The Class Size/Achievement Issue: New Evidence and a Research Plan

by Leonard S. Cahen and Nikola N. Filby

Using "meta-analysis," Gene Glass and Mary Lee Smith have discovered important student achievement gains when class size is reduced to 15 or below. Cahen and Filby are now involved in intensive field study of the whys and hows of these gains.

On the average, student achievement increases as class size is reduced, and the advantage rises sharply for a class of 15 and below. Reductions in size of from, say, 28 to 25, are projected to make only a small difference in average achievement. These are perhaps the most significant conclusions reached in a new "meta-analysis" of half a century of research, performed as part of a project in class size and instruction being conducted by the Far West Laboratory for Educational Research and Development with National Institute of Education funds.

Gene Glass and Mary Lee Smith of the University of Colorado were responsible for the meta-analysis. At the same time, a complementary and converging approach to the question of class size/achievement relationships was undertaken and is continuing. A small number of field studies were designed in which class size is experimentally controlled and intensive observation of classroom procedures is being conducted. A chief object is to find out what aspects of instruction in smaller classes account for the achievement advantages. The remainder of this article will detail the procedures, findings, conclusions, and policy implications of the Far West project.

For the research synthesis, we felt that the new approach called meta-analysis would prove to be a powerful way of resolving some of the inconclusive findings reported in the literature. Glass, a primary developer of meta-analysis methodology, reported that the class size/student achievement literature might lend itself to the technique.

Meta-analysis provides a method for the statistical integration of data across many studies. Studies of psychotherapy and tutoring, among other fields, have already been integrated via meta-analysis. Meta-analysis proceeds by calculating the size of one or more measures of effect in each study, then pools these measures as data points for further analysis. In the case of class size studies, each data point is a measure of the difference in achievement between two classes of different size.

Glass and Smith first obtained and read some 300 reports, publications, theses, etc., that reported findings on class size and achievement. The search was made through ERIC, dissertation abstracts, research reports and reviews, and from nominations and suggestions from other researchers. Glass and Smith found current reviews by Doris Ryan and T. Barr Greenfield and C. D. Lafleur, R. J. Sumner, and E. Witton to be very helpful. Only 77 of the 300 documents could be used. They yielded 725 comparisons of achievement in different class sizes. Many studies yielded multiple sets of data. For example, one might report achievement data for reading, mathematics, and science for three grade levels, thus yielding nine comparisons. The studies provided a data set based on nearly 900,000 pupils and spanned, over half a century. Sixty-five percent of the comparisons were obtained from journals, approximately 16% from books, and 11% from unpublished sources. Approximately 8% came from theses, a source not generally tapped in prior examinations of the literature. Approximately 56% of the comparisons were obtained on children whose ages ranged from 5 to approximately 11 years.

As expected most of the studies compared class size in the range of 20 or larger. Comparisons of classes of about 26 pupils with classes of more than 30 were common, 10 with 20 far less so. For many years researchers expected to see dramatic differences between class sizes of 25 and 28.

Glass and Smith define class size as the pupil-to-instructor ratio (P/I). One teacher with 30 pupils gives a P/I of 30, two with 30 a P/I of 15. One teacher doing supplementary math instruction with four pupils gives a P/I of four. The search for an appropriate descriptive ratio has a long history in the research on class size. Any ratio is, at best, a crude indicator of how much teacher attention any pupil receives. One hopes that as the total number of pupils in a class decreases, the teacher will be able to provide more appropriate, personal instruction for every pupil. How to help teachers take advantage of reduction in total class size becomes a crucial issue, to be discussed later.

Glass and Smith define "delta" as a key concept. A statistical index of the achievement advantage of one size class over another size class, delta is defined as the mean achievement score for the smaller class in a study minus the mean of the larger class in the study, the difference then being divided by the within-group
achievement test. The Glass-Smith curve and

2) square of the number of students in the
design characteristics - e.g., where pupils
on the quality of the research design. Ef-
fctions are equal,
given and when the standard devia-
toward when means and standard deviations
69th percentile.

The calculation of delta is straightforward when means and standard deviations are given and when the standard deviations are equal, but these conditions are not always met. Glass worked out formulas for estimating delta from other common statistics, such as a correlation coefficient. Problems can arise in defining the within-group standard deviation when the groups differ widely in variability. In this case the estimate of delta may be biased. Continued work on the methodology of meta-analysis, as developed by Glass, will need to study the effects of heterogeneous variability on the magnitude of deltas and the relationship of the deltas to other variables in the studies being examined.

Of the 725 deltas calculated, 60% were positive, indicating that achievement was higher in the smaller class. The average delta was .09. Further analysis revealed two important interactions: The size of the difference depended on the sizes of the classes being compared; it also depended on the quality of the research design. Effects were stronger in studies having good design characteristics – e.g., where pupils were randomly assigned to classes or were “matched,” or where the same teachers or pupils participated in both the smaller and the larger class. The average delta in well-designed studies was .40.

To take into account the influence of different class sizes, a regression model was developed to predict delta. After preliminary models were tried, the final system predicted delta (advantage of smaller over larger) from three variables: 1) number of students in the smaller class; 2) square of the number of students in the smaller class; 3) difference between the number of students in the smaller class and the number of students in the larger class. The regression model was used to generate a graph of predicted achievement. Predicted achievement scores were transformed to a percentile rank on a hypothetical nationally normed standard achievement test. The Glass-Smith curve for well-designed studies is shown in Figure 1.

In this figure the curve starts to rise most dramatically when class size is reduced below 15 pupils. The average pupil in class sizes of 40, 20, 15, 10, and five would be expected to score at the 50th, 55th, 58th, 65th and 75th percentiles respectively. The predicted outcome difference can be described in grade-equivalent units over one school year: 1.00 years of growth for class size 40, 1.15 years for size 20, 1.24 years for size 15, 1.45 years for size 10, and 1.72 for size five. These data show an impressively large advantage for smaller classes.

The overall difference in results between the well-controlled and poorly controlled studies was dramatic. The curve for the poorly designed studies was almost flat, indicating, at best, a very small advantage to smaller classes. Almost half of the deltas came from the poorly designed studies. Little wonder class size research has been so inconclusive.

Glass and Smith analyzed data separately for elementary and secondary pupils. Small-class advantages were slightly stronger at the secondary level. It is our opinion that the advantages are too small to lead to a conclusion that elementary pupils would profit less than secondary pupils if class size were reduced. There also appeared to be no difference in results for different subject matters, such as reading or mathematics.

The meta-analysis reports that there is no correlation between class size and achievement advantage in the studies performed before 1940. Over half the 725 deltas were from pre-1940 studies. It is not surprising that surveys of the literature prior to World War II typically concluded that reducing class size had no effect on achievement.

Over the next few months we plan to fit the Glass-Smith equation to data not analyzed when the model was developed. This will allow us to estimate the error in the model for different class size comparisons. We anticipate that new studies will be identified, but these, like those available to Glass and Smith, will contain few data for class size smaller than 20, the range we believe to be crucial.

How does one judge the importance of the differences shown in Figure 1? Is the percentile advantage in achievement between class sizes of 15 and 30 big enough to make it worthwhile to reduce size by this much? Policy makers will have to decide. As researchers, we encourage the concept of utility. We regard the delta difference between class size 30 and 25 as relatively trivial. But the difference between class size 30 and 15 has utility. Enough pupils should profit to warrant pursuing ways of creating the smaller class. We acknowledge the economic difficulty of putting this judgment into practice. But we encourage investigations of reduced instructional group size for parts of the school day. More on this later.

A cautious comment about the small changes in achievement above a class size of 20: Achievement tests measure only one aspect of instruction. They do not capture the quality or humanness of the classroom environment. Certainly larger classes permit less relaxed interaction with individual pupils. Teachers often feel overwhelmed and frustrated.

It is also important to point out that the Glass-Smith meta-analysis shows the relationship of class size and achievement without any attempt to see how this relationship is conditioned by a set of variables we shall call quality of instruction. It would be useful to find out whether, and how, good and poor teaching or environmental conditions alter the curve.

The Field Studies

The Glass-Smith meta-analysis...
The field studies may show how to achieve even greater gains in small classes.

increases as class size decreases. If this is true, it must be because of some change in classroom instruction. With fewer pupils to attend to, a teacher should be able to improve the quantity and/or quality of instruction.

P. J. Ponvoll and others have concluded that many qualities of classroom instruction, such as increased individualization, are improved when class size is reduced. Glass and Smith are now doing a meta-analysis of studies relating class size to classroom processes, student attitude, and teacher satisfaction. Teacher satisfaction is an important outcome to consider in its own right. It appears from the literature search that relatively few studies have systematically examined the question of how class size influences student achievement. The field studies undertaken by the Class Size and Instruction Project address this question.

The basic plan of the current field studies is to reduce class size experimentally and see what changes take place in the classroom. In each of two schools, we work with two second-grade classes, each taught by a single teacher. Midway through the year a third teacher is hired and some students from each class are moved to the new third class. Many methodologies are used to learn about the nature of schooling in the larger class situation (before the split), and this can be compared with what we learn when the classes are made smaller after the split.

An important aspect of the field studies is the role of the classroom teachers. We hope to make them collaborators in the investigation of an important educational question. As a research team, we shall form hypotheses about what might be different in a smaller class, and we shall collect evidence about what actually changes. The teachers are encouraged to "tinker," i.e., try new techniques. This means that the field studies are not a "clean" experimental test of class size but are instead a combination of class size experiment and in-service training for teachers. It is exactly this combination that we consider it important to study. Many people have suggested that reducing class size will have no effect if teachers do exactly the same thing in a small class as in a large one and that it is important to help teachers take advantage of the opportunity of a small class.

In the field studies we work with the teachers to find out what it is possible to do in a small class. If we could consistently improve the quality of instruction as we reduce class size, then the increases in achievement should be even greater than those shown in the Glass-Smith curve.

A major source of our perspective in describing classroom instruction is our previous work on the Beginning Teacher Evaluation Study (BTES). In our current work we hope to elaborate and extend the BTES model of instruction, thus building a cumulative research program. BTES researchers, working with second- and fifth-grade classes, looked at a series of questions about pupil learning in mathematics and reading and how this learning was related to teaching behaviors and characteristics of classroom learning environments. The BTES study convinced us that the teacher controls learning conditions that are positively associated with pupil learning. For example, larger pupil achievement gains, were associated with teacher monitoring of pupil behavior, the teacher's ability to diagnose pupil status and prescribe appropriate educational tasks (quantity and quality), and teacher feedback. Classes with larger gains were typically associated with teachers who held academic goals for their pupils and provided relatively large amounts of direct instruction. It was also observed that the teaching/learning environments in these classes were supportive. Teachers did not have to be punitive in order to have children learn. As we began to design our plan for the Class Size and Instruction Project, we wondered how the learning environments in classes could be changed if we reduced the number of pupils for whom the teacher had responsibility. If class size could be reduced by one-third or one-fourth, would the teacher be able to provide a more individualized form of instruction? Would the teacher be able to diagnose pupil needs better, assign more appropriate work, and monitor the work more frequently? Would pupils' "wait time" (waiting for teacher direction or help) be reduced? Would pupil/pupil and pupil/teacher interaction change?

How would teachers feel about teaching and their pupils when class size was reduced? Would there be more time for informal discussions with pupils? Would there be changes in the curriculum or learning activities such as more and different types of art or science lessons? Would pupils now be allowed to talk to each other as they worked?

The following categories of questions provide a framework for our inquiry: instruction, pupil/pupil interactions, pupil/teacher interactions, teacher planning, classroom environment, rule setting, interruptions and disruptions, diagnosis, assessment and pupil evaluation, teacher feedback, reward systems, teacher expectations, prior to Splitting, and teacher evaluation of conditions before and after splitting.

Two schools are participating in the study. One is a rural school near Charlottesville, Virginia, directed by Gail Me Cutcheon of the University of Virginia. Pupils are primarily low socioeconomic level blacks (60%). Before they were split in January, 1979, each class had about 19 pupils. Splitting reduced the classes to approximately 13 students each. Parent volunteers assist the teachers.

The second school is located in Oakland, California. Both second-grade classes prior to splitting were composed of 34 students, so size dropped to approximately 23 students per class after splitting in February, 1979. Classroom aides are used. There is a staggered reading schedule, meaning that half the students in a class come for an hour in the morning and the other half remain at the end of the day for their smaller group instruction in reading. Nikola Filby, one of the authors, teaches the class created by the split.

Methods of Data Collection

The central activity in the field study will be to document and describe differences in instruction before and after splitting. Research on teaching today is multidisciplinary and uses many approaches to knowing. Some researchers advocate the experimental method as the most powerful way of detecting teaching/learning relationships. Others feel that understanding can best be attained by spending many hours in classrooms watching the process, talking to teachers, etc. Many researchers like ourselves think it is wise to combine many methods: We observe and record what we see, we measure some dimensions, we ask our teachers to help us understand what we see. Our methodology includes both qualitative and quantitative approaches. The Oakland and Virginia researchers have developed descriptions of different approaches to inquiry being used in the study under the following headings: 1) "case study" observation, 2) interviews with teachers, 3) systematic, quantitative observation, 4) teacher journals, 5) achievement testing, 6) samples of student work, 7) photographs, and 8) later follow up.

Reporting the Findings

The detailed case studies of each class will be a major form for reporting our study findings. The case studies will document any changes between instruction in the large-class phase and the small-class phase. We hope to discover whether changes in instruction are a function of reduced class size. The case studies will also address more general questions about important characteristics of classroom in-
Issues and Policies

To date, major reviews of the literature on class size have reported conflicting findings in the research. Some studies supported smaller class sizes; others did not. Reviewers generally found the literature complex and inconclusive. Some reviewers became pessimistic about the value of smaller classes. The Glass-Smith meta-analysis, is unique because it reveals general trends. Previous reviews and the conclusions drawn from them were primarily reached from an "armchair" synthesis of the literature. Studies were classified as supportive of smaller class size, larger class size or inconclusive. The classifications were guided by the statistical significance reported. No evaluation was given in the counting procedure to studies nearing conventional levels of significance. For example, studies showing probability levels greater than .05 would typically be classified as nonsignificant and thus be placed in the inconclusive category. In contrast, Glass and Smith used all the available data to develop a continuous distribution of effects and therefore move their analysis beyond the nominal classification of supportive (favoring smaller classes), nonsupportive (favoring larger classes), and inconclusive (failure to reject the null hypothesis). We feel that the new findings by glass and Smith present a convincing case that average achievement increases as class size decreases, especially when class size is below 20 pupils per class. Earlier arguments that smaller classes cannot be justified on the basis of test scores must be reexamined in light of the Glass-Smith findings.

We must point out, however, that there are many exceptions to the general end. Smaller is not always better. Previous reviews of the literature have done a commendable job of describing the limitations of past studies of class size and explaining how research in the area must depict the problem as interactive—a fraction of pupil characteristics, teachers quality of teaching, subject matter taught, etc. Their reviews have also pointed out the need for understanding the complexities listed above and their interaction with different outcome measures: achievement, classroom processes, teacher morale, and pupil affect. The Glass-Smith analyses did not find any general interactions in the data; that is, class size effects were not noticeably different for children of different ages or abilities or studying different subjects. But there were 'many instances in the data where small classes did not produce superior achievement. Two possible explanations are the nature of the teaching that takes place and imprecision in the construct "class size."

As discussed earlier, a number of people have pointed out that the effect of class size 'depends on the intervening classroom 'instruction. Poor teaching will not be effective, even in a small class; Teachers may need help in learning to use the potential available in the small-class situation. We are exploring this issue in the field studies. Certainly anyone who plans to reduce class size should plan also to support and educate personnel to realize the potential.

From discussions of class size in the literature, it is clear that better designs are needed if we are to understand the complexities of instruction and how these complexities are influenced by the sometimes poorly defined global term "class size." Donald Pidgeon has described other characteristics of students and classrooms that influence the size of the job facing teachers. He mentions homogeneity of pupils, classroom space available, and ancillary assistance available in the classroom. The concept of teacher load is discussed in the literature. While the term is usually used to describe the teaching responsibilities of secondary teachers, it applies to the elementary school as well. A teacher who has responsibility for grading essays probably has a different out-of-school workload than a shop teacher. The teacher who has many students learning English as their second language adds additional teaching burdens. The problems created by disruptive students must be reckoned with in assessing teaching load/ responsibilities. The Class Size Committee of the Lodi (California) Education Association has attempted to weight factors in the classroom—i.e., number of slow learners, hyperactive pupils, bilingual pupils, etc.) in adjusting class size so that it better reflects the range of teaching responsibilities. All of these issues create complications in simplified indices such as class size.

In the end, one must face the central question: If smaller is generally better, is it generally worthwhile to make the change? At this period in our history we have many unemployed teachers. We also have a financial and political climate that resists spending. In the late spring of 1979, in collaboration with Gene Glass and Mary Lee Smith, we shall commission reaction papers to the meta-analysis on class size and achievement and the second meta-analysis dealing with the relationship of class size and classroom processes, teacher satisfaction, and pupil affect. Within our funding restriction we shall seek reaction papers from teachers, administrators, economists, and researchers. These papers will serve to clarify and highlight the different viewpoints on class size and the trade-offs that must be made. In the end, individual states, communities, or parents must make their own value judgments.

We would hope that in discussions of class size many different alternatives will be considered. The data suggest that there is relatively little pay-off for small overall reductions (e.g., 28 to 25). Attention should be given to ways to make larger reductions in more limited situations. Flexible arrangements within a school might allow the creation of smaller instructional groups for part of the school day or for those students most in need of closer supervision or individual attention. Some school districts use a staggered schedule so that students spend part of the-day in a smaller class. Paraprofessionals can help. The use of nonprofessional instructional staff (aides, parent volunteers, and pupil tutors) deserves careful attention. R. G. Stennett, A. L. Hyer and Robert M. McClure, and Beatrice A. Ward and William J. Tikunoff have discussed issues relating to the use of noncertificated personnel in classroom instruction.

We would also hope that schools examine ways to rehire some of the many talented teachers who have lost their positions or cannot find teaching positions. We share the positions of John Corbally and Herbert Walberg and Sue Pinzur Rahe that the large number of unemployed teachers should be viewed as an underutilization of talent, not as a surplus. It is interesting to ponder what instruction in schools could be with two professionals teaching 30 pupils, at least for reading and mathematics in the primary grades.

We are concerned that the Smith-Glass curve may be interpreted by "budget at any cost" school administrators and citizens to mean that class size can be increased beyond 30 pupils without achievement deficit or other consequences. We should emphasize that the present findings consider only student achievement as an outcome. Glass and Smith are presently completing a second meta-analysis for our project. This analysis will

(Continued on page 538)
Class Size/Achievement

(Continued from page 495)

examine the relationship of class size, instructional processes, teacher morale, and pupil affect. The studies will include pieces of research that previously have not been integrated into most of the literature reviews. Their report is scheduled for publication later this spring. We can anticipate that this further analysis and the field studies will give a richer picture of the benefits of smaller classes. Certainly many teachers are convinced of the need for smaller classes. NEA President John Ryor has said that wages and class size were primary strike issues in 1978-79. In November, 1978, half of the Fresno, California, public school teachers struck in a dispute over class size. The school board had rejected the Fresno Teachers Association proposal to add an aide in elementary school classes with more than 33 students.

We need to consider a broad range of outcomes — the relationship between class size and the quality and humanness of the nation’s schools. These concerns may make even small changes in class size worthwhile and may increase the impetus to find ways to create some small classes. We encourage educators and the public to think seriously about what we want our schools to be and how smaller classes might help make that image a reality.

6. Porwoll, op. cit.
7. Ibid. See also B. McKenna, “Measures of Class Size and Numerical Staff Adequacy Related to a Measure of School Quality” (Doctoral dissertation, Teachers College, Columbia University, 1955); Pugh, op. cit.; and Donald H. Ross and Bernard McKenna, Class Size: The Multi-Million Dollar Question (New York: Teachers College, Columbia University, 1957).