" in the title, evidence enough that several dissertations were missed"because their titles lacked the key wards. Still another complication concerns the use of class-size as an incidental variable in studies focused on other issues. There
are probably many such studies, and only a few of : whes were located.

## The Texture of the Literature

In what follows in this integrative analysis, one can easily lose touch with precisely what kinds of research are being integrated. : The statistics and graphs that represent the findings of this meta-analysis of class-size research will seem far removed from the original studies themselves. And, in a very real sense, what will be done fo" the sake of arriving at general conclusions places the reader in benign jeopardy of losing qualitative and personal familiarity with the research. In this section, the general texture of the class-size literature will be described, and a few studies typical of various eras will be reported.

The research on class-size and its relationship to achievement falls into four stages: the pre-experimental era (1895-1920); the primitive experimental era (1920-1940); the large-group technology era (1950-1970); and the individualization era (1970-present). The boundaries of the eras are not impenetrable, and even today an ata!istic throwback to the 19 th century will appear in a doctoral thesis. At each new stage, the sophistication of research methodology increased, and the question of class-size and its effect on achievement was examined with different motives. One discerns in the narration accompanying the numbers the cult of efficiency of the early part of this century, the rising birth rate of the post-war ' 40 s , the advent of teaching technology in the ' 60 s , and most recently the teacher labor movement combined with declining enrollments. What
was said about the data changed as new intermintations served emerging purposes, even when the data changed litt $\mathrm{T}^{\mathrm{i}} \mathrm{e}$ them:

The first empirical study on educac ucesses and their effects on achievement included an examination of the class-size question (Rice, 1902). No strong relationship of class-size to attainment was observed. But unfortunately, Rice reported virtually no numbers; and it is impossible to determine now whether the relationship Rice found was genuinely small or whether it was moderately, large but only seemed small to Rice, who may have expected much more. Rice's study was followed by several similar analyses. on new data collected between 1900 and 1920. These studies are typified by their rugged non-experimental logic. A study by Cornman (1909) can serve as an example.

Cornman examined the promotion records for January 1909, in District No. 6, Philadelphia. Before the day of "social promotion," the passage from one grade to the next higher indicated adequate achievement at the lower grade. Cornman categorized classes into three groups: under 40 pupils, 40 to 49 , 50 or more. The rate of promotion was calculated for each class-size category. At grade 3, 88 percent of 400 pupils in classes of 40 or fewer were promoted, 85 percent of 1,300 pupils in classes size 40 to 49 were promoted, and 81 percent of 640 pupils were promoted in classes of over 50 pupils. Cornman also investigated "satisfactory conduct" ratings by teachers in classes of different sizes. The discussion of results showed little sensitivity to questions of experimental control; such concerns were doubtless not wide-spread at the time.

Beginning in the early 1920s, the class-size and achievement question was approached with better methods: Studies began to appear that used matching of
pupils in large and small classes on ability and achievement; content and methods were standardized in the two clascra rensionally the same tronhers taught
 between class-size arsd achievement in grammar and English at the high-school level in Grand Junction, Colorado. In the Fall of 1922, three. English classes of 44, 34, and 20 pupils were formed. Their Terman Group Test IQs were nearly identical at the first, second; and third quartiles. "After thoroughly establishing our classes, our method of conducting the experiment was merely to proceed with the year's work in the usual way, except that we found it necessary to depend rather more than usual on test grades, because the number of pupils in the large class made it impossible for each pupil to make many daily recitations each period" (p. 127). The experiment was run for nine weeks. Then the Starch Grammar Test and Kirby Grammar Test were administered along with some specially designed classroom tests on clauses. The findings slightly favored the two smaller classes over the class of 44.

In the 1940s, class-șize research went dormiant when educational researchers went to war. It was revived along with the rest of the field in the 1950 s and 1960s. Researchers seemed intent on demonstrating, particularly at the college level, that lecture classes could be doubled or tripled in size without loss of effectiveness. At about the same time, massive empirical studies of education were undertaken to inform national education policy: the Coleman study of equality of educational opportunity (1966); Project TALENT; the International Assessment of Education in mathematics and reading; and surveys of government-
funded programs of compensatory education (Title I). These large empirical studies typically included, as incidental features, data on the relationship of class-size and achievement. The study by Nelson (1959) is representative of the first kind of study' to appear in the 'fus and '60s; the Colemar $: 2$. ('966) study is like many studies of the second type.

In 1959, Nelson reported on a study of large-group coilege instruction. Four instructors were involved, each teaching one large and one small section of elementary economics. The pupils in each instructor's classes were matched on major (e.g., business, engineering), level (freshman, sophomore), and sex. The course was taught three hours a week for a semester. The class-sizes compared were 20 vs. 138,16 vs. 141,20 vs. 94,20 vs.. 90,17 vs. 309,17 vs. $94 ; 19$ vs. 85. A common final examination was administered to al.1 14 classes. Achievement outcomes were adjusted by covarying on students' prior grade-point average. The means favored the larger classes by three one-thousandths standard deviation!

The Coleman study is famous. Tens of thousands of pupils in grades $1,3,6$, 9, and 12 were surveyed. Achievement tests were administered and "school resources" were neasured at the level of the school, e.g., teachers' experience, use of special programs. Among these resource variables was pupil/instructor ratio. The P/I ratio was correlated with pupil achievement. The correlations. were generally negative. When Mayeske et al. (undated) partialed out three or four other variables which might have obliterated these correlations, the r's remained consistently negative.

The research relevant to class-size that appeared in the 1970 s showed a concern for establishing the benefits of individualization. Experiments were performed that involved radically reduced instructional group sizes, one teacher with two or three pupils. Studies of individual pupils taught by computer or machine have also become common; they were not considered in this integrative analysis since the particular concern here is with the processes of human instruction. (For a meta-analysis of tutoring and compüter-assisted instruction in mathematics that produced surprising findings, see Hartley, 1977.) An experiment typical of studies of radically reduced group size was condicted by Bausell et al. (1972). Pupils in grades 4 and 5 were randomly assigned to receive either individual tutoring on exponential arithmetic for one hour across two days or instruction by randumly comparable teachers for the same amount of time in a class of 25 pupils. Instruction was a part of an on-going school program. A test designed to cover only the content of the instruction was administered to all pupils. Pupils in "class-size 1 " scored approximately one-hatf standard deviation above pupils in classes of 25 on the achievement tests.

## Methods

In this section, the methods are described by which the stur'ies were coded and the quantitative fịndings integrated.

## Defining the Field

The problem of this meta-analysis is to determine what the available research proves about the relationship of class-size to achievement. Drawing boundaries around this topic wis simple compared to the difficulties encountared in defining psychotherapy, for example (Smith and Glass, 1977). Conventional definitions of "achievement" seem scarcely to have changed over eighty years. "Class-size". can be described and quantified in several different ways, but it was relatively easy to select one approach. Definitions of class-size differ in terms of how close they are to the reality of the . child's experience in the classroom. Some definitions, such as "Numerical Staff Adequacy," reflect the ratio of staff to pupils on a district-wide basis. Such definitions are relatively distant from the classroom. On theother hand, within a conventional classroom -unit, several instructors can be present, thus reducing the actual instructional group size for a particular student. Instructional group size is very close to the child's experience in the classroom. Because of an interest in the classroom processes that pres sumably mediate the relationship of class-size to achievement, we chose a definition which is close to classroom reality". In this review, "class-size" is defined as the ratio of pupils to instructors, or instructional group size. In most studies, this was the same as the size of the classroom unit, but in some it was not.

## Coding Cnaracteristics of Studies

 $\cdot$Whe quantification of characteristics of studies permits the eventual statistical description of how properties of studies affect the principal findings. Such questions can be addressed as "How does the class-size and achieve"ment relationship vary as a function of age of pupils?" or "How does it vary between reading and math instruction?" The first step in coding studies is to identify those properties of s.tudies that might interact with the relationship between class-size and achievement. There is no systematic and logical procedure for taking this step. One simply reads a few studies from the literature of interest, talks with experts, and then makes a best guess; modifications can always be made later if needed. The best guesses as to which conditions might mediate the relationship fell into five broad categories: Study Identification, Instruction, C.lassroom Demographics, Study Conditions, and Outcome Variable. About 25 specific items fell into these categories. Some were more fruitful than others; several items were seldom reported in the research publications: A
cr ting chect ias devi anty which the information about each study could be transcribed. A single study might fill several coding sheets, depending on how many different class sizes were compared in pa:rs, how many different achievement tests were reported, whether data were reported separately for different ages or : IQs, and so forth.

The major items of the coding sheet are reported below:

## IDENTIFICATION:

1) Year. This item was included to check on whether there is a time trend in the class-size and achievement relationship.
2) Source of Data: Whether fiom a journal, book, thesis, or unpublished source.

## INSTRUCTION:

3) Subject. The subject taught (reading, math, etc.) was recorded.
4) Duration of Instruction. The amount of teaching was recorded in hours and in weeks.
5) No. of Pupils. The numbers of pupils on which the small and largeclass achievement means were based were recorded. This number was not the same as the "c'ass-size" since there might be several small or large classes used in the study.
6) No. of Instructional Groups. (See \#5 above.)
7) No. of Instructors. (See \#5 above.)
8) Pupil/Instructor Ratio. This measure is the measure of class-size. One teacher with a group of 30 counts as a P/I ratio of 30 ; two teachers in a class of 30 gives a P/I of 15 .

## CLASSROOM DEMOGRAPHICS:

9) Pupil Ability.: Average IQ of the pupils was estimated when not reported; three broad categories were used: $\mathrm{IQ} \leq 90 ; 90<\mathrm{IQ}<110$; $\mathrm{IQ} \geq 110$.
10) Ages and Average Age. These two variables permitted discriminating instances in which all pupils were of one age from studies in which pupils of several ages were represented and the average age was used to describe their level since data were not reported separately. This variable was used to distinguish data from elementary and secondary school levels.

STUDY CONDITIONS:
11) Assignment of Pupils and Teachers to Groups. The assigriment of pupils and teachers to classes of different sizes was described.as either "random," "matched," "repeated measures," or "uncontrolled." These variables were important in describing the degree of experimental: control exercised in the study. "Random" is.obvious; "matched" refers to attempts to equate small and large classes by other than random means on pretests of achievement or ability; "repeated measures" refers to using either the same pupils or teacher in both small and large classes, e.c., 10 pupiits might be taught alone and then in a group of 40 and their achievement compared; "uncontrolled" should"be obvious.
12) Type of Achievement Measure. Outcomes were measured by standardized achievement tests, specially designed (ad hoc) tests, or teachers' assessments of achievement. The latter two categories were grouped.
13.) Quantification of Outcomes. In some instances, a degree of experimental control could be attained by expressing achievement as gains from pretest to posttest or covariance adjusting. posttest means for pretest differences. If this was done, it was noted.

## Quantifying Outcomes

A simple statistic is desired that describes the relationship between classsize and achievement as determined by a study. No matter how many class-sizes are compared, the data can be reduced to some number of paired comparisons, a smaller class against a larger class. Certain differences in the findings must be attended to if the findings are later to be integrated. The most obvious differences involve the actual sizes of "smaller" and "larger". classes and the scale properties of the achievement measure. The actual .class-ṣizes compared must be preserved and become an essential part of the descriptive measure. The measurement scale properties can be handled by standardizing all mean differences in achievement by dividing by the within group standard deviation (a method that is complete and discards no information at all under the assumption of normal distributions). "The eventual measure of relationship seems straightforward and unobjectionable:

$$
\Delta_{S-L}=\frac{\ell_{S}-\chi_{L}}{\hat{\sigma}},
$$

where:
$X_{S}$ is the estimated mean achievenent of the smalier class which contains S pupils;
$\bar{X}_{L}$ is the estimated mean achievement of the larger class which contains L pupils; and
$\hat{\sigma}$ is the estimated within-class standard deviation, assumed to be homogeneous across the two classes.

As a first approximation to'studying the class-size and achievement relationship, it is considered irrelevant that the particular types of achievement that lie behind the variable $\underline{X}$ are quite different knowledges and skills measured in quite different ways.

If distributional assumptions about $\underline{X}$ are needed to add meaning to particular values ci $\Delta_{S-L}$, normality will be assumed. For example, suppose $\Delta_{S-L}=+1$. Then assuming normal uistributions within classes, the average pupil in the smaller class scores at the 84 th percentile of the larger class. These interpretations are occasionally helpful, but seldom critical, and our investment in the normality assumption is not great. It would be no surprise nor any concern if the assumption proved to be more'or less wrong, and it's probably not far off in most instances.

Calculating $\Delta_{S-L}$ -
Reports of research frequently omit such basic descriptive measures as means and standard deviations. This omission frequently complicates the caiculation of $\Delta_{S-L}$, but seldom obviates it. Transformations of commonly reported statistics
(t, $\underline{F}$, etc.) into $\Delta$ 's can be derived (Glass, 1978). A specia? problem in calculation of $\Delta_{S-L}$ concerns studies in which class-size is correlated with achievement across many classrooms (e.g., Coleman, 1966; Robinson, 1963). In these instances, $\Delta_{S-L}$ was calculated as follows. The distribution of class-sizes was determined by assuming normality and noting the mean and standard deviation: The regression coefficient was calculated for the regression of achievement (assumed to be calculated on a unit-normal scale) onto class-size via $\hat{\beta}=r_{A, C S} / \hat{\sigma}_{C S}$. Then the class-sizes at the 25 th and 75 th percentiles, assuming normality, were deter̂mined. These became the "smaller" and "larger" classes. Finally, the achievement in these classes was determined via the formula $\hat{B}(\dot{X}-\mathbb{X}$.) where $\underline{X}$ is "class-size.". The value of $\Delta_{S-L}$ is then" readily calculated, Some studies involved only a dichotomous achievement measure (e.g., "promoted (to the next grade) vs. not promoted"). Proportions thus derived were transformed into metric information and then into values of $\Delta_{S-L}$ by means of the probit transformation (see Glass،, 1978).

## Describing the Class-Size and Achievement Relationship

There exist several alternative statistical techniques for integrating a large set of $\dot{\Delta}_{S-L}$ 's so as to describe the aggregated findings on the class-size and achievement relationship. A large, square matrix could be constructed in which the rows and columns are class-sizes and the cell entries are average values of $\Delta_{S-L}$; nearly equal values of average deltas could be connected by lines
to form "iso-deltas" in much the manner as economic equilibrium curves are used to depict three variable relationships. Or a variation of psychometric scaling could be employed: a square matrix of class-sizes could be constructed for which each cell entry would be the proportion of times the row class-size gave achievement greater than the column class-size. This matrix could be scaled by means of Thurstone's Law of Comparative Judgment, which would locate the classsizes along an achievement continuum. (This method was used and the results were reaṣnably satisfactory; but they add little to findings obtained by more direct means, that are reported here.) . Finally, regression equations could be constructed in which $\Delta_{S-L}$ is partitioned into a weighted linear combination of $\underline{S}$ and $\underline{L}$ and functions thereof and error. There is much to recommend this latter procedure, and the technique eventually employed is a variation of it. But the regression of $\Delta_{S-L}$ onto only $\underline{S}$ and $\underline{L}$ requires three dimensions to be depicted. Anything more complex than a simple two-dimensional curve relating achievement to the size of cláss was considered undesirably complicated and beyond the easy reach of most audiences who hold a stake in the results.

The desire to depict the aggregate relationship as a single line curve is confounded with the problem of essential inconsistencies in the design and results of the various studies. A single study of class-size and achievement may yield several vatues of $\Delta_{S-L}$. In fact, if $\underline{k}$ different class-sizes are compared on a single ach evement test, $k(k-1) / 2$ values of $\Delta_{S-L}$ will result. This set of ${ }^{\circ}$ $\Delta^{\prime}$ s from a single study will form a consistent set of values in that they can be joined to form a single connected graph depicting the curve of achievement as a
function of class-size. However, various values of $\Delta_{S-L}$ arising from different studies can show confusing inconsistencies. For example, suppose that Study \#l gave $\Delta_{10-15}, \Delta_{10-20}$, and $\Delta_{15-20}$, and Study \#2 gave $\Delta_{15-30}, \Delta_{15-40}$, and $\Delta_{30-40}$. A few moments reflection will reveal that there is no obvious or simple way to connect these values into a single connected curve.

The eventual solution to these problems proceeded as follows: $\Delta_{S-L}$ was regressed onto a quadratic function of $\underline{S}$ and $\underline{L}$ by means of the least-squares criterion; then that set of values of $\hat{\Delta}$ that could be expressed as a single, connected curve was found.

The regression model selected accounted for variation in $\Delta_{S-L}$ by means of $\underline{S}$. $\underline{S}^{2}$ and $\underline{L}$. Obviously, something more than a simple linear function of $\underline{S}$ and $\underline{L}$ was needed, otherwise a unit increase in class-size would have a constant effect regardless of the starting class-size $\underline{S}$; and the $\underline{S}^{2}$ term seemed as capable of filling the need as any other. The size differential between the larger and smaller class, L-S, was used in place of $\underline{L}$ for convenience. Thus, the $\Delta_{S-L}$ values were used to fit the following model:

$$
\begin{equation*}
\Delta_{S-L}=\beta_{0}+\beta_{1} S+\beta_{2} S^{2}+\beta_{3}(L-S)+\varepsilon \tag{1}
\end{equation*}
$$

Fitting this model by least-squares will result in the curved regression surface

$$
\begin{equation*}
\hat{\Delta}_{S-L}=\hat{\beta}_{0}+\hat{\beta}_{1} S+\hat{\beta}_{2} S^{2}+\hat{\beta}_{3}(L-S) \tag{2}
\end{equation*}
$$

- The problem now is to find the set of, $\hat{\Delta}$ 's in this surface that can be depicted as a single curved-line relationship in a plane. The property that must hold for a set of $\hat{\Delta}$ 's before they can be depicted as a connected graph in a plane is what might be called the consistency property:

$$
\Delta_{n_{1}-n_{2}}+\Delta_{n_{2}-n_{3}}=\Delta_{n_{1}-n_{3}}
$$

for $n_{1}<n_{2}<n_{3}$. If this property is not satisfied, then one is in the strange
situation of claiming that the differential achievement between class-sizes $10^{\circ}$ and 20 is not the sum of the differential achievement from 10 to 15 and then from 15 to 20.

When the consistency property is imposed on (2), it follows that:

$$
\begin{align*}
& \hat{\beta}_{0}+\hat{\beta}_{1} n_{1}+\hat{\beta}_{2} n_{1}^{2}+\hat{\beta}_{3}\left(n_{2}-n_{1}\right)+\hat{\beta}_{0}+\hat{\beta}_{1} n_{2}+\hat{\beta}_{2} n_{2}^{2}+\hat{\beta}_{3}\left(n_{3}-n_{2}\right) \\
& =\hat{\beta}_{0}+\hat{\beta}_{1} n_{1}+\hat{\beta}_{2} n_{1}^{2}+\hat{\beta}_{3}\left(n_{3}-n_{1}\right) \tag{3}
\end{align*}
$$

Simple algebraic reduction of (3) produces the following:

$$
\begin{equation*}
\hat{\beta}_{0}+\hat{\beta}_{1} n_{2}+\hat{\beta}_{2} n_{2}^{2}=0 \tag{4}
\end{equation*}
$$

The two solutions to the quadratic equation in (4) are points $n_{2}$ such that if $\hat{\Delta}_{S-L}$ is measured with $n_{2}$ as either the larger, $\underline{L}$, or smaller, $\underline{S}$, class-size, then the resulting set. of $\hat{\Delta^{\prime}} s$ will lie on the four dimensional regression curve in (2) but can be depicted as a single line curve in a plane. Since $n_{2}$ becomes the point around which values of $n_{1}$ and $n_{3}$ are selected, it will be called the pivot point. That there are two solutions for $n_{2}$ is perplexing; fortunately, in the analyses to be reported the two corresponding curves were virtually parallel in practice.

A single line curve in a plane can be constructed by solving for one or the other values of $n_{2}$ in (4) and constructing a set of $\hat{\Delta}$ values. These values will give the standardized mean differences in achievement between $n_{2}$ and any other class-size. The curve that connects these $\hat{\Delta}$ 's has no non-arbitrary starting point. One can assume for convenience sake that the achievement curve ( $\underline{z}$ ), instead of the differential achievement curve $(\hat{\Delta})$, is centered around an arbitrary
class-size, e.g., something like the national average in the low $20^{\prime}$ s. Finally, for descriptive purposes, the metric of percentile ranks was chosen over the metric of $\underline{z}$-scores; thus the curve $\underline{z}$ was transformed into a curve of percentile ranks by assuming a normal distribution of achievement.

## Comment on Statistical Inference

In the analyses that follow, ordinary matters, of statistical inference have been ignored. The application of usual interval estimation procedures or statistical tests makes little sense for two reasons. The data base is laced with a complicated structure of interdependent observations; several comparisons arise from a single study when more than two class-sizes are compared, and there is no sensible way to reduce each study to one observation. Even if a'study involves comparing only two class-sizes, there might have been comparisons of reading and math achievement. It makes far less sense to average these than to let each be separately entered in the data base. The data bases of most metaanalyses are complex nested and multi-level arrangements. The methods of analyzing them fully await a full explication; methodological work on these problems has been launched in promising directions (Burstein, 1978). Secondly, randomization is absent from the data set in any form that would make probabilistic models based on it anplicable. To the extent that one might care to infer to populations of pupils, the sample size is so large that significance tests would be an empty pro form ritual. To the extent one might wish to infer to populations of studies, it must be recognized that the studies included have in no way been.
sampled from any conceivable population. Error and instability of various odd sorts exist in the data set; how they should be dealt with is not at all apparent.

## Findings

The report of findings falls into two broad categories: (1) description of the data base and (2) regression analyses relating achievement and class-size.

Description of the Data Base
In $\mathrm{a} 11,77$ different studies were read, coded, and analyzed. These studies yielded a total of $725 \Delta$ 's. The comparisons are based on data from a total of nearly 900,000 pupils spanning 70 years research in more than a dozen countries. (The entire set of data is reproduced in the appendix to this report.)

The total body of evidence can be described parily in quantitative terms through use of frequency distritutions of characteristics of the studies. These tabulations will be presenter in terms of $\Delta$ 's rather than studies. The descriptive data do not only communicate an understanding of the evidence upon which the conclusions rest; they point to the relatively over-studied and under-studied aspects of the topic and can help guide future research on class-size and achievement.

In Table 1 appears the frequency distribution of $\Delta^{\prime}$ 's by year in which the study appeared. It is clear from Table 1 that class-size research was an active

Table 1
Class-Size Comparisons ( $\Delta$ ) by Year of Study

| Year | No. of $\Delta^{\prime} \mathrm{s}$ | \% | $\begin{gathered} \text { Cumulative } \\ \% \end{gathered}$ |
| :---: | :---: | :---: | :---: |
| 1900-1909 | 22 | 3:0\% | 3.0\% |
| 1910-1919 | 184 | 25.4\% | 28.4\% |
| 1920-1929 | 138 | 19.0\% | 47.4\% |
| -1930-1939 | 47 | 6.5\% | 53.9\% |
| 1940-1949 | 1 | 0.0\% | 53.9\% |
| 1950-1959 | 62 | 8.6\% | 62.5\% |
| 1960-1969 | 150 | 20:8\% | 83.3\% |
| 1970-1979 | 121 | 16.7\%. | 100.0\% |
|  | 725 | 100.0\% |  |

32
early topic in educational research, was largely abandoned for 30 years after 1930, and has been resurrected in the last 15 years.

In Table 2 appear data on the publication source from which the comparisons were drawn. Although published journal articles are the major source of data, about $20 \%$ of the data were found in theses and unpublished reports -- both of which have not been well covered in previous reviews.

In Table 3 appear the frequencies of comparisons categorized by the school subject taught in the study. Nearly half of the comparison came from studies in which elementary school pupils were taught all subjects in classes of varying sizes. There is surprisingly little work on reading alone; however, the 342 "all subjects combined" comparisons typically include reading as an important element.

In Table 4 are reported the numbers of hours of instruction given in the classes being compared. The range is enormous, from a single hour for a very small scale tutoring study, to 9,000 hours; representing five years of elementary school instruction. The "hours of instruction" distribution shows three modes: 50, 180, and 900 hours. These times correspond to a three credit-hour semesterlong course, a five credit-hour year-long course; and a year of teaching five hours per day. The literature does not lack studies conducted over significant intervals of time. The average duration is 536 hours with a standard deviation of 1033 hours and a skewness of 5.58 .

In Table 5 appears the distribution of comparisons for various ages of pupils. 'Research is spread fairly evenly across the elementary and secondary grades. The first four years of school are only slightly underrepresented. The

## Table ? <br> Class-Size Comparisons ( $\Delta$ ) by Publication Source

| Source | No. of $\Delta^{\prime} \mathrm{s}$ | \% |
| :---: | :---: | :---: |
| Jourrial | 474 | 65.4\% |
| Book | 114 | 15.7\% |
| Thesis | 60 | 8.3\% |
| Unpublished | 77 | 10.6\% |
|  | 725 | 100.0\% |

Table 3
Class-Size Comparisons ( $\Delta$ ) by Subject of Instruction

| Subject Taught | No. of $\Delta^{\prime} \mathrm{s}$ | \% |
| :---: | :---: | :---: |
| All Subjects Combined <br> (i.e., elementary school classes) | 343 | 47.2\% |
| Reading | 39 | 5.4\% |
| Mathematics | 84 | 11.6\% |
| Language | 144 | 19.9\% |
| Psychology | 23 | 3.2\% |
| Natural/Physical Sciences | 28 | 3.9\% |
| Social Sciences and History | 40 | 5.5\% |
| All Others | 25 | 3.4\% |
|  | 725 | 100.0\% |

Table 4
Class-Size Comparisons ( $\Delta$ ) by Hours of Instruction


## Table 5

Class-Size Comparisons ( $\Delta$ ) by Age of Pupils

| Age | No. of $\Delta^{\prime} s^{\text {i }}$ | \% | Cumulative Percent |
| :---: | :---: | :---: | :---: |
| 5-6 | 56 | 7.7\% | 7.7\% |
| 7-8 | 55 | 7.6\% | 15.3\% |
| 9-10 | 198 | 27.3\% | 42.6\% |
| 11-12 | 98 | 13.5\% | 56.1\% |
| 13-14 | 81 | 11.1\% | 67.2\% |
| 15-16 | 109 | 15.0\% | 82.2\% |
| 17-18 | 108 | 14.9\% | 97.1\% |
| 19 \& older | 20 | 2.8\% | 100.0\% |
|  | 725 | 100.0\% |  |

average age represented in the 725 comparisons is 12.3 years with a standard deviation of 4.0 years.

The next few items of information concern the experimental validity of the comparisons, i.e., the incidence of various experimental controls and ex post facto adjustments. In Table 6, the comparisons are tabuiated by the type of assignment of pupils to the different size classes. The type of assignment labeled "repeated measures" refers to the use of the same group of pupils in both a small and a large class and the comparison of their achievemer.t in the two classes. Each of the first three types of assignment represents reasonably good attempts at eliminating gross inadequacies in design; these three conditions account for slightly more than half of all the comparisons. Even though half of the comparisons involved comparing naturally constituted and non-equivalent large and 'small classes', some of them were based on ex post facto statistical adjust'ments for pre-existing differences. So the data are not half worthless; indeed whether the experimental inadequacies are important mediators of findings is an empirical fact -- rather than an a priori judgment -- which will be examined in detail later in this réport.

Many studies attempted to control for the initial non-equivalence of small and large classes by correcting the achievement dependent variable, either by calculating simple gain-scores or by covariance adjusting means. We hasten to point out that an uncorrected dependent variable does not necessarily indicate a comparison of poor quality. Corrections might be quite irrelevant in a study that matched or randomly assigned pupils to classes.

Table 6
Class-Comparisons ( $\Delta$ ) by Assignment of Pupils to the Small and Large Classes
$\qquad$

| Type of Assignment | No. of $\Delta^{\prime} \mathrm{s}$ | $\underline{\%}$ |
| :--- | ---: | ---: |
| Random | 110 | $15.2 \%$ |
| Matched | 235 | $32.4 \%$ |
| "Repeated Measures" | 18 | $2.5 \%$ |
| Uncontrolled | $\frac{362}{725}$ | $\frac{49.9 \%}{100 \%}$ |

Finally, the collparisons can be described by whether achievement was measured with a "standardized test" (i.e., a published test for a national market) or an ad hoc instrument designed specifically to measure achievement in the imnediate context of the instruction given (see Table 7).

In Table 8 appears the joint distribution of smaller and larger class-sizes on which the $725 \Delta$ 's are based. For example, six $\Delta$ 's derive from comparisons of group sizes 1 ánd 3. The table contains only 550 entries instead of 725 , since comparisons would not be recorded in this tabulation if $\underline{S}$ and $\underline{L}$ were contained within the same broad category (e.g., if $\underline{\subseteq}=18$ and $\underline{L}=22$ ). . Such comparisons were incorporated in all subsequent analyses, but the need to keep Table 8 downto a reasonable size precluded the classification of all $725 \Delta^{\prime} s$. It is apparent in Table 8 which size comparisons have been relatively overstudied and which have been neglected. The dearth of comperisons of instructional group sizes in the range from 2 to 10 pupils is particularly apparent.

## Regression Analyses

The dependent variable, $\Delta_{S-L}$, in the regression analyses had the following statistical properties:

## Properties of Distribution of $\Delta_{S-L}$

a) $N=725$.
b) Mean $=.088$; Median $=.050$.
c) $40 \%$ of the $\Delta_{S-L}$ were negative; $60 \%$, po'sitive.
d) Standard deviation $=0.401$.
e) Range: -1.98 to 2.54 .
f) Skewness $=1.151 ;$ Kurtosis $=7.461$

Table 7

Class -Comparisons ( $\Delta$ ) by Type of Achievement Measure

|  |  |  |
| :---: | :---: | :---: |
| Type of Achievement Measure | No. of $\Delta^{\prime} \mathrm{s}$ | \% |
| Standardized st | 318 | $43.0 \%$ |
| Ad Hoc Measure | $\frac{407}{725}$ | $\frac{56.1 \%}{100.0 \%}$ |

"

Table 8
Joint Distribution of Smaller and Larger Class-sizes in the Comparisons $\Delta_{S-L}$


On the average, the $725 \Delta_{S-L}$ 's were positive, i.e., over all comparisons available -- regardless of the class sizes compared -- the results favored the smaller class by about a tenth of a standard deviation in achievement. This finding is not too interesting, however, since it disregards the sizes of the classes being compared. One interesting feature of the $\Delta$ 's is that only $60 \%$ of them are positive, i.e., favor the smaller class in achievement. This is so, even though every effort was made in compiling the data base to include studies spanning the full range of class-sizes from individual tutorials to huge lectures. One suspects that the odds of observing a positive $\Delta_{S-L}$ in the typical class-size range so often studied ( 15 to 40 , say) are even smaller, perhaps as low as $55 \%$ to $45 \%$.

In these rough estimates, one of the fundamental problems is revealed that has made the class-size literature so difficult for reviewers. If the relationship one seeks has only 55 to 45 odds of appearing and one looks for it without all the tools of statistical analyses that cạn be mustered, the chances of finding it are small. One need not wonder why narrative reviews of a dozen or two studies produced little but confusion.

To make sense of the class-size and achievement reiationship, one must account for the magnitude of the $\Delta^{\prime} s$ and their variance in terms of the actual sizes of the smaller and larger classes. These are the purposes of the regression analyses. In the remainder of this section, such regression analyses are reported for the entire data set and for the data set stratified on several important characteristics of the studies (e.g., age of pupils, validity of the study).

1. Regression Analysis for Entire Data Set.

The model $\Delta_{S-L}=\beta_{0}+\beta_{1} S+\beta_{2} S^{2}+\beta_{3}(L-S)+\varepsilon$ was fit by least-squares for the 725 points. The results were as follows:


Correlations

|  | $\underline{S}$ | $\underline{S^{2}}$ | $\underline{L-S}$ | $\underline{\Delta}$ |
| :---: | :---: | :---: | :---: | :---: |
| $S$ | 1 | .932 | .004 | -.271 |
| $S^{2}$ |  | 1 | .011 | -.135 |
| L-S |  |  | 1 | .047 |

Regression Analysis
Multiple $R=.426$
Source of Variation Nf MS

$$
\begin{array}{lll} 
& \begin{array}{l}
\text { Regression } \\
\text { Residual }
\end{array} & \begin{array}{l}
3 \\
\hat{\beta}_{0}=.
\end{array} \\
\hat{\beta}_{0} & 6.684 \\
\hat{\beta}_{1}=-.03860
\end{array} \quad \hat{\beta}_{2}=.00059 \quad \hat{\beta}_{3}=.00082
$$

The regression equation for estimating $\Delta_{S-L}$ is

$$
\hat{\Delta}_{S-L}=.57072-.03860 S+.00059 S^{2}+.00082(L-S)
$$

Based on the entire data set, the following table of standardized comparisons for selected class-sizes can be constructed:

| Sma11 Class Size | Large Class Size | Standardized <br> Differential <br> Achievement, $\Delta_{\text {S-L }}$ |
| :---: | :---: | :---: |
| 1 | 40 | .565 |
| 10 | 40 | .268 |
| 20 | 40 | .051 |
| 30 | 40 | -.048 |
|  |  |  |
| 1 | 25 | .552 |
| 5 | 25 | .409 |
| 10 | 25 | .256 |
| 15 | 25 | .133 |
| 20 | 25 | .039 |

These data show that the difference in achievement between class-size 1 , i.e., individual instruction, and cilass-size 40 is more than one-half standard deviation. The difference between class-size 20 and class-size 40 is only about five hundredths standard deviation. Class-size differences at the low end of the scale have quite important effects on achievement;'differences at the high end have little effect.

The curved regression surface, can be reduced to a single line curve in a plane by imposing the consistency condition and solving for the pivot points. The two pivot points are the solutions to

$$
.57072-.03860(P)+.00059\left(P^{2}\right)=0
$$

In this instance, the pivot points equal approximately 43 and 23. The lower value, 23 , was selected as the pivot point around which to coristruct the connected curve; the choice was arbitrary and calculations not reported here revealed it to be largely immaterial. The values of $\Delta_{S-P}$ and. $\Delta_{P_{-L}}$ are as follows for $P=23$ :

$$
\begin{aligned}
& \Delta_{1-23}=.551 \\
& \Delta_{2-23}=.513 \\
& \Delta_{5-23}=.407 \\
& \Delta_{10-23}=.254 \\
& \Delta_{20-23}=.037 \\
& \Delta_{23-30}=.001 \\
& \Delta_{23-40}=.009
\end{aligned}
$$

Hence, on this curve the difference between achievement in class-sizes 1 and 40 is. $551+.009=.560$. The curve is presented in Figure $1 .$. The ordinate is represented by a standard score metric; the zero point is arbitrarily fixed at a class-size of 30 .

In Figure 2, the curve in Figure 1 is translated into a metric of percentile ranks on the ordinate by assuming a normal distribution of achievement. There it can be seen that the difference in average performance from class-sizze 1 to class-size 40 is from above the 70 th percentile to just below the 50th. There is nearly à ten percentile rank difference between instructional groups of sizes 10 and 20 pupils.


Figure 1. Consistent regression line for achievement (in 2 -score units) onto class-size.


Figure 2. Consistent regression line for achievement (percentile ranks) onto class-size (all data).
2. Regression Analyses for Sub-sections of the Data.

Regression analyses were performed for many smaller portions of the entire data set in an attempt to determine which characteristics of the studies might mediate the size of the class-size and achievement relationship. More thian a dozen factors were employed in splitting the data base: year of study, subject taught, age of pupils, IQ, type of test, etc. Few of these characteristics were systematizally related to the strength of the class-size and achievement correlation. Among those factors of discrimination that produced virtually identical regression lines were "source of data," "subject taught," "duration of instruction," "pupil IQ," and "type of achievement measure." From among these few characteristics that appeared to interact with the relationship, three stand out as particularly interesting: year of the study, level of schooling (elementary vs: secondary), and internal validity of the siudy. The complete regression analyses will be reported below for the latter two characteristics. Details of the "year of study" analyses will not be reported here; suffice it to note that there is no correlation between class-size and achievement in those studies carried out before 1940 and a strong relationship favoring smaller classes in post-1960 studies. The two eras differ in many respects, most notably in terms of the sophistication of both experimental design and measurement.

Elementary vs. Secondary. The curvilinear regression model in (2) was fit separately for pupils of age 11 years or younger (elementary) and 12 years or older (secondary). The summary statistics and solutions are as follows:


Some particularly interesting values of $\Delta$ on the two regression surfaces are listed below:
Smaller Class Size
1
10
20
30
Larger Class Size
40
40
40
40

Differential Achievement Elementary Secondary
.490 : . 749
.241 . 357
.063 . 057
-.011 -. 102
1
10
. 387
.716
$\begin{array}{ll}3 & 10 \\ 5 & 10\end{array}$
.324 . 619
10 . 265 . 527
The class-size and achievement relationship seems consistently stronger in the secondary grades than in the elementary grades. This interaction is also seen in Figure 3 where the consistent curves are drawn around pivot points of 19 for elementary and 22 for secondary. The ordinate scale in Figure 3. is percentile ranks.

Well-Controlled vs. Poorly-Controlled Studies. The comparisons were distinguished on the basis of degree of experimental control exercised in the study: Although many features of experimental control could have been noted and analyzed, the method of assignment of pupils to classes of different sizes proved to be the most important. Over one hundred $\Delta$ 's came from/studies in which pupils. were assigned at random to larger and smaller classes; over three hundred comparisons were "uncontrolled," ie., naturally constituted larger and smaller classes were compared. The summary statistics and solutions of the regression models are as follows:


Finure 3. Consistent: Regression Lines for the Recression of Achievement (Expressed in Percentile Ranks) onto Class-sizize for Elementrary and Secundary Grades.


The pivot points for the consistent regression curves are 17 and 48 for the poorly-controlled studies and 17 and 32 for the well-controlled studies. These curves calculated around class-size 17 appear in Figure 4 where the ordinate is expressed in percentile ranks.

The curves in Figure 4 show large differences in the class-size and à̀chievement relationship depending on whether pupil assignment was random or uncontrolled. This finding contrasts sharply with similar analyses of the association between experimental design quality and effects in the field of psychotherapy (Smith and Glass, 1977). The difference is probably due to the magnitude of the effects that are the object of the research in the two fields. The typical psychotherapy effect (therapy vs. control group) is between three-quarters and a full standard deviation (Smith, Glass and Miller, 1979); the typical class-size study was seeking to establish an effect of less than one-tenth standard deviation. It is little surprise, then, that in one field experimental design quality proves critical; and in another field it does not.

In an area of research where the quality of methodology interacts with the findings of studies, the results of the best designed studies should be given more weight in drawing conclusions. The curve for the well-controlled studies in Figure 4 , then, is probably the best representation of the class-size and 8 achievement relationship.

Concern was expressed by several persons who examined the preliminary analyses that the curve for the well-controlled studies in Figure 4 might depend excessively on the twenty or thirty comparisons of very small class-sizes (one and two up to five, say) in the data base. When all those comparisons for which

$\underline{S}=1$ were removed, the curve in Figure 4 for well-controlled studies was even steeper than that shown; this finding is contrary to the claim that tutoring studies skewed the curve unnaturally. When all comparisons for which $\underline{S}$ was less than 6 were removed, the curve for well-controlled studies became less steep; however, it still rose from the 50 th percentile at size 40 to the b0th at size 10 , the 67 th at size 5 and the 74 th at size 1 .

## Conclùsions

Research on class-size and achievement is a particularly complex body of findings to integrate and understand. The integration of this literature has required more sophisticated analysis than has previously been applied to the problem. The meta-analysis of the research reported here has drawn heavily on precise quantitative description and analysis. A clear and strong relationship: between class-size and achievement has emerged. The relationship seems slightly stronger at the secondary grades than the elementary grades; but it does not differ appreciably across different school subjects, levels of pupil IQ, or several other obvious demographic features of classrooms. The relationship is seen most clearly in well-controlled studies in which pupils wére randomly. assigned to classes of different sizes. Taking all findings of this meta-analysis into account, it is safe to say that between class-sizes of 40 pupils and one pupil lie mare than 30 percentile ranks of achievement. The difference in
shievsicent resulting from instruction in groups of 20 pupils and groups of 10 can so larger than 10 percentile ranks in the central regions of the distribution. There is little doubt that, other things equal, more is learned in smaller čayses.

61

## REFERENCES

Burstein, L. The role of levels of analysis in the specification of educational effects. Los Angeles: Graduate School of Education, University of California - Los Angeles, 1978.

Glass, G.V Integrating findings: The meta-analysis of research. Review of Research in Education, 1978, 5, 351-379.

Hartley, S.S. Meta-analysis of the effects of individually paced instruction in mathematics. Ph.D. Thesis, University of Colọrado, 1977.

Lafleur, C.D.; Sumner, R.J.; and Witton, E. Class Size`Survey. Canberra: Australian Government Publishing Service, 1974.

Porwell, P.J. Class size: A summary of research. Arlington, VA: Educational Research Service, Inc., 1978.

Rice, J.M. Educational researcin: A test in arithmetic. The Forum, 1902, 34, 281-297.

Ryan, D.W. and Greenfield, T.B. The Class Size Question. Toronto, Ontario: The Ministry of Education, 1975.

Smith, M.L. and Glass, G.V Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Smith, M.L.; Glazi, G.V; and Miller, T.I. The Benefits of Psychotherapy. Baltimore: The Johns Hopkins University Press, 1979.

## REFERENCES FOR STUDIES INCLUDED IN THE META-ANALYSIS

Anderson, F.H.; Bedford, F.; Clark, V.; and Schipper, J. A report of an experiment at Camelback High Schoo. The Mathematics Teacher, 1963, Vel. 56, pp. 155-159.

Averill, L.A. and Mueller, A.D. Size of' class and reading efficiency. The Elementary School Journal, May 1925;, Vol. 25, 'pp. 682-691.

Below, I.H: A longitudinal evaluation of reading achievement in small classes. Elementary Education, Feb. 1969, Vol. 46, pp ، 184-187.

Bates, D.A. The relation of the size of class to the efficiency of teaching. Master's thesis, University of Chicago, 1928.

Bausell, R.B.; Moody, W.B.; and Walze, F.N. A factorial study of tutoring versus classroom instruction. American Educational Research Journal, 1972, 9, 591-598.

Bostrom, E.A. The effect of class size on critical thinking skills. Ed.D. thesis, Arizona State University, 1969, 69-1276.

Boyer, P.A: Class, size and school progress. Psychological Clinic, 1914, Vol. 8, pp. 82-90.

Breed, F.S. and McCarthy, G.D. Size of class and efficiency of teaching. School and Society, Dec. 23, 1916, Vol. 4, pp. 965-971.

Brown, A.E. The effectiveness of large classes at the college level: An experimental? study involving the size variable and size-procedure variable. University of Iowa Studies in Education, 1932, Vol. 7, pp. 1-66.

Camarosano, JR. and Santopolo, F.A. Teaching efficiency and class size. School and Society, 1958, Vol: 86; pp. 338-340.

Christensen, J.J. The effects of varying class size and teaching procedures on certain levels of student learning. Ph.D. thesis, Wayne State University, 1960, 60-2698.

Clarke, S.C.T. and Richer, S. The effect of class size and teacher qualificalions on achievement. Research Monograph \#5. Edmonton, Alberta: Alberta Teachers Association, 1953.

Coleman, J.S., et al. Equality of Educational Opportunity. Washington: U.S. Government Printing Office, 1966.

Cook, J.J. and Blessing, K.R. Class size and teacher aides as factors in the achievement of the educable mentally retarded. Madison: Wisconsin State Department of Public Instruction, 1970, ED 047484.

Cornman, 0.P. Size of classes and school progress. The Psychological Clinic, Dec. 15, 1909, Vol. 3, pp. 206-212.

Cram, B.M. An investigation of the influence of class size upon academic attain-, ment and student satisfaction. Ed.D. thesis, Arizona State University, 1968, 131. pp., 29/04-A.p. 1066, 68-14988.

Davis, C.0. The size of classes and the teaching load in the high schools accredited by the North Central Association. School Review, Vol. 31, June 1923, pp. 412-429.

Davis, E. and Goldizen, M. A study of class size in junior high school history, The School Review, May 1930, Vol. 38, pp. 360-367.

Dawe, H.C. The influence of size of kindergarten group upon performance. Child
Development, $1934,5,295-303$.
DeCecco, J.P. Class size and co-ordinated instruction. British Journal of Educational Psychology, Feb. 1964, Vol. 34, pp. 65-74.
届 Eash, M.J. and Bennet, C.M. The effects of class size on achievemerit and attitudes. American Educational Research Journal, 1964, Vol. 1, pp. 229-239.
Eastburn, L.A. Report of ćlass size investigations in the Phoenix Union High School, 1933-34 to 1935-36. Journal of Educational Research, Oct. 1937, Vol. 31, pp. 107-117.

Edmonson, J.B. and F.U. Mulder. Size of class as a factor in university instruction. Journal of Educational Research, Jan. 1924, Vol. 9, pp. 1-12.

Ellsôn, D.G.; Barbér, L.; Engle, T.L.; and Kampwerth, L. Programed tutoring: A teaching aid and a research tool. Reading Rėsearch Quarterly, 1965, 1, 77-127.

Ellso., D.G.; Harris, P.; and Barber, L. A field test of programed and directed tutoring. Reading Research Quarterly, 1968, 3, 307-367.

Feldhusen, J.F. The effects of small and large group instruction on learning of subject matter, attitudes, and interests. Journal of Psychology, 1963, 55, 357-362.

Flinker, I. Optimum class size: what is the magic number? Clearing House, 1972, 46, No. 8, 471-473.

Flynn, D.L.; Hass, A.E.; Al-Salam, N.A. An evaluation of the cost effectiveness of alternative compensatory reading programs. Vol. III: Cost-effectiveness . analysis. Bethesda, MD: RMC Research Corporation, April 1976.

Frymier, J.R. The effect of class size upon reading achievement in first grade. The Reading Teacher, 1964, 18, 90-93.

Furno, O. and Collins, G.J. Class size and pupil learning. Baltimore City Public Schools, Oct. 1967, ED 025003.

Glaș, G. ${ }^{\text {i }}$ et al. Data analysis of the 1968-69 Survey of Compensatory Education. Boulder: Laboratory of Educational Research, University of Colorado, August 1970.:

Haertter, L.D. An experiment of the efficiency of instruction in large and small classes in plane geometry. Educational Administration and Supervision; Vol. 14, No. 8, Nov. 1928, pp. 580-590.

Harlan, C.L. Size of class as a factor in schoolroom efficiency. Educational Administration and Supervision, 1915, 1, 195-214.

Haskell, S. Some observations on the effects of class size upon pupil achievement in geometrical drawing. Journal of Educational Research, 1964, 58, 27-30.

Holland, B.G. The effect of class size on scholastic acquirement in educational psychology. School and,Society, June 2, 1928, Vol. 27, pp. 668-670.

Hoover, K.H.; Baumann, V.H.; and Shafer, S.M. The influence of class-size variations on cognitive and affective leárning of college freshmen. Journal of Experimental Education, 1970, 38, 39-43.

Horne, K. Optimum class size for intensive language instruction. Modern Language Journal, March 1970, Vol. 54, pp. 189-195.

Husen, T. International Study of Achievement in Mathematics, Vol. II. Stockholm: Almquist \& Wikse11, 1967.

Jeffs, G.A. and Cram, B.M. The influence of class size on academic attainment and student satisfaction. Las Vegas, Nev.: Edward W. Clark High School, - 1968, ED 021252.

Johnson, M. and Scriven, E. Class size and achievement gains in seventh and eight'n grade English and Mathematics. The School Review, 1967, Vol. 75, pp. 300-310.

Judd, C.H. Report of the consultive committee. Bulletin: Department of Secondary School Principals, 1929, 25, 49-61.

Kïrk, J.R. A study of class size, teaching efficiency, and student achievement. Phi Delta Kappan, Aug: 1929, Vol. 12, pp. 59-61.

Little, A.; Mabey, C.; and Russell, J. Do small classes help a pupil? New Society, Od. 1971, Vol. 18, pp. 769-771. Der Rm HM 1 N48.

Lundberg, L.D. Effects of smaller classes. The Nation's Schools, May 1947, Vol. 39, pp. 20-22.

Macomber, F.G. and Siegel, L. A study in large-group teaching procedures. The Educational Record, 1957, 38, 220-229.

Martin, G.M. The effect of class size on the development of several abilities involved in cricical thinking. Ed.D:- thesis, Tempie University, 1969 (7110853).

Mayeste; G.W., et al. A Study of Our Nation's Schools. U.S. Office of Education (undated).

Meredith, V.H.; Johnson, L.M.; and Garcia-Quintana, R.A. South Carolina first Grade Pilot Project 1976-77: The effects of class size on reading and mathematics achievement. Columbia, S.C.: South Carolina Department of Education, January 1978.

Metzner, A.B: and Berry, C. Size of class for mentally retarded children: Training School Bulletin, 1926, Vol. 23, pp. 241-251.

Miller, P.S. A quantitative investigation of the efficiency of instruction in high school Physics. Journal of Educational Research, Feb. 1929, Vol. 19, pp. 119-127.

Moody, W.B.; et al.: The effect of class size on the learaing of mathematics: A parametric study. Journal of Research in Mathematics, 1973, 4, 170-176.
Moss, F.A.; Loman, W.; and Hunt, T. Impersonal measurement of teaching.
$\because$ Educational Record, Vol. 10, No. 1, Jan. 1929, pp. 40-50.
Nachman, M. and Opochinsky, S. The effects of different teaching methods: A methodological study. Journal of Educational Psychoiogy, 1958, 49, 245-249.

Nelson, W.B. An experiment with class size in the teaching of elementary economics. Educationel Record, 1959, Vol. 40, pp. 330-341.

Perry, R.F. A teaching experiment in geography. Journal of Geography, March 1957, Vol. 56, pp. 133-135.

Rivera, L.R.D. The effects of increasing class size on achievement and the reactions of 'students' and faculty toward this practice at the Catholic University of Puerto Rico. Ed.D. thesis, Lehigh University, 1976, 77-10, 706.

Robinson, J.S. A study of the relationship of selected school and teacher characteristics to student performance on the BSCS Comprehensive Finai Examination 1961-62. Boulder, Colo.: BSCS, May 1963.

Rohrer, J.H. Large and small sections in college classes. The Jourina of Higher Education, May 1957, Vol. 28, pp. 275-279.

Ronshausen, N.L. The programed math tutorial -- paraprofessionals. provide one-to-one instruction in primary school mathematics. Washington, D.C.: American Educational Research Association, Annual Meeting. 1975. ERIC No. Ed 106743.

Shaver, J.P. and Nuhn, D. The effectiveness of tutoring under-achievers in reading and writing. Journal of Educational Research, 1971, 65, 107-112.

Siegel, L.; Macomber, F.G.; and Adams, J.F. The effectiveness of large group instruction at the university level: Harvard Educational Review, 1959, 29 216-226.

Silver, A.B. English department, large-ṣmall class study: English 50-60. Revised, July 1970, ED 041586.

Simmons, H.F. Áchievement in intermediate algebra associated with class size at the University of Wichita. College and University; 1959, Vol. 34, pp. 309315.

Smith, D.I. Effects of class size and individualized instruction on the writing of high school juniors. Ph.D. thesis. Florida State University, 1974 (7425461).

Smith, D.V. Class Size in High School English. Minneapolis: The University of Minnesota Press, 1925.

Spitzer, H.F. Class size and pupil achievement in elementary schools. Elementary School Journal, 1954, 55, 82-86.

Stevenson, P.R. Class-size in the Elementary School. Bureau of Educational Research Monographs No. 3. Ohio State University Studies, Vol. 2, No. 10, November 20, 1925.

Summers, A.A. and Wolfe, B.L. Equality of educational opportunity quantified: A production function approach. Philadelphia: Department of Research, Federal Reserve Bank of Philadelphia, 1975.

Tope, R.E.; Groom, E. ; and Beeson, M.F. Size of class and school efficiency. Journal of Educational Research, Feb... 1924, Vol: 9, pp. 126-132.
Verducci, F. Effects of class size upon the learning of a motor skill: Research Quarterly, May 1969, Vol. 40, pp. 391-395.
-
Wasson, W.H. A controlled experiment in the size of classes. Masters thesis, Ur:iversity of Chicago, June 1929.

Weitzman, D.C. Effect of tutoring on performance and motivation ratings in secondary school students. California Journal of Educational Research, 1965, 16, 108-115.

Wetzel, W.A. Teaching technique and size of class. School Life, 1930, 15, 181182.

Whitney, L: and Willy, G.S. Advantages of small classes: School Executives Magazine, Aug. 1932, Vo 1. 51, pp. 504-506.

Wilsberg, M.; Castiglione, L.V.; and Schwartz, S.L. 'A program to strengthen early childhood education in poverty area schools. Educational Research Committee, Nov. 1968', Center for Urban Education.

Woodson, M.S. Effect of class size as measured by an achievement test criterion. IAR Research Bulletin, Feb. 1968, Vol. 8, pp. 1-6.

Wright, F.N., et al. Effects of Class Size in the Junior Grades. Toronto: Ministry of Education, Ontario, 1977.

## APPENDIX

DATA LISTING

The raw data on which the analyses are based are listed on the following pages. The key to decoding the variables appears in Table 3.1 ir the section of the report on Methods. Horizontal lines separate the studies on the first page of the listing only. The variables are numbered on the first page of the data listing. The itities of the variabies correspunding to these numbers are as foliows:

```
    1. ID#
    2. Year
    3. Source
    4. Subject taught
    5. Hours of instruction
    6. Weeks of instruction
    7. N for small classes
    8. No. of teachers for small classes
    9. Class-size (P/I) for small
    10. Accuracy of P/I
    11. N for large cilasses
    12. No. of teachers for large classes
    13. Class-size (P/I) for large
    14. Accuracy of P/I
    15. IQ
    16. Age
    17. Assignment of pupils
    18. Assignment of teachers
    19. Type of achieyement measure
    20. Subjeci of achievement measure
    21. Quantification of outcomes
    22. Congruence of instruction and achievment measure
    23. Delta( S-L )
    24. No. of times S greater than L
    25. No. of times L greater than S
```





 $\dot{i}_{n}$ $\because \quad 73$








|  | 715936 | 180 | 36 |
| :---: | :---: | :---: | :---: |
|  | －715936 | 180 | 36 |
|  | 7159313 | 180 | 36 |
| $\therefore$ | 7159313 | 180 | 36 |
|  | 715939 | 180 | 36 |
|  | 715939 | 150 | 36 |
|  | 7159311 | 180 | 30 |
|  | 71593.11 | 180 | 36 |
|  | 7276319 | 50 | 18 |
|  | 7276319 | 50 | 18 |
|  | 727639 | 50 | 18 |
|  | 727639 | 50 | 18 |
|  | 727639 | 50 | 18 |
|  | 727639 | 50 | 13 |
|  | ．． 7276319 | 50 | 18 |
|  | 7276319 | 50 | 18 |
|  | 727 ¢317 | 50 | 18 |
|  | 7276310 | 50 | 18 |
|  | 7276310 | 50 | 18 |
|  | 7369320 | 40 | 13 |
|  | 7460317 | 50 | 18 |
|  | 752913 | 90 | 13 |
|  | 7心75320 | －0 | 4 |
|  | 7675320 | 20 | ． 4 |
|  | 7675320 | 20 | 4 |
|  | 777434 | 30 | 6 |
|  | $77743^{3}, 4$ | 30 | 6 |
|  | 777434 | 30 | 6 |
| $\leqslant$ | 77743 4 | 30 | 6 |
|  | 777434 | 30 | 6 |
|  | 777434 | 30 | 6 |
| － | 777434 | 30 | 5 |
|  | 777434 | 30 | 1 |
|  | 777434 | 30 | ． 6 |
|  | 777434 | 30 | 6 |
|  | 777434 | $\cdot 30$ | 6 |
|  | 77743 1 | 30 | － 5 |
| － | 777434 | 30 | 6 |
|  | 777434 | 30 | 0 |
|  | 77743 4 | 30 | 6 |
|  | 777434 | 30 | 6 |
|  | 777434 | 30 | 6 |
| － | 777434 | 30 | 6 |


| 380 | 21 | 182 | 840 | 21402216241 | 612 | 07 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 430 | 18 | 243 | 600 | $18 \cdot 343216241$ | 612 | 00 |
| 380 | 21 | 182 | 840 | 21402217241 | 312－ | － 22 |
| 430 | 18 | 243 | 600 | $18 \quad 34321724$ | 312－ | － 55 |
| 380 | 21 | 182 | 840 | 21402217241 | 912 | ． 9 |
| 430 | 18 | 243 | 500 | $1836321724 i$ | 912 | $9 \%$ |
| 380 | 21 | 182 | B40 | 21402217241 | 112 | － 24 |
| 430. | 18 | 243 | 600 | 1834321724 | $1112-$ | － 3 |
| 40 | 1 | 403 | 116 | 11163319222 | 933 | 10 |
| 50 | 1 | 503 | 126 | 11263319222 | 933 | 08 |
| 64 | 1 ： | 643 | 116 | 11163319222 | 933 | 93 |
| 16 | 1 | 163 | 66 | 1553319222 | 933 | 33 |
| 44 | 1 | 443 | 1；9 | 11173319222 | 933 | 02 |
| 28 | 1 | 283 | 43 | $143: 319222$ | 933 | 23 |
| 62 | 1 | 62.3 | 146 | 114 E3319222 | 1933 | 02 |
| 31 | 1 | 313 | 105 | 1105．＂－1922， | 193\％ | － 72 |
| 45 | 1 | 453 | 103 | 11083319222 | 733 | 39 |
| 36 | 1 | 363 | 124 | 11243319222 | 1033－ | － 05 |
| 26 | 1 | 263 | 38 | 1383319222 | 1033 | 02 |
| 45 | 3 | 153 | 45 | 14533181312 | 2032 | 07. |
| 27 | 1 | 273 | 03 | － 5 ここ319232 | 1733－ | － 30 |
| 22 | 1 | 22.3 | 44 | $1443 \% 14242$ | 335－ | － 15 |
| 28 | 1 | 282 | 38 | 1 ЗEこ21743120 | 2023 | 35 |
| 23 | 1 | 282 | 58 | 1．592こ 17431 | 2023 | $02^{\circ}$ |
| 18 | 1 | 332 | 58 | 1562217431 | 2023－ | 34 |
| 84 |  | 13 | Q4 | 61431.16121. | 432 | 55 |
| 84 |  | 13 | 84 | 6143216131 | 432 | 133 |
| 84 |  | 13 | 84 | 6143316131 | 432－ | 15 |
| 84 |  | 13 | 8.4 | 6143115132 | ¢33 1 | 156 |
| 84 | － | 13 | $8{ }^{4}$ | 6 1432：61．32 | 43こ | 78 |
| B4 |  | 13 | 84 | 6．143315：32 | 433 | 27 |
| 94 |  | 13 | 1.20 | 4 302i1613： | 432 | 91 |
| 84 |  | 13 | 120 | 430.321 .6151 | 4321 | 105 |
| 84 |  | 13 | 120 | 4303316131. | 432 | 27 |
| $\therefore 8$ |  | 13 | 120 | $4303: 1613:$ | 433 | 151 |
| 84 |  | 13 | 120 | 4303216132 | 433 | 85 |
| 84 |  | 13 | 120 | $4303316: 32$ | 433 | 69 |
| 84 | 6 | 143 | 120 | 4305116131 | 432 | 36 |
| 84 | 6 | 143 | 120 | $4.30 こ 215131$ | 432－ | 2 B |
| 84 | 6 | 143 | 120 | 4 2゙331613i | 432 | 42 |
| 84. | 6 | 143 | 120 | 4 303116132 | 433－ | 05 |
| $84^{\circ}$ | $E$ | 143 | 120 | 4303216132 | 433 | 07 |
| B4 | 6 | 143 | 120 | 4303316132 | 433 | 22 |



## $\mathcal{F}$


[^0]:    

    - Reproductions supplied by EDRS are the best that cer be made. *
    from the original document.

