



PERGAMON

International Journal of
Educational Research 29 (1998) 691–710

International Journal of
**Educational
Research**

Chapter 1

Research on class size effects: a critique of methods and a way forward

Peter Blatchford*, Harvey Goldstein, Peter Mortimore

*Department of Psychology and Special Needs, Institute of Education, University of London, 25 Woburn Square,
London WC1 OAA, UK*

Abstract

Research findings on class size differences need to be examined carefully because of their far reaching pedagogical and resource implications. An examination of research on the possible causal effect of class size differences on pupil progress, and on possible mediating factors, also raises important issues for educational research in general. In this chapter critically different approaches to research on class size effects are reviewed and an approach used in current research at the Institute of Education, that seems most likely to inform this controversial topic is described. Researchers need to pay more attention to providing a more reliable and valid measure of class size itself. In keeping with the bulk of research in this area the emphasis is on quantitative approaches, although possible limitations and alternative research approaches are considered. Limitations of earlier cross sectional research are examined, and the well-known STAR research is used to highlight some limitations of experimental designs. Finally, the advantages of longitudinal research, using multi-level modelling techniques, are examined.

© 1998 Elsevier Science Ltd. All rights reserved.

In the UK, as in many other countries, the educational effect of class size in schools is an important and often controversial topic. A main reason for disagreement is the conflict between, on the one hand, the common assumption among teachers, parents, and governors of schools (Bennett, 1996) that smaller classes provide a more productive educational environment than larger classes, and, on the other hand, the unwillingness of Government agencies and other policy makers, as well as some researchers (e.g., Slavin, 1989), to agree that class size in itself is a main determinant of educational progress. A main reason for caution is that any commitment to smaller classes necessarily involves more teachers, and this has considerable resource implications.

* Corresponding author. E-mail: P.Blatchford@ioe.ac.uk.

In this context the research evidence on class size effects has become a main source of material to support or refute arguments about policy. In the UK there has been a dearth of research that can provide information on the effects of class size. There can be few areas where research findings are so out of step with professional and lay opinion. Why is it that the research evidence on the educational effects of class size has been so inconclusive? Is it that the research has been inadequate? Or have researchers not asked the right questions or looked at the right outcomes? Or perhaps class size is not as important as is often assumed. Perhaps, as the UK Government and the head of the government agency OFSTED have maintained, other factors — particularly the teacher — are more important.

The recent STAR research in Tennessee has done much to bring back attention to research evidence on class size effects. It has provided evidence that small classes do have benefits for young children in schools, and this has influenced policy in the US and other countries. But, as discussed below, the STAR research also raises important questions about the validity and generalizability of results from experimental research designs.

More research is needed in the UK on the effects of class size. There are at least three reasons for this: first, because of the importance of class size in educational planning and resourcing; second, because of the likely importance of class size in the experience of teachers and pupils; and third, because of the selective use of research evidence or the lack of it in order to service particular views and policy on class sizes. At the same time, however, utmost care concerning the use of appropriate research methods is required. The generally inconclusive results from research may be in part attributable to the inadequate research designs and measures that have been employed. There is a need to consider carefully the measures that have been used to represent size of class or pupil–teacher ratios, the methods used to establish causal links between class size and educational outcomes, and the methods used to study possible mediating processes. Important issues for research on school progress more generally are raised.

This chapter looks critically at measures used to describe class size and pupil–teacher ratios, and at the available official statistics. The focus is on changes over time in the UK and also international comparisons. Research methods used to study the causal effect of class size on educational outcomes are reviewed. Research findings are not reviewed in depth (this has been done elsewhere: see Blatchford and Mortimore, 1994; see also NAHT, 1996, Slavin, 1989); rather, the emphasis is on the methods used. Limitations of earlier correlational research,¹ which are now well understood, are pointed out. Additionally, difficulties with experimental research which are perhaps less well understood are discussed. Finally some difficulties in examining mediating processes that might explain associations between size of class

¹ The term correlational research is used here to describe methods that look at relationships between measures of class sizes, characteristics of pupils, teachers, and schools as they exist, without experimental manipulations. The term correlational is interchangeable with observational, a term preferred in the statistical literature - see Goldstein and Blatchford, 1997.

and educational outcomes are pointed out. In keeping with the bulk of research in this area quantitative research is emphasized. In the final section the limits of such research are assessed and a possible integration with alternative more qualitative approaches is offered.

1. Measuring class size

In order to study the effect of class sizes it is first important that reliable measures of class size be used. Although this may appear straightforward, in practice there are a number of complications. In the past there has been a tendency to use the most obvious and available measures, but these are not always the most reliable or precise. It is possible that this tendency has contributed to the generally inconclusive findings from research.

There are two main indicators. Class size in the countries of the UK is measured on the basis of an annual snapshot of classes taught during an allotted period during one school day — in England in January. Information is obtained on the number of children taught in classes with one teacher and with more than one teacher. Pupil–teacher ratios (PTRs) are calculated by dividing the full-time equivalent pupils on a school's roll by the full-time equivalent number of qualified teachers, excluding short term cover and teachers who are absent for a term or more. PTRs are different from class size because they take no account, for example, of non-contact time. All the children on roll are divided by all the teachers in the school. It should not be assumed that teachers entered into the calculation are teaching for all the time. Consequently, the pupil element in the PTR is a smaller figure than in the class size figures. In 1996 in England the average PTR for primary and secondary schools combined was 18.5 (this excluded sixth form colleges), while the equivalent average class size figure was 24.9 (information provided by the DfEE).

There are limitations with existing class size and PTR measures. Class size figures do not take into account the contribution of non-teaching staff, which in the UK appear to have been growing in recent years. (Collection of Government statistics on this, in connection with class size in the UK, has only just started). It might appear more realistic to calculate an adult–child ratio (where adults would include teaching and non-teaching staff) but this would assume that non-teaching staff were equivalent to teaching staff — an assumption that many teachers would challenge. In this regard the STAR research found no difference in pupil achievement between so-called “regular” classes (22–25 pupils) and “regular” classes with a full-time teacher assistant (Word et al., 1990). Although class size figures are probably more helpful as a guide to what pupils experience in schools, figures on PTRs are commonly given and, for some purposes, class sizes are not available. International comparisons appear to be only available in terms of PTRs, as seen below. In the UK, comparisons between the maintained and the independent (fee-paying) sector are also only available in terms of PTRs.

There are other important limitations with these measures. First, they are taken from a survey of the number of children supposed to be in a class at one point in time.

The extent to which this figure actually matches the everyday experience of pupils and teachers is questionable. The number of children actually in the class at any time may be different from the number according to the class register; children may be away, for example, and the extent of absences may vary from school to school. Moreover, over the course of the school year the number of children may change. This is particularly true of the first year of primary schooling in England — the “reception” class. Entry procedures vary from school to school and between Local Educational Authorities (LEAs). Sometimes all children enter at the beginning of the school year in which they become five years (in September); sometimes entry will be termly so that classes may grow in size over the year, or re-arranged each time; and sometimes the younger children will start on a part-time basis, only gradually building up to full-time attendance. Even once children move beyond the reception year and there is more stability in class sizes, the actual number present at one time can vary between lessons and because some may go out for certain purposes. There is therefore a distinction between the theoretical class size and what might be called the “experienced” class size.

Another limitation of published Government statistics is that they are not broken down in terms of age levels. Figures for 1996 for the first time are broken down in terms of the two main stages within the primary sector (Key Stage (KS) 1 (5–7 years) and Key Stage (KS) 2 (7–11 years). Prior to this, figures referred to the whole primary phase. There are mixed-age classes which make the calculation of age-related class sizes difficult, but if it is true that class size has most effect in the earliest years in school (see below), then it is important that we have more detailed information on class sizes for these years. In the UK, for example, actual reception class sizes are not available even though there is growing disquiet at the consequences of increasing numbers of four year olds entering school and forming large reception classes.

These limitations in published class sizes are not trivial. A measure of class size must be closely tied to a child’s experience if it is to be precise enough to be examined in relation to educational progress. In our current research at the Institute of Education we are seeking a more precise and continuous account of class size differences. This is being done by a termly survey of class sizes. One precise and yet practical measure comes from school registers, which describes attendance during each morning and afternoon session (this would take into account, for example, part-time attendance of the youngest children in school), and is closer, therefore, to the “experienced” class size. Summary measures can therefore be calculated (for example, the extent of variation in size, the average size, and any trend in size over the year) in order to give a figure for each child. These measures are a considerable improvement on a yearly snapshot. In our current research we are also calculating “experienced” pupil–teacher ratios and pupil–adult (teaching and non-teaching staff) ratios in order to provide as full a picture as possible.

Even these more sensitive measures do not take into account the organization of children into groups within a class of a given size. This would require a different level of description, but would be even closer to the experienced classroom environment. In our current research we are conducting a termly survey of group sizes and their composition, and the curricular area and nature of the tasks covered, in order to

examine relations with class size. Is it the case, for example, that in larger classes there tend to be more groups?

2. Official statistics on class size and PTRs

Putting aside for a moment the limitations of class size data, what do the published figures tell us? Here we provide a selective update of figures discussed in previous publications (Mortimore and Blatchford, 1993; Blatchford and Mortimore, 1994), with particular emphasis on international comparisons. UK statistics show that class sizes differ between primary and secondary sectors so that in 1996, for example, the average class size in England at primary was 27.5 and at secondary, 21.9 (here and elsewhere figures used are for all classes whether taught by one or more teacher; the vast majority of classes, according to the annual snapshot, are taught by one teacher — 94% at primary and 96% at secondary in 1996 — and results are similar). Secondary classes therefore have on average about six fewer pupils than primary classes. For the first time in 1996 the primary stage figures are given for KS 1 and KS 2 separately. At KS 1 the average class size was 26.8, which indicates that the youngest children in school are still in much larger classes than at secondary level.

Average primary class sizes over the period 1982–1996 increased by 2.1 pupils (see Table 1). Table 1 also shows the proportions of classes falling into three groups: 1–20, 21–30 and 31 + pupils. The biggest change is in the proportion of pupils in classes with 20 or fewer pupils. The percentage has more than halved from 1982, indicating that many of the smallest classes have been phased out. More than a quarter of classes now have over 30 pupils. In 3% of classes (or 4617 classes), class sizes exceeded 35. In 837 classes, numbers of children exceeded 40.

Changes over time in secondary class sizes are shown in Table 2. Over the period 1982–1996 average class sizes at secondary level changed very little. In contrast to primary schools, there has not been the same increase in the proportion of pupils in large classes of more than 30. There are more than four times as many large classes in primary schools.

Table 1
Primary class sizes (England) from 1982 to 1996

Class size	1982	1987	1990	1994	1996
%1–20	20.4	16.8	13.5	9.5	8.6
%21–30	57.1	62.6	65.9	66.2	64.3
%31 +	22.5	20.6	20.6	24.3	27.1
Av. size	25.4	25.8	26.4	27.2	27.5
Number of classes	154 431	142 793	146 499	150 846	153 955

Note: It includes children under 5 in primary schools. *Source:* DfEE Analytical Services Branch.

Table 2
Secondary class sizes (England) from 1982 to 1996

Class size	1982	1987	1990	1994	1996
%1–20	45.3	46.9	47.2	39.1	36.5
%21–30	45.7	46.5	47.5	55.3	57.3
%31 +	9.0	6.6	5.3	5.5	6.2
Av. Size	21.3	21.0	20.7	21.6	21.9
Number of Classes	169 313	150 133	132 753	123 953	132 716

If official statistics in the UK have a number of limitations, then comparisons between countries are even more difficult. Figures compiled for OECD countries allow approximate comparisons to be drawn but it should be noted that figures are in terms of pupil–teacher ratios. It should not be assumed that if one country has a lower ratio of students to teachers than another country this necessarily means that classes are smaller or that pupils receive more teaching time. The relationship between class sizes, PTRs, and the amount of teaching time is complicated by a number of factors, including variations between countries in the length of the school day, the number of hours a student attends each day, the number of classes or students for which a teacher is responsible, and how a teacher's time is divided into teaching and other duties. Differences also exist in the way that PTRs are collected in each country. These factors mean that interpretation of differences between countries is difficult and conclusions need to be expressed cautiously.

Table 3 summarizes the latest available figures (1994). In most countries the student–teacher ratio decreases with stage of education, so that the older the child the more teachers to pupils. This suggests that the UK is not alone in turning on its head what would appear to be in pedagogical terms the more obvious arrangement — that is, smaller classes for younger children. However, results not shown in Table 3 indicate that while in many countries there has been a significant decrease in student–teacher ratios since 1985, at both primary and secondary stages, this trend appears not to be true of the UK (OECD, 1996).

It can be seen in Table 3 that at primary level the UK comes out badly when compared internationally. In 1994 it had one of the worst PTRs, 21.7:1, compared with the OECD average of 17.5:1. At secondary level the UK was in the middle band with a PTR of 15.2:1, still above the OECD average of 13.8:1. In comparison with other European Union countries, the UK had the second highest pupil–teacher ratios at both primary and secondary (the highest was Ireland).

International comparisons are always difficult and conclusions need to be expressed with caution. Just as with the UK official statistics, it would be helpful if data on class sizes, and more precise estimates, were available. But when taken together the figures described in this section suggest that at primary level the number of children in large classes in the UK are a cause of concern. Class sizes in the UK are increasing, there are many more larger primary classes (over 30 pupils) than at secondary level, and internationally PTRs compare unfavorably.

Table 3
Average number of pupils per teacher (PTRs) (1994)

Phase	20 or more	PTR 15–20	Less than 15
Primary	Turkey	France	Italy
	Ireland	Spain	Denmark
	UK	Australia	Austria
	Germany	Japan	Sweden
	New Zealand	Canada	Belgium
	Mexico	Greece	Portugal
		Czech Rep.	Hungary
		Switzerland	
Secondary	Turkey	Japan	France
	Korea	Ireland	Australia
		UK	Sweden
		New Zealand	Denmark
		France	Austria
		Canada	Italy
		Mexico	Belgium
			Greece
			Spain
			Germany
			Czech Rep.
		Portugal	
		Hungary	

Note: Figures were not available for the US and Norway for 1994.

Source: OECD (1996).

3. Research on the causal effect of class size on educational outcomes

For many years in the UK it has often been remarked that research on the effect of class size differences on educational outcomes has been inconclusive. Historically this led to a period when little research on class size effects was conducted. Perhaps not surprisingly, given the resource implications, this has been seized on by politicians: “I do not believe there is any proven connection between class size and quality of education.” (Eric Forth, Minister of State for Education, reported in *The Independent*, 3.3.93). A past Secretary of State for Education (John Patten) said: “None of it (i.e., research) has shown any link between class size and pupil performance.” And the recent Education Secretary Gillian Shephard argued that class size reductions were not effective in raising educational standards (reported in the *Times Educational Supplement*, 28.2.97).

3.1. Cross sectional correlational studies

Perhaps the most obvious research approach is to look at associations between a measure of class size or PTR, on the one hand, and measures of pupil attainment, on

the other. This approach does not attempt to change or control variables; rather it assesses the situation as found. This was the basic design of earlier research on class size in Britain (see review in Blatchford and Mortimore, 1994). These studies were predominantly cross sectional in design, that is, relationships between variables at a single point in time were examined. Results from these studies, even when large data sets were used, tended to find little or no relationship, or that pupils in larger classes did somewhat better than pupils in smaller classes. Attempts were made to control for other variables that might explain the association, but the advantage of pupils in larger classes was still there.

These results are obviously disconcerting because they seem counter-intuitive. But the now well understood problem with this kind of research, which looks at naturally occurring associations between size of class or PTRs and pupils' performance, is that we often do not know whether the results can be explained by another factor. For example, poor attainers tend to be allocated to smaller classes, more experienced teachers are given larger classes, and less competent teachers are allocated smaller classes. In other words, there may be reasons which could explain the results other than the implausible conclusion that large classes are better for pupils. Studies are also difficult to interpret if they are confined to cross sectional designs. It is better to set up longitudinal studies, within which account is taken of pupil differences at the outset, and therefore the effect of class size on progress can be assessed. More sophisticated types of design and analysis are needed to control for other factors, in order to arrive at more definite conclusions about the effects of class size.

One example of large scale correlational and cross sectional research is a recent report from OFSTED (1995). Given the very public way the conclusions were conveyed in the UK (The Times newspaper interpreted the study as showing that "class size makes little difference"), and the way it has been used by Government ministers to support the view that class size is not important, it is important to look closely at the methods used. This study was based on a large number of inspections by OFSTED inspectors: 594 in secondary schools in 1993/94 and 1173 in primary schools in 1994/5. At the beginning of the report there is a disclaimer that inspectors working for OFSTED were not "academic" researchers and because they were not specifically investigating the effects of class size, the results were therefore somehow more reliable. However, as Bassey (1996) has said, it is not the quantity of data or how disinterested the researcher that counts so much as the relevance of the questions asked and the quality of the research design and methods. In fact the OFSTED research is subject to the same problems of interpretation as other correlational research. In particular the cross sectional design, for reasons given above, makes it impossible to provide a secure interpretation.

Given these limitations what did the research find? Inspectors' gradings of the quality of teaching and learning seen in lessons were used as the response variable, and the percentage of lessons graded 'good or better' was the main measure used. At Key Stage 1 (5–7 years) both the quality of learning and the quality of teaching were rated as better in smaller classes and, as far as we can tell, these results were statistically significant. Also lessons rated as 'less than sound' in terms of teaching and learning were seen less often in small classes. There were no clear differences in favor of

small classes with older children. Therefore, although in the UK much was made of how the results proved that class size had little effect, even in its own terms evidence was found for a link between class size and the quality of teaching and learning at KS 1 — a finding that appears consistent with other studies that support a class size effect. But given concerns about the research method used doubts must exist about the credibility of the results.

3.2. *Experimental studies*

In response to a strong case presented for the benefits of small class sizes with young children, the state of Tennessee funded a large scale intervention project which included over 7000 pupils in 79 schools. It was based at the Center of Excellence for Research in Basic Skills, Tennessee State University. The difficulty with simple correlational research, as has been seen, is that it is difficult to overcome the problem that it could be something about the kinds of pupils (or teachers) in small or large classes which might explain any differences found. Another way of expressing this point is to say that the allocation of pupils and teachers to classes of different sizes may be non-random, and may be biased. This is why the designers of the STAR research decided to employ an experimental design involving the random allocation of pupils and teachers to three types of classes in the same school: “small” classes (13–17), “regular” classes (22–25), and “regular” with full-time teacher aide. Any later differences between groups, according to the logic of experimental designs, could not then be attributable to pre-existing differences between teachers or pupils.

The results of the STAR project will only be briefly summarized here. Pupils were followed from kindergarten (aged 5) to third grade (aged 8). In both reading and maths pupils in small classes performed significantly better than pupils in regular classes. Moreover, small classes with one qualified teacher had pupils who did better than pupils in regular classes with an assistant. In fourth grade the pupils returned to regular classes and the experiment ended, but gains were still evident after a further three years, that is, grades 4–6 (Nye et al., 1993, Word et al., 1990).

The study appeared to provide clear findings on the independent effect on pupil achievement of small classes, but given the huge resource and pedagogical implications it is not surprising that it has been criticised. Some have argued that effect sizes are small and that funds would be better applied to more effective educational initiatives such as one to one tutoring (e.g., Slavin, 1989). Some have questioned the validity of the findings (Mitchell et al., 1991). A recent reanalysis of the STAR data using more sophisticated multi-level modelling techniques, provides perhaps the most convincing appraisal to be made. It supported the central finding of a difference between small and regular classes (both with and without teacher aide) for the kindergarten year, though the effect sizes of 0.17 for maths and 0.18 for reading are modest (Goldstein and Blatchford, 1997).

The STAR project was undoubtedly a bold and timely study. The results are compelling and have provided the basis for a number of educational initiatives and policies in the US and other countries, for example, in the Netherlands (Roel Bosker, personal communication), and also in the British Labour Party's commitment to

reducing class sizes during KS1 (5–7 years). But there are some difficulties and limitations with the STAR project. Elsewhere the limitations of experimental or randomized controlled trials (RCTs) as a way of studying the causal effect of class size differences on pupils' attainment have been examined (Goldstein and Blatchford, 1997). Here some main concerns that have been influential in our preference for a different approach to the study of class size effects are summarized.

3.2.1. Specific limitations of the STAR research

There are two main limitations. First, there was no baseline assessment of children on entry to the study. Although the project requires the random allocation of pupils to classes, without a baseline measure it is difficult to know if this condition was attained. It is also not possible to analyze progress over the kindergarten year and to study whether class size effects varied with initial attainment at baseline. Second, the statistical analysis did not take account of the hierarchical nature of education systems (separate levels of child, class/teacher, and school) which could have had an effect on results found.

3.2.2. How generalizable are the results?

One of the big problems with experimental studies such as the STAR project is that what they gain in control over the main variable of interest they can lose in the generalizability of results. This is seen in several ways. One limitation is the narrow range of class sizes studied. What the STAR study compared, by UK standards at least, was very small classes of around 15 children with small classes of around 23. There are important pedagogical questions about the value of small classes, but concern in the UK is more commonly with the pedagogical consequences of large classes of, say, more than 30 or 35 children.

There are also difficulties in generalizing beyond the particular types of schools involved in the STAR project. To what extent can we be sure that effects found in the large, three same-age class entry schools used in the STAR research will also be found in smaller schools, possibly with mixed-aged classes? Perhaps most obviously, there are difficulties in generalizing results from Tennessee to other countries.

3.2.3. The validity of findings from experimental research

One difficulty with experimental studies is the effect that allocation to experimental conditions may have on the validity of conclusions. Although the effects resulting from the manipulation of a variable (in this case, class size) might appear to be easily interpretable this is not necessarily the case. In the STAR project, for example, where different class types were set up in the same school, the teachers and pupils involved would have known which group they are in. As a result any effects found may be due to, for example, the expectations or resulting attitudes of those involved, rather than class size differences as such. A prior difficulty is that involvement in such a time consuming and invasive project is likely to mean that to some degree at least schools are self selected and therefore not necessarily typical. An alternative, between-school

design, in which classes of different sizes are compared in different schools, would be difficult to interpret because class size and school characteristics would be partly confounded. There are likely to be particular problems of interpretation when experimental conditions are not equally attractive to participants. It might be worth the reader putting himself or herself in the place of a teacher assigned for a school year to a larger class than a colleague in the same school.

From this point of view an evaluation of alternative approaches, all with a different educational rationale, might be preferable. However, it is important to ensure comparisons are fair. In the case of the STAR project the only change was to reduce the number of pupils; other interventions may involve a particular approach to instruction. To be fair we would need to test the effects of class size reduction with and without additional instructional input.

One cannot assume that changes brought about by experimental manipulations will necessarily apply outside the experiment. Experimental manipulations can be difficult to interpret because groups are necessarily isolated from the “real world” (Robson, 1993) of education. Education systems are complex and hierarchical and there are threats to validity if class size is studied independently of the wider system.

There are also practical difficulties with experimental research on class size. If a main experimental condition involves class size reductions this probably entails employing more teachers; this can make the costs involved prohibitive. A few years ago we costed an experimental study involving the creation of small classes in schools. Staff costs alone, needed to hire an extra 24 teachers and 24 extra classroom assistants, amounted to 1.5 million pounds. Other difficulties involved in the creation of new classes can be the physical limits of school buildings.

These points are not meant to imply that experimental designs cannot be of value, but it does mean that results have to be interpreted with more care than is sometimes the case. These kinds of worries about the validity of the experimental research have made us reappraise the possible advantages of an alternative approach.

3.3. *Longitudinal correlational studies*

There are difficulties with correlational research and also of experimental research. The reader might be forgiven for wondering at this point what hope there might be for research in this area. This touches on a difficult dilemma over alternative approaches that lies at the heart of much debate about quantitative educational research. There may be an obvious attraction to the adage “to understand something, change it”, but on balance it may be more productive to study class size effects by seeking to capture the real and complex world of education rather than control one feature of it. It may, in other words, be more valid to seek better understanding by measuring and examining relationships between factors, including class size, as they occur in the real world, and to make adjustments for relevant factors such as baseline attainment. In this connection Slavin (1990) compared the effect sizes of randomized experiments with seven correlational studies which matched classes or schools in terms of student characteristics, and found that the effect sizes (about 0.2–0.3 sds) were similar for both. This suggests that correlational approaches were effective in adjusting for intake

differences and supports the view that they may be as informative as experimental studies. If they are to be as informative, however, several issues need to be addressed.

First, there is a need to study pupils over time using longitudinal designs, obtaining measures of pupil attainment on entry to school. As has been seen, a possibly needless doubt has been cast on the STAR project because of the absence of baseline attainment data. One might add that the STAR project is consistent with other research in strongly suggesting that effects of class size are most likely to be evident with young children immediately after entry to school; this is when longitudinal research on class size effects should begin. There is little evidence that class size reductions on their own are effective as an initiative with older children who have been in school for some years.

Second, research would need to be based on a conceptual framework drawn from previous research on class size. It is important that care is taken over the operationalization and organization of variables. Previous research has tended not to have a strong rationale for the choice of variables.

Third, it is important to use statistical modelling that does justice to the complexity of educational systems in which class size may have an effect. In particular it is important to build on recent advances in multi-level statistical modelling (see Goldstein, 1995). The statistical analysis used in the STAR research was rather simple in the sense that it did not take account of separate levels of child, class/teacher, and school, all of which could have had an effect on results found. In a recent reanalysis of the STAR data using multilevel modelling techniques (Goldstein and Blatchford, 1997) it was found that for reading at grade 1 that there was significant variation between schools in the class size effect. This means that in some schools children in small classes did much better than those in regular classes, while in other schools children in small classes did not. It also means that this effect differed according to school subject (reading but not maths). These findings were not detected in the original analysis but are important in suggesting school level and subject level processes (as far as we know unmeasured in the STAR study) that affect class size effects.

The current Institute of Education research, described in more detail below, has been developed with these features in mind. This is not meant to imply that even with these improvements difficulties with correlational studies have been completely overcome; rather, it means that in contrast to experimental research the results may be more valid and there may be advantages in gaining a more realistic appraisal of class size effects. In particular, it is unrealistic to assume that class sizes are constant, as the STAR project required, and this real life variation may be an important influence.

4. What factors might explain the effect of class size on educational outcomes?

In the last section the causal link between class size differences, on the one hand, and educational outcomes, on the other, was examined. An equally important educational issue involves identification of factors that might explain any link found. In other words, it is important to ask what factors might mediate associations between

class size and outcomes. There has been little research that can provide information on this issue. Almost all the studies are from the US, and doubts exist about the reliability of some of the studies (see Blatchford and Mortimore, 1994). As seen, the STAR research was not set up to investigate processes that might explain any differences found between small and regular classes and it is therefore difficult to be sure why differences were found in the study. This lack of information is unfortunate because without it there are difficulties in explaining class size effects; it is also difficult to offer practical guidance on how to maximize the opportunities provided by classes of different sizes.

As discussed in Blatchford and Mortimore (1994), knowledge about mediating processes might also help explain why previous research has not always found a link between class size differences and outcomes. It may be, for example, that when faced with a larger class teachers might alter their style of teaching — they might tend to use more whole class teaching and concentrate more on a narrower range of basic topics. As a consequence, children's progress in these cases might not be different (and may even be superior to) children taught in smaller classes. More generally, it may be that when faced with larger classes teachers "compensate" in a number of ways (e.g., by working harder to maximize feedback to individual pupils). If this is true, pupil progress may not be affected adversely; however, there may be more covert costs, seen in more teacher stress, lower morale, and fewer opportunities for teacher planning. Another possibility is that some teachers do not alter their teaching to take advantage of smaller classes (as found in Shapson et al., 1980). This might explain why class size differences have little effect. In order to more closely examine these possibilities, detailed information on classroom processes would be needed.

The intention here is not to review research on mediating factors (for reviews see Blatchford and Mortimore, 1994, Cooper, 1989, NAHT, 1996). Instead, the discussion is broadened in order to identify some relevant methodological issues. First, what kind of research would be helpful in identifying mediating factors? One basic objective would be to collect information on classroom processes in order to see if they are affected by class size differences and whether they then affect educational outcomes. To take a simple example, it may be that in larger classes teachers have less opportunity to interact with individual pupils and offer them feedback on their work, and it may be this which explains why children in such classes make less progress. What would be needed here, therefore, would be identification and measurement of the mediating variables — in this case the amount of individual attention and feedback experienced by pupils. This approach may appear straightforward but there can be difficulties. One problem is defining whether a variable is a mediating or outcome variable. Pupil's difficult behavior, for example, may be considered a factor affecting the class size influence. A teacher in a class with more difficult children may devote less time to the rest and hence they may make less progress. On the other hand, difficulties of adjustment to school might be seen as an outcome, in the sense that children's difficulties may be brought into being or exacerbated by larger classes.

Another problem is the difficulty that can be faced in producing reliable and valid measures of mediating processes. In the review by Blatchford and Mortimore (1994)

the following factors were identified on the basis of past research as likely to be important processes: individualization of teaching, quality of teaching, curriculum coverage, pupil attention, better teacher control and less time spent on managing pupils' behaviour, more time and space, better morale, and better pupil–pupil relations. In some cases measures may be tangible and relatively easily measured — for example, the amount of teacher attention to individual children can be assessed using systematic observation methods (see Blatchford et al., 1987). However, this form of data collection can be very time consuming, can be intrusive in schools, and there are difficulties arising out of the changeability of behavior from day to day. At the very least, collection of data at multiple time points is required. Other mediating factors may not be easy to operationalize. It is notoriously difficult, for example, to measure “quality” of teaching and therefore examine how it is affected by class size differences. Developing adequate measures of teacher morale and stress may also prove difficult.

It is apparent that great care is required in identifying, operationalizing, and organizing measures of classroom processes. As described above, much previous research, when it has attempted to explain class size effects, has tended not to have a strong rationale for the choice of variables. It is important to select variables on the basis of previous research and to organize them in a coherent way, so that there are, for example, sets of variables, conceptually linked, perhaps cutting across different forms of data collection. It is likely that multiple methods will be needed, for example, classroom observations, teacher ratings of pupils, as well as pupil and teacher self reports.

One way of conceiving possible explanatory factors is to divide them, as Mitchell et al. (1991) do into “direct” and “indirect” effects. Direct effects relate to the kind of processes within classrooms discussed in this section. They include such variables as teaching methods, curriculum coverage, pupil attention, and relationships in class. In Mitchell et al's study, three direct effects were examined: greater instructional overhead, increased student interaction time, and decreased access to instructional resources (though in their study measures were not based on empirical research but derived on logical grounds). But Mitchell et al proposed a separate set of explanatory factors, which they call ‘indirect’ explanations. These derive from the spread of pupil abilities within a class and comprise what they call “class heterogeneity”, “instructional pacing”, and student grouping or achievement modelling. Mitchell et al. find most support for indirect explanations and none for direct effects. These are provocative results, in the sense that, if true, they mean class size differences in achievement are not attributable to pedagogical or educational factors at all. But there are doubts about the approach adopted. Perhaps most importantly it is not clear if similar results would be found with “real” data.

There are a number of models of influences on school progress that could be drawn upon. One of the earliest and most widely cited is Dunkin and Biddle's (1974) model which supported research in the “process–product” tradition. There were four stages — presage, context, process and product. Class size, in this model, is a context variable. Bennett (1996) uses this model in his report of a Teacher Association survey of teachers', parents', and school governors' views on class size effects. Other models

abound. Two recent ones are in Creemers (1994) and in Willms (1992). Willms (1992) draws on Tagiuri's (data) taxonomy of organizational climate which comprises four categories: ecology (physical and material environment of classroom, school and district), milieu (characteristics of individuals), social system (e.g., relations between teachers and pupils) and culture (e.g., norms and values of the school). Whichever model is used, variables would need to be specified at the separate levels of child, class/teacher, school, and even local authority. As described above, in our reanalysis of the STAR data there were differences between schools in the class size effect in reading, which suggests that processes at the school level that might explain these findings need to be studied.

Assuming that mediating processes can be measured reliably and organized in a conceptual framework then it is possible to set up statistical models in order to test the most important mediating factors, and this is the approach adopted in our current study. Systematic observation techniques, for example, are being used in our current research to record pupil attention to the teacher, the amount of pupil initiations and responses, the amount of teacher talk to the class, groups, and individuals, the amount of interactive feedback to pupils, and the extent to which the teacher-child interactions are task or procedure related, involve praise or discipline and so on, as well as aspects of pupil-pupil contact and behavior when not interacting (the observation method is described in Blatchford et al., 1987, Tizard et al., 1988). Other forms of data collection are also being used to study facets of classroom experience likely on conceptual grounds to be related to class size. The breadth and difficulty level of curriculum coverage, at both the child and class level, has been found in previous research to be an important predictor of school progress with young children (Tizard et al., 1988), and is being measured in the current research using teacher-completed checklists. Other forms of data collection are directed at within-class grouping strategies (Kutnick and Blatchford, 1996), and the extent of teacher time devoted to National Curriculum areas, and preparation.

It is sometimes assumed that a more effective strategy is to set up experimental studies in order to examine the processes affected by class size. The STAR project is the most famous example of such a study, though, as has been seen, process data do not appear to be available to explore intervening processes in a systematic way. The attraction of such studies is the assumption that differences between groups (for example, larger and smaller classes) in behavior or attitudes are likely to be attributable to differences in class size (assuming random allocation to groups). The difficulties in interpreting results from such designs, which seek to assess the causal effect of class size on outcomes, have already been examined. There may be particular difficulties in short term intervention studies, for example, where teachers are asked to teach in a class of a given size for only a short length of time. In such designs, changes in behavior and attitudes may be more the function of change itself. This is particularly likely when teachers are studied in artificial situations outside their normal classroom experience. A preferable design might be to select naturally occurring classes for comparison that are different in size but matched as far as possible with regard to other characteristics such as the social backgrounds of the students.

5. Alternative approaches to research on class size effects

A huge gap exists between commonly held views about the advantages of smaller class sizes, on the one hand, and the generally inconclusive results from research, on the other. It is therefore understandable if people question the nature and validity of the research that has been done. For the most part, research has been of a quantitative nature, and it is important to consider whether such approaches miss important processes and effects. One of the authors (PB), as part of the pilot work for our current project, has recently spent time in a relatively small reception class (children aged 4–5 years) of fewer than ten pupils and sought to identify factors associated with what appeared to a very productive experience for the pupils. Here are some of the conclusions drawn from this case study.

- The teacher appeared able to maintain control and a level of attention and noise that helped the pupils' work. She could control and reinforce appropriate behavior, for example with regard to calling out answers. This was much easier with fewer children.
- The teacher appeared able to organize the pupils' work efficiently and to get them ready for a task so they were clear what to do. She was able to help them see how to do a task, for example, cutting out fish shapes from paper and understanding which side was the "good" side and the "bad" side.
- The teacher appeared able to develop a teaching point, e.g., on letter sounds. She could explain the point and consolidate the children's understanding of it.
- The teacher appeared able to offer the pupils feedback, pitched to their individual levels of understanding. She was able to check their knowledge and grasp of a topic, offer feedback that was appropriate and encouraging, and then move them on. Thus, the sensitivity and effectiveness of feedback was enhanced.
- The teacher appeared able to pick up on and use events as they came up and then develop them into a teaching point; for example, the arrival of the classroom assistant with an aquarium allowed the teacher to instigate talk about fish and the planned trip tomorrow to buy them. In a larger class these opportunities can be potentially disruptive and so may be missed or avoided.
- The pupils appeared to be developing good relationships with each other, reinforced by the teacher, so they could help and encourage each other in their work and school behavior, and point out errors and inappropriate behavior. It was apparent that their relationships with each other (and the teacher) were friendly, but this appeared to have been carefully nurtured by the teacher.
- The pupils' ability to get the teachers' attention quickly and get difficulties dealt with appeared to be enhanced. They did not have to wait, and hence become discouraged or distracted; consequently, their concentration was more continuous and applied.

These are just a few points from a long list of impressions. Overall the small class appeared to facilitate more intensive teaching and sensitive feedback to pupils, and enhance relationships between pupils in the class, as well as the quality of pupils' work. No claims are made here for the objectivity or reliability of these impressions, or their generalizability to other classrooms. With only one class it is in any case difficult

to disentangle the effect of being in a small class with other factors that might be important, in particular qualities of the teacher herself. In this case, salient factors might be the fact that she knew the small class was only for a term (and therefore to be enjoyed while she had the chance), the fact that she was experienced, resourceful and enthusiastic, as well as fairly new to reception age children; and so on. Rather, the impressions are listed here to give some examples of the kinds of classroom behavior and processes that would need to be researched in order to understand differences between small and large classes.

But how can we measure or capture these features? As described in the last section, some can be captured by quantitative methods and studied in a larger study, perhaps using the kinds of statistical models discussed above. But some features are difficult, or require extensive resources, to capture quantitatively — for example, the quality of work and relationships in a class, and the efficiency of work preparation. There is likely to be a need to capture such features through more qualitative, case study approaches that describe events, relationships, and processes as they develop over time in a small number of classes. It may be that some advantages of smaller classes — which teachers have in mind when discussing small vs large classes — can only be captured in this latter way. There is a need to be clear which features can be measured quantitatively and which need to be studied qualitatively, and how the two types of data and analysis can be designed to inform each other.

This raises important questions about the limits of quantitative (and qualitative) research, that we do not have space to enter into here. At the very least it presents a challenge to researchers using quantitative approaches and raises questions about the limits and complementarity of different “paradigms” of research (Shulman, 1986). An alternative, more qualitative approach, adopted in the current research, involves case studies of a small sample of classes, chosen because of particular characteristics (e.g., that they have fewer than 20 and more than 30 pupils). The aim here is to research individual classes over time in order to build up a continuous and particular account of classroom processes and features of teaching and learning that are related to differences in class size. Each class is examined on common dimensions and then field notes analyzed to draw out illustrative experiences and common themes. Teachers are understandably keen that research in this area captures and does justice to their experiences, and no doubt suspicious that some of the force of that experience can get lost in research which collates results across schools. The type of case study research described here aims to test their experience by seeking a rich yet valid description of classroom processes in individual classes. Such research can also be effective in generating specific theories that can be evaluated and tested. Teachers are also being involved in case studies, in the context of a framework set by the wider research project.

However, it should not be assumed that ethnographic and other types of qualitative research can avoid issues of measurement and reliability (Hammersley, 1987). McIntyre and Macleod (1986) have criticized ethnographic research for taking quantitative researchers to task for an over reliance on numerical data while themselves unwittingly using quantitative (but anecdotal) statements in support of their own arguments. There are difficult issues involved in the credibility of evidence provided in

quantitative and case study research that can only be touched on here. As a general strategy it seems worth considering different approaches not in opposition to each other but as complementary. Without wishing to gloss over the difficulties involved, there is much to be said for seeking to integrate them in a study of class size effects — the aim being to use a common frame of reference for each in order that each will mutually inform the other. This is being attempted in our current Institute of Education research.

6. Conclusions

In this chapter some concerns with methods used in research on class size have been highlighted. There is a need for better measures of class size and allied measures like student–teacher ratios that would bring them closer to the day to day experience of pupils and teachers. There are well known limitations of cross sectional correlational research, but there also are perhaps less well understood difficulties with experimental research. Although applauding the timeliness and boldness of the STAR research, and supporting its main findings, some limitations do exist. These limitations have shaped our approach which seeks to improve on previous research by developing better measures of class sizes; using a longitudinal design with baseline assessment, by which to measure progress; building the research on a conceptual framework drawn from a review of previous research; and making use of multi-level modelling techniques to analyse effects at the conceptually separate, but possibly interacting levels, of child, class/teacher, and school.

Our current research has two main aims: first, to examine the connections between size of class in school and pupils' educational progress; second, to examine connections between size of class in school and classroom processes such as curriculum coverage and pupil attention in class. A third aim is to provide a valid and reliable description of class size itself. As might be anticipated from what has been said thus far, the research seeks to combine the strengths of quantitative and qualitative approaches. With regard to the former, it involves a large number of classes in schools (around 300) in a number of LEAs in England, and the collection of a range of information on pupil, class, teacher, and school characteristics, and classroom processes, some of which are described above. It involves in the first instance a two-year longitudinal follow up of children from school entry to the end of the second year in school (that is through the reception year and year 1), though it is hoped to follow the children beyond this point if funds are available. It also involves planned comparisons of classroom processes in a sub-sample of larger and smaller classes, using systematic observation techniques, in ways described above. The project also involves case studies of selected classrooms, again as described above.

The project is the result of extensive collaboration and discussion between the Institute and staff in Local Education Authorities, who are jointly funding it. We are impressed by the willingness in LEAs and in schools to support the drive for systematic information on class size effects, especially when for participating schools this has added to their already considerable workloads. Together we are aiming to

provide perhaps for the first time in the UK evidence which will be able to inform policy and practice on this vexed topic.

References

- Bassey, M. (1996). Inspection, unlike research, is disinterested! *Research Intelligence*, 55.
- Bennett, N. (1996). Class size in primary schools in primary schools: Perceptions of headteachers, chairs of governors, teachers and parents. *British Educational Research Journal*, 22(1), 33–55.
- Blatchford, P., Burke, J., Farquhar, C., Plewis, I., & Tizard, B. (1987). A systematic observation study of children's behaviour at Infant school. *Research Papers in Education*, 2(1), 47–62.
- Blatchford, P., & Mortimore, P. (1994). The issue of class size in schools: What can we learn from research? *Oxford Review of Education*, 20, 411–428.
- Cooper, H.M. (1989) Does reducing student-to-teacher ratios affect achievement? *Educational Psychologist*, 24, 79–98.
- Creemers, B. (1994). *The effective classroom*. London: Cassell.
- Dunkin, M.J., & Biddle, B. (1974). *The study of teaching*. New York: Holt, Reinhart and Winston.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Goldstein, H., & Blatchford, P. (1997). *Class size and educational achievement: A methodological review*. Report commissioned by UNESCO, Geneva.
- Hammersley, M. (1987). Some notes on the terms validity and reliability. *British Educational Research Journal*, 13, 73–81.
- Kutnick, P., & Blatchford, P. (1996). *Studying pupil groups in primary school classrooms: Where is the pedagogy?* Paper to British Educational Research Association Annual Conference, Lancaster University, September.
- McIntyre, D., & Macleod, G. (1986). The characteristics and uses of systematic observation. In M.Hammersley (Ed.), *Controversies in classroom research*. Milton Keynes: Open University Press.
- Mitchell, D. E., Beach, S. A., & Baduruk, G. (1991). *Modelling the relationship between achievement and class size: A re-analysis of the Tennessee project STAR data*. Riverside, CA: California Educational Research Co-operative.
- Mortimore, P., & Blatchford, P. (1993). The issue of class size. In National Commission on Education Briefings (pp. 167–182). London: Heinemann.
- NAHT (1996). *Class size research and the quality of education*. Sussex: National Association of Head Teachers.
- Nye, B. A., Boyd-Zaharias, J., Fulton, B., DeWayne, Achilles, C. M. (1993). The lasting benefits study. Sixth grade technical report. Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University.
- Organisation for Economic Cooperation and Development (OECD) (1996). *Education at a glance*. Paris: OECD.
- OFSTED (1995). *Class size and the quality of education*. London: OFSTED.
- Robson, C. (1993). *Real world research*. Oxford: Blackwell.
- Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, 17, 144–152.
- Shulman, L. (1986). Paradigms and research programs in the study of teaching: a contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching*, (3rd ed.). New York: Macmillan.
- Slavin, R. E. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist*, 24, 99–110.
- Slavin, R. (1990). Class size and student achievement: Is smaller better? *Contemporary Education*, 62, 6–12.
- Tizard, B., Blatchford, P., Burke, J., Farquhar, C., & Plewis, I. (1988). *Young children at school in the inner city*. Hove: Lawrence Erlbaum Associates.
- Willms, J. D. (1992). *Monitoring school performance. A guide for educators*. London: Falmer.

Word, E. R., Johnston, J., Bain, H. P., Fulton, B. D. et al. (1990). The State of Tennessee's student/teacher achievement ratio (STAR) Project, Technical Report 1985–90. Nashville TN: Tennessee State University.

Peter Blatchford is Reader in Education at the Institute of Education. Currently he is directing a large scale research project on the educational effects of class size differences at Key Stage 1, and an allied ESRC funded study of the nature and uses of grouping practices in primary school. He has worked with Prof Harvey Goldstein on a reanalysis of the STAR class size project. Previously, he worked with colleagues from Thomas Coram Research Unit on a major longitudinal study of pupils' progress in London primary school (cf. Tizard et al., *Young Children at School in the Inner City*, Lawrence Erlbaum, 1988) and has published papers on the same pupils' academic self-assessment, self-concept and views on school throughout their school careers. He has published widely on pupils' causal relations in school and break-time/recess.