Rethinking Class Size

Blatchford, Peter, Russell, Anthony

Published by University College London

Blatchford, Peter and Anthony Russell.
Rethinking Class Size: The complex story of impact on teaching and learning.
Project MUSE. muse.jhu.edu/book/81855.

For additional information about this book
https://muse.jhu.edu/book/81855

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2777523
In preparation for this and the subsequent chapters we first extend what we said in Chapter 1 about wanting to move debate and research beyond just whether class size affects academic outcomes, toward a richer understanding of the classroom processes that might be at play. In this chapter we provide more background to how we have come to think about class size as a classroom contextual influence. This chapter therefore extends the discussion of our aims in Chapter 1, in particular Aim 3, by providing the background to a new conceptualisation of class size effects on classroom processes. This chapter sets the scene for Chapters 3 to 9, in which we present our findings on class size and classroom processes, to be followed by the presentation of our summary model of findings in Chapter 10.

In the second half of this chapter we provide more details on the research projects on which the book is based and identify the three main methods of data collection, along with an explanation of the mixed method approach we used.

Theories of class size effects

Perhaps the most obvious starting point when considering learning in school-aged students is within child factors such as intelligence, motivation and recent insights from neuroscience and genetics. There is an understandable rationale here in that these factors seem most likely to be the most direct influences on pupil learning. The theories of Piaget,
Vygotsky and Bruner have been influential in understanding learning and cognitive development (see Illeris 2007), and there is now a burgeoning literature on psychological processes connected to learning (see chapters in Harris et al. 2012). Claims for the important role of genetics in human development are becoming ever stronger (for example, Plomin 2018).

An additional and common way of examining factors that influence learning and school performance has been more through a sociological lens, for example, on the influence of the family and demographic factors like social class, ethnicity and mothers’ educational level. There has also been extensive interest in school and educational system features that are effective in terms of pupil academic performance, fuelled recently by the PISA results published by the OECD.

Despite the importance of these influences on development, we argue that they are not sufficient to account for learning in classrooms because, first, in the case of individual psychological approaches, the concern with intrinsic within-child characteristics is insufficient to account for classroom learning, and, second, in the case of sociological and school-level analyses, they are too ‘distal’ from the child’s learning. What is needed is a way of capturing the more immediate classroom-based influences on learning, and this includes the number of pupils in the classroom.

Historically, there are a number of models of influences on school progress which recognise some aspects of classroom processes. Dunkin and Biddle’s (1974) early model was supported by research in the ‘process–product’ tradition and had four stages – presage, context, process and product. Class size, in this model, is a context variable, teacher–pupil interactions are a process variable, and the ‘product’ is pupil attainment. Another model, by Pianta et al. (2002), divides influences on education into ‘distal’ versus ‘proximal’ and predictably finds that relatively distal structures like class size have less influence than proximal factors like classroom processes, teaching and the emotional quality of the classroom setting.

One of the limitations of these kinds of models is that classroom contextual features like class size are seen as background, static and relatively distant factors and are given a minor role, if any, as an influence on learning. The results in this book suggest an alternative is to position class size as a factor that might not be as directly important as teaching and other facets of classroom life, but which will interact with and influence them on a moment by moment basis. The important question is not therefore which is important, class size or teaching, but how to
describe the interconnections between them. We need a more dynamic and nuanced approach to classroom contexts and effects on learning.

Another kind of framework for understanding educational influences, which has the virtue of being closer to pupil learning, is the field of teacher effectiveness. As with so-called ‘process–product’ research, quantitative methods have typically been used to identify, usually through correlational analysis, the most important aspects of teaching affecting pupil outcomes (Creemers 1994; Ko et al. 2013; Kyriacou 2009; Muijs and Reynolds 2011). There has also been more recent specifically psychological work, for example on instruction in relation to thinking and reasoning in science (Sinatra and Chinn 2012), instruction for the development of learning strategies (MacArthur 2012), and problem-based learning (Loyens et al. 2012), and approaches to teacher–pupil relationships, informed by work on mother–child relations (for example, the CLASS system of Pianta and colleagues – see Hamre and Pianta 2010). Despite their many strengths, these strands of research also have in common a lack of interest in specific classroom contextual influences on teaching. In some studies, there is coverage of dimensions like supportive classroom environments, but this does not approach the classroom context in the more fundamental way we think is necessary. For the most part there is an underlying assumption in many studies of a direct model, where teaching affects, in a causal way, pupils’ achievements and learning. But, as shown repeatedly in this book, teachers do not meet pupils out of context, and class size can be seen as one contextual influence on classroom life, to which teachers and pupils will inevitably have to adapt, and which will affect their behaviour and the nature of the interactions between them.

We need, therefore, conceptual frameworks and theories that better help us account for class size effects. As Mitchell et al. (1989) have said:

... without an adequate theoretical conception of how changing class size might influence student achievement we are likely to ... draw the wrong conclusions. (37)

Unfortunately, however, there have been surprisingly few efforts made to provide such theories (Grissmer 1999), perhaps in part because disagreement over the effects of class size has inhibited efforts to develop theories to account for them.

Though not directed at class size effects as such, there are insights to be gained from previous models and theories of classrooms more
generally. As we saw in Chapter 1, one approach that nicely captures the dynamic nature of classroom life was put forward by Doyle (1986) and it has been influential in showing there are important elements in place, over and above the characteristics of particular teachers and pupils. He identified distinctive elements of classroom environments, including ‘multidimensionality’ (the classroom is often a crowded place, and there is a large quantity of events and tasks in the classroom); ‘simultaneity’ (many things happen at once in classrooms, perhaps especially in primary schools); and ‘immediacy’ (there is a rapid pace in classroom events). Doyle argues that these dimensions create pressures that shape the task of teaching. Their effect varies no doubt, but the pressures operate in all classrooms regardless of how teachers organise activities. It seems plausible that some of these dimensions will be affected by class size, for instance, a larger class may well mean the first two dimensions, multidimensionality and simultaneity, increase in intensity, but as far as we know this possible direction has not been explored.

Doyle’s work is part of a long tradition of research built on close observations of classroom life. There are a number of other early and still insightful accounts of classroom processes that repay reading (for example, Kounin 1970; Jackson 1968), as well as a rich tradition of more qualitative sociological, ethnographical and linguistic studies of classroom discourse and teacher–pupil interaction (for example, classroom language studies by Barnes and Todd 1981; Cazden 2001; Edwards and Westgate 1994; Mercer 2000; Sinclair and Coulthard 1975; Stubbs 1983; classroom interaction studies by Nuthall 2007; Pollard et al. 1994; ethnographical studies of classrooms by Mehan 1979; Woods 1986). As we have said, in our view it is unfortunate that there seems to have been a reduction in these kinds of descriptive, educational studies of classroom life and processes. But valuable as these descriptive studies are, they do not to date allow us to get very far in understanding the influence of the classroom contextual feature of the number of pupils.

Specific accounts of class size effects

So far, we have looked at more general approaches that might be helpful in conceiving of class size effects. In this section we briefly describe some main models which have been developed specifically to account for class size effects.
Cahen et al. (1983) in their book-length treatment of class size effects uncover many ways in which class size is important in teaching. They do not present a formal model of effects, though they do identify three key summary processes affected by class size: behaviour management, individualisation and effective coverage of the curriculum.

Zahorik et al. (2002, 6) provide a more formal model of class size effects on teaching and learning. Reduced class size means less discipline/more instructional time, more knowledge of students and more teacher enthusiasm. This in turn leads to two lines of causal effects. First, more individualisation (in terms of personalised procedures emphasising articulation and critique of understandings, common content, and one-to-one, small group and class participation) leads to more student self-direction, thinking and responsibility, and this leads to more student achievement. A second, parallel line of influences leads to more hands-on activities, which leads to deeper and more content, which also leads to more student achievement.

Finn, in a number of publications, has also provided models of class size effects (Finn et al. 2003; Finn and Shanahan 2016; Finn 2019). Finn identifies the following factors as influenced by class size: closer relationships, including pupils being more supportive of each other; covering the curriculum in more depth; having more time for instruction in terms of individualised instruction, and assessment of student performance.

Another model of classroom processes affected by class size was put forward by Anderson (2000). He proposed that there are three main ways in which reduced class size has its effect on student achievement: first, there are fewer disciplinary problems and therefore more instructional time and greater opportunity to learn; second, there is greater knowledge of students and therefore more appropriate personalised instruction and greater student engagement in learning; and, third, there is greater teacher satisfaction and enthusiasm and therefore greater teacher effort and more in depth treatment of content.

Anderson conjectured that increased knowledge, greater teacher satisfaction and time, and hard work resulting from smaller classes enables teachers to teach better without necessarily teaching differently (2000, 16). The effect of class size is to change the substance not the form of teaching. From this point of view, smaller classes therefore provide opportunities for teachers to teach better, but they do not cause teachers to do so. As we say in Chapter 3, Anderson suggested his model to be a starting point for a conceptualisation of class size effects.

These specific accounts of class size effects are helpful but in our view are not sufficient, because they are not based on detailed study
of classroom process connected to class size, they have not analysed closely the interconnections between classroom processes and class size, and because they have not formally developed the kind of contextual approach to class size effects we feel is necessary.

In Blatchford et al. (2003b) we developed a model as a way of summarising results from the CSPAR KS1 study. This collated our knowledge at the end of the KS1 phase of the research, when children reached seven years of age, but it now needs to be extended to account for the further work we have done, as described in this book.

Social psychology

Perhaps the most obvious discipline with potential for understanding class size effects is social psychology (Finn et al. 2003). We look briefly at social psychological approaches to group performance and processes in Chapter 5, when examining peer relations and class size. Social psychological theories would suggest (but to date have not shown) a negative effect from larger groups. A decrease in effort stemming from being part of a group has been labelled ‘social loafing’ (Latané et al. 1979), which would be expected to increase with the size of the group or class. An allied concept is that of the ‘free rider’ effect, where group members contribute little or nothing to the group activity and product. A connected theory is a ‘dilution’ effect, that is, increases in the numbers of pupils have necessary effects on the amount of attention a pupil receives from a teacher. Borland et al. (2005) argue that dilution of the teacher’s time across a large number of students lowers the impact of the teacher on any individual student, and so lowers academic achievement. Finn et al. (2003) point to a similar well-known social psychological theory – diffusion of responsibility – which is when people tend to be less likely to help others in distress when part of a group. As group size increases, more diffusion of responsibility takes place. Finn et al. argue that this applies to the classroom, though no research to date has been conducted.

Another social psychological theory, cited by Finn et al. (2003) in relation to class size, is ‘group cohesiveness’ or team spirit. When applied to class size the idea would be that students in small classes are more likely to support each other, while in larger classes groups may divide, hindering teaching and learning. Group cohesiveness is similar to an allied notion of ‘psychological sense of community’ (PSOC), for instance, a perception of similarity and interdependence with others, and feeling part of a larger stable structure. Bateman (2002) argues that PSOC is enhanced in smaller classes.
Although social psychological theories look to have important implications for class size effects it is noticeable that there has been little systematic effort to apply or test these concepts in relation to understanding class size effects in schools. In a small-scale study based on eight teachers’ views, Englehart (2006) examined the relevance of several social psychological theories and found some support for social loafing only. Further work would benefit social psychology and education.

But there is also the point, made by Heft (2001), that social psychological or group-based perspectives do not go far enough in helping us understand how the immediate environment affects action and behaviour. This is because we need conceptual frameworks that help identify and explore the particular contexts and settings within which people find themselves and groups develop and operate. Individual and even group-based approaches are insufficient for this task.

Contextual approach

We need then a more thorough analysis of the contextual basis of classroom learning. We have seen that the main traditions of research have tended to consider the effects of teaching and teacher–pupil interactions independently of the environment in which these interactions occur. Some time ago, we made the point that further progress in conceptualising class size effects requires a more fully worked recognition of ‘a contextual approach to learning, within which class size differences have effects on both teachers and pupils’ (Blatchford et al. 2003a, 709–10).

What we can take from social psychological approaches, and Doyle’s model, is the basic orientating point that behaviour is affected by the situation as well as by individual characteristics and personalities within the situation. A basic tenet of social psychology (Ross and Nisbett 1991) is that it is easy to underestimate the effect of the situation on behaviour – this is what is called the ‘basic attribution error’.

In terms of a conceptualisation of causal influences, this interest in classroom contextual effects turns on its head the usual way of conceiving classroom effects; that is, in terms of teaching (in research terms, the independent variable) affecting pupil learning or attainment (the dependent or outcome variable). Instead, one can consider interactions and behaviour as dependent variables, with the context or environment of the focus of attention as the independent variable.

One well-known expression of a contextual approach can be found in Bronfenbrenner (1979), although in our view Bronfenbrenner’s model
Ecological psychology

One helpful approach, we believe, is ecological psychology, because it helps provide a fuller conceptualisation of classroom influences, including the role of class size. Harry Heft, in a book (2001) and a recent article (2018), argues that ecological psychology is an impressive though sadly neglected programme of research and theory. The general underpinning idea is that settings or regions within which daily life takes place are important and qualitatively distinct, and influence in profound ways the actions of people in the settings. The roots of this approach are explained in Heft’s *Ecological Psychology in Context: James Gibson, Roger Barker, and the legacy of William James’s Radical Empiricism* (2001). This hardly seems the most instantly accessible book, but Heft nicely shows how ecological psychology has its roots in William James, one of the founding fathers of psychology, though its more recent precursor is in the social psychology of Kurt Lewin. Heft provides a fascinating account of the dual histories of ecological psychology in the works of James Gibson and Roger Barker (a PhD student of Lewin). He outlines how the latter in particular helped show that psychology has been handicapped historically because of a lack of a coherent framework to describe the environment within which individuals live. Psychology has been almost exclusively concerned with individual subjectivities and rarely with analysing the environment that would enable commonalities between individual perceptions.

We believe that Barker (1968) pointed out something quite profound: psychology is unique among sciences in that it began explicitly as an experimental discipline, and unlike other natural sciences it has never had a well-developed descriptive phase. He regrets this limitation, and it led to his efforts to study naturally occurring behaviour. It reinforces for us the value of descriptive, observational studies of what goes on in naturally occurring contexts within schools (and in everyday life).

In the book *One Boy’s Day*, Barker and Wright (1951) provided detailed written records of the observed activities of an individual child over the course of his day. They found these narrative records were extremely effective as a way of describing multiple attributes of behaviour and the immediate situation. The basic observation data – ‘behavioural streams’ – are interesting but perhaps limited in value. It was by examination of the records that they determined more useful
‘episodic units’. Barker realised that children’s behaviour was structured and indeed to a degree predictable if, instead of looking for causes of behaviour, he looked at the proximal environmental factors around the behaviour.

He realised, in other words, that children’s behaviour changed as they moved from one region or setting to another – say from the classroom to the hall, or the corridor to the playground – and that the behaviour of different children within the same setting was more similar than that of one of them in different settings. Interestingly, Golding (2017) recently showed how different school settings, that is, the corridor, the lunchroom and the playground, involve different forces and different behaviours. Barker goes on to argue – still provocatively – that:

… we could predict some aspects of children’s behaviour more adequately from knowledge of the behaviour characteristics of the drugstores, arithmetic classes and baseball games they inhabited than from knowledge of the behaviour tendencies of particular children. (Barker 1968, 4)

Perhaps Barker’s key idea is the identification of a discrete, immediate and dynamic unit, which he called a ‘behaviour setting’. This is a region in the community (including schools) which can be characterised as:

… an emergent, dynamic structure constituted by interdependent, joint actions among individuals and features of the material environment (milieu) considered over some extended period of time.

A person who inhabits the setting is a component part, a fixture of the behaviour setting, and as such is:

… anonymous and replaceable, and his behaviour is subject to the non-psychological laws of the superordinate unit. At the same time, however, every inhabitant of a behaviour setting is a unique person subject to the laws of individual psychology, where his own private motives, capacities, and perceptions are the causal variables … (Barker 1968, 17)

Heft gives the example of a primary school language lesson. This involves a group of students and a teacher in a specific location at a particular time with supportive materials (for example, books, chairs) for the express
purpose of conducting and participating in the lesson. Although we can’t predict exactly how a child will behave in the lesson there are noticeable constraints on what is likely, for example, sitting, reading, listening, writing etc., and not usually running, shouting or tossing a ball. These are typical classroom behaviours we take for granted but they emerge as congruent with the locale where they are observed at a given time.

Ecological psychology: Application to education and class size

Barker’s work shows us that individual and group approaches to understanding children’s behaviour in schools tend to miss the overall reality and influence of the ecological unit within which children spend their time:

For students of education this means that schools must be studied as carefully as the behaviour of the individual children within them. (1968, 15)

This idea introduces a powerful way of looking at classrooms and one which we believe is highly relevant to our task of developing a conceptualisation of class size effects. That is, the focus is not just on the behaviour of individuals within the classroom but regards behaviour in classrooms as understandable through a higher order conceptualisation of the interdependencies between factors in the classroom.

Thankfully, Barker and his colleagues Paul Gump and Phil Schoggen did much to apply ecological psychology to schools. The main use of ecological psychology in relation to education is Roger Barker and Paul Gump’s book *Big School, Small School: High School Size and Student Behaviour* (1964). From the point of view of our focus on class size, the key idea is that of ‘underpopulated’ settings. Barker argues that there is an optimal functional level (number of people) in a setting and when the number is reduced below this optimal level there are certain predictable consequences for the people in the setting. The two main consequences are: (1) an increase in the strength of the forces acting upon the individual inhabitants; and (2) an increase in the range of the direction of the forces acting upon the individual inhabitants (Barker and Gump 1964, 21). The basic idea is that when there are fewer available people, they are pressed more strongly to produce the same number and variety of behaviour units. And fewer behaviour units are pressed to produce the same number and variety of achievements.
We can take some valuable points from ecological psychology, for example, the importance of the idea of space and context; the identification of a meaningful ecological unit, for example, the ‘behaviour setting’; and insights into the effects of ‘underpopulated’ settings on behaviour. These ideas help with the development of an account of how class size works. Different class sizes may well induce different dynamics, which influence both teachers and pupils. So class size is more than a static, presage factor, as in Dunkin and Biddle’s model, and is, rather, something that influences teachers and pupils on a moment by moment basis.

There is much to commend the ecological psychology approach, but there are two obvious issues when it comes to understanding class size effects. First, it is interesting that the idea of underpopulated settings is applied to school size when it is perhaps more obviously applicable to the size of the school class, because this is the more direct, proximal context for the child and for teaching. As far as we know, however, there are no studies specifically using ecological psychological ideas in relation to class size. Second, the idea of ‘underpopulated’ is specified in relation to an optimal number, so that anything below that is seen as to a degree dysfunctional – that is, it is largely about how the system adapts when it is abnormally low in numbers. One issue here is the problem of determining an ‘optimal’ number of pupils, and an allied issue is that small classes are regarded by many teachers as anything but dysfunctional! One is also led to wonder about the consequences from an ecological psychology point of view of the most obvious problem for teachers, that is, larger class sizes, or what might in ecological psychology terms be called ‘overpopulated settings’.

Ecological psychology is not therefore developed enough to fully comprehend the effects of class size on classroom processes. It helps us conceptualise the context or setting within which action occurs, but needs to be developed in order to capture a full analysis of the kinds of classroom influences and processes affected by class size.

Summary

In this early part of the chapter we have tried to provide the basis for rethinking the way that class size has an effect on teachers and pupils. We have moved away from a model in which class size is considered exclusively in terms of associations with academic attainment to one which seeks to map out how class size works as part of a contextual approach. We reviewed some specific class size models. Though helpful, they do not in our view fully capture the factors related to class size.
identified in this book, or their interconnections. We paid particular attention to social psychology and ecological psychology for the valuable insights we think it gives to such a contextual approach to classroom learning.

In the following chapters we present our findings on class size and classroom processes, and in Chapter 10 we return to the more general perspective on class size and provide a summary description of our findings. We also there introduce the social pedagogical perspective that we believe is helpful when seeking to understand class size effects.

In the second part of this chapter we explain the research basis for this book.

The IOE research programme and methods of data collection and analysis

What is this book based on? As already mentioned, it is based on our reading of the literature on class size and on our extensive experience of teaching (AR) and research in classrooms (PB and AR). But most importantly this book is based on a pioneering large-scale programme of research extending over many years.

In the interests of narrative and accessibility we do not intend in this book to provide full technical details of the design and forms of data collection and analysis. But we will now describe the key features of our approach in order to allow the reader to appreciate the scale and significance of the methods, understand the basis for our findings and our interpretations, and feel confident in the conclusions we draw (or provide the basis for a critique of them).

Let us describe the projects within the programme.

CSPAR

The main source for this book is the Class Size and Pupil Adult Ratio (CSPAR) project, probably the largest study worldwide of class size and classroom processes and, we believe, unique in terms of the depth, detail, scale and rigour of data collection.

The origins of this project go back to 1996. PB, working with Peter Mortimore, then Director of the IOE; Harvey Goldstein, Professor of Statistics at the IOE; and Clare Martin, the first researcher on the project, put together a consortium of English Local Authorities who were willing to collaborate in a study designed to obtain systematic and objective
information on class sizes in schools and their effects. This willingness was energised by the lack of interest in the issue by the Government of the day. We were aware of previous research on the topic and in particular the Tennessee STAR project and, indeed, organised a seminar in London in order to hear the STAR principal investigators, including Chuck Achilles and Jeremy Finn, speak about the methods and results.

It became clear to us that a new approach to the topic was required. The IOE CSPAR project was set up to answer, for the first time in the UK, questions about two things: the connection between class size and pupil academic outcomes, and the connection between class size and classroom processes like teaching, grouping practices and pupil behaviour. It had a number of features that were designed to build on what we saw as the gaps and limitations of previous research. In line with the limitations concerning experimental designs, which we describe in Chapter 3, an ‘observational’ or ‘naturalistic’ approach was adopted rather than an interventionist one involving random assignment. This also involved a longitudinal design. Baseline assessment and start of school year scores were used in order to adjust for possible purposive or non-random selection of children into classes on the basis of their pre-existing achievement.

The first stage of the study followed a large sample of children from school entry over the first three years of school in English schools (that is age 4–7 years – Reception, followed by KS1 – covering Years 1 and 2). There were 220 schools, with 368 classes and 9,330 children in eight Local Education Authorities (LEAs) involved. Schools were randomly selected from within the participating LEAs, drawn from a wide range of social backgrounds, and were situated in urban, suburban and rural areas. All children entering the first year in selected schools were included in the study.

The CSPAR project then continued over KS2, that is from when the pupils were 7 years through to when they were 11 years. In this stage there were 202 schools and 332 classes in Year 4 (age 8–9 years), which, through attrition, reduced to 173 schools and 261 classes in Year 5 (age 9–10) and 153 schools and 224 classes in Year 6 (age 10–11). (Data collection did not take place when the pupils were in Year 3.) In these schools we followed 8,728 pupils in Year 4, 6,607 in Year 5 and 5,755 in Year 6. This later stage was funded by the UK Government.

As described in more detail in Chapter 3, for the analysis of class size and academic progress we used multilevel statistical procedures to model effects of class size differences on pupil attainment. We controlled
for extraneous, potentially confounding, sources of variation that might affect the relationship, such as gender, child earlier attainment, and family income. In using an ‘observational’ design, we were able to capture the nature of the relationship between class size and achievement across the full range of observed class sizes (not just a restricted range). This enabled us to see whether certain class sizes or bands of class sizes had stronger effects than others. We also employed a more sophisticated approach to modelling the relationship between class size and achievement than that conducted in previous research (see Blatchford, Goldstein et al. 2002).

The CSPAR was therefore what we have called in this book a ‘dedicated’ study of class size effects, in which the team developed new methods able to measure class size and allied classroom factors in a reliable way as well as classroom processes connected to class size. The study developed a range of sources of data. In summary, these were:

- **Data on class size, pupil–adult ratios and presence of teaching assistants (TAs)** from (termly) questionnaires completed by teachers
- **Teacher questionnaires** which asked for information on biographical details and views and experiences on a range of issues
- **Headteacher questionnaires** which asked for information on a number of issues including allocation of teachers to classrooms
- **teaching assistant questionnaires**
- **Assessments of pupils** at the start of the first year of school and at end of each subsequent school year in maths and literacy
- **Pupil background details** including age, gender, free school meal entitlement
- **Class characteristics** including whether the pupils were taught in sets or classes, amount of time teaching, and number and size of within-class groups
- **Case studies** on a sub-sample of small and large classes in Year 5 and Year 6, comprising semi-structured observations and interviews with pupils, teachers and TAs
- **Systematic observations** on a sub-sample of small (under 25) and large (over 30) Year 6 classes in Year 6 (10–11 years).

In this book we concentrate in particular on three forms of data collection designed to provide insights into classroom processes related to class
size: (1) the annual questionnaire surveys of teachers, (2) systematic observations of pupils and (3) detailed case studies. We say more about these shortly.

Publications from the KS1 study have been on relationships between class size and attainment over the Reception year (Blatchford et al. 2002a); class size and within-class groupings (Blatchford et al. 2001); class size and teaching (Blatchford et al. 2002b); class size and pupil attentiveness and peer relations (Blatchford et al. 2003c); as well as a book-length treatment and research article on the whole Reception and KS1 study (Blatchford et al. 2003b; Blatchford et al. 2003a).

There have been fewer publications from the KS2 stage of the research (though see Blatchford et al. 2005; Blatchford et al. 2007). The aim of this book is to give expression to the extensive analysis at KS2 of the three forms of data collection. This book builds on earlier publications, but stands alone, as we shall see, because we present new results from the CSPAR KS2 (7–11 years) stage.

DISS

We also draw on data from another project we conducted – the Deployment and Impact of Support Staff (DISS) study (2003–9) – funded by the English and Welsh governments. This complemented the CSPAR study, and is the largest study worldwide of the deployment and impact of teaching assistants (TAs). It was voted in the top 40 educational research projects in the last 40 years by the British Educational Research Association (BERA).

The key research task in DISS was to establish the causal role of support from TAs in relation to pupils’ attainment and other pupil outcomes. A traditional approach would be to use an experimental design, contrasting groups with and without support. But, as with research on class size effects (Goldstein and Blatchford 1998), experimental manipulations can have a narrow range of applicability and do not easily capture the everyday ways that support staff are used and deployed. An alternative naturalistic design was therefore used that sought to measure the amount of support received by pupils under normal circumstances and then examine relationships with academic and behaviour outcomes. It had a similar design to the CSPAR study in that it involved a longitudinal study of pupil progress connected to the amount of TA support for each pupil, controlling for other factors, including pupil characteristics like prior attainment and SEND status that might be expected to affect
the relationship. It was again a multi-method study involving national questionnaire surveys, a systematic observation component and in-depth case studies. Overall, it combined numerical data on connections with pupil and teacher outcomes along with qualitative, interpretive analysis to obtain a detailed and integrated account of the deployment and impact of support staff. There were seven different age groups, across the primary and secondary school stages. ‘Wave 1’ took place in 2005/6 and focused on pupils in 76 schools in Years 1 (age 5–6), 3 (age 7–8), 7 (age 11–12) and 10 (age 14–15), and ‘Wave 2’ took place in 2007/8 and involved an increased sample of pupils in 77 schools in Years 2 (age 6–7), 6 (age 10–11) and 9 (age 13–14). In total there were nearly 5,000 pupils across the seven age groups. Schools were nationally representative.

There are several reasons why the DISS study is relevant to this book. The most obvious reason is that, like the CSPSAR study, the DISS study included a systematic observation component in which observations were carried out over 2005/6 in 49 mainstream schools. The rationale for the coding methods is explained in Chapters 3 and 4. Unlike CSPAR, observations were conducted in all class sizes, rather than preselection of large versus small classes. A measure of class size was included in the observation schedules for each observation point, as we see below, allowing a powerful analysis of the relationship between class size and teacher and pupil behaviour across the full distribution of class sizes. The DISS study also extended the systematic observation component of the CSPAR study because it involved four year groups in primary and secondary schools, Year 1 and Year 3 (5–6 and 7–8 years) and Year 7 and Year 10 (11–12 and 14–15 years).

Further details of the DISS study (especially the systematic observation component) will be presented below and in the following chapters and can also be found in the book by Blatchford et al. (2012) and research articles by Blatchford et al. (2009); Blatchford et al. (2011a); Blatchford et al. (2011b).

A more substantive reason for the relevance of the DISS project is that over recent years a major shift has taken pace in the UK, with a recent increase in paraprofessionals working in classrooms, which has relevance for the class size debate. A key use of TAs is to help pupils who are struggling and/or with SEND. Results from the STAR project (Finn and Achilles 1999) suggested that the presence of a TA did nothing to improve attainment in regular class sizes (the key factor was being in a small class). But pupils with SEND in the UK are much more likely to be found in mainstream classrooms nowadays (in contrast to the time
of STAR) and, as we shall see, a key way of meeting this inclusion agenda, especially with large class sizes, has been to hire TAs. Given this, we wanted to examine systematically whether pupil outcomes and classroom processes were affected by TAs. We also examined the effect of the amount of support for individual pupils, rather than just their presence or not in a classroom, as in STAR.

**MAST and SENSE**

In this book we also draw on results from the Making a Statement (MAST) and Special Education Needs in Secondary Education (SENSE) projects, which tracked the everyday classroom experiences of pupils with SEND in primary schools in Year 5 (9–10 years) and secondary mainstream and special schools in Year 9 (13–14 years), respectively. Both studies were funded by the Nuffield Foundation, and co-directed by Peter Blatchford and Rob Webster. Whereas the DISS study looked at TAs, the MAST and SENSE studies focused on the day to day experiences of pupils most likely to be supported by TAs.

In the first phase (MAST), we collected data on 48 pupils at Year 5 (9–10 years old) who had a Statement for either moderate learning difficulties (MLD) or behaviour, emotional and social difficulties (BESD). In 2014, the SEND Code of Practice was revised. Statements began to be replaced by Education, Health and Care Plans (EHCPs) and the categories of SEND were reorganised. So, for the second phase of data collection (SENSE), in order to offer some consistency between the two cohorts, we prioritised the recruitment of pupils with needs relating to cognition and learning. In the SENSE study, 49 Year 9 pupils (aged 13–14 years) were tracked.

Observations were also collected on comparison pupils. The aim was to observe a sample of typically developing pupils, average in the class in terms of their academic attainment, in order to provide a point of reference for the results on the pupils with Statements/EHCPs. Primary school teachers were asked to identify at least three average-attaining pupils in the class, and one of these pupils was used as the comparator for each lesson observation. Comparison pupils were matched to the pupils with Statements/EHCPs in terms of gender. Observations were collected on 151 control pupils: 115 boys and 36 girls.

In the second phase of data collection in secondary schools, anticipating that many secondary schools set pupils by attainment for core subjects, researchers observed in classes defined as ‘average-attaining’. With guidance from the class teacher, researchers selected
one average-attaining pupil to observe for the duration of the lesson. Comparison pupils were again matched in terms of gender to the pupil with the Statement, who was the primary focus of the school visit. Observations in this phase of the study were collected on 112 average-attaining pupils, again matched by gender: 83 boys and 29 girls.

In the first phase of data collection, researchers visited a total of 45 primary schools across London, the South-East and East of England regions. The majority of schools (84 per cent) were situated in predominantly urban areas. Analyses indicated few differences between the pupil SEND groups, and so results were combined. In the second phase, researchers visited fewer schools overall \( n = 34 \), but there was a greater geographical spread. Most of these pupils at Year 9 (82 per cent) had a Statement/EHCP for needs relating to cognition and learning. Both samples were broadly consistent with the national picture for pupils with SEND.

A total sample of just under 100 pupils with SEND may not appear to be a large sample, but each child was treated as a case study comprising week-long systematic observations, along with semi-structured interviews with school special educational needs coordinators (SENCos), teachers, TAs, each pupil’s parents/carers and, in the SENSE study, pupils themselves, as well as documentary evidence. Given the range and intensity of data collected, data collection and analysis were extremely labour intensive. Taken together, the MAST and SENSE studies probably represent the largest systematic observational research project on the everyday behaviour and interactions of pupils with SEND ever conducted in the UK.

Findings from the case studies are presented in Webster and Blatchford (2015, Blatchford and Webster 2018). In this book, we report results from the study’s main method of data collection, the systematic observations.

Three types of data collection

Finn et al. (2003) argue that methods for studying classroom processes related to class size have tended to be anecdotal and informal, with few systemic observation studies designed to provide systematic information on classroom behaviour and interactions. The aim of the CSPAR KS2 stage was to build on what we had learned from the KS1 stage and provide systematic information and advance understanding of class size effects. Although, as we have seen, there were a number of different
forms of data collection, in this book we draw on three main sources of
data collection, now described in more detail.

1. Teacher questionnaires

A main form of data collection is from practitioners themselves, as
expressed in the open-ended responses to the CSPAR annual national
questionnaires completed by teachers, headteachers and TAs. In this
book we particularly focus on responses to the teacher questionnaire
(TQ).

The annual questionnaires from teachers (the TQs) were
returned by 486 teachers altogether, 206 in Year 4, 184 in Year 5 and
96 in Year 6. There were also annual questionnaires completed by TAs
(340 altogether across the three year groups) and headteachers (437
altogether across the three year groups). In each of these three school
years, there were a set of questions which asked teachers to comment
on whether, and if so how, the number of children in their class size
had affected their teaching, pupil learning and behaviour, grouping
practices, pupil relations with other pupils, and classroom tasks and
the curriculum. These questions were first trialled and then adapted
where necessary where there was ambiguity or uncertainty amongst
respondents. In some cases, questions in later years were adapted
following answers and feedback from earlier analyses, for example, to
split a question into two to get more precise information. In general,
however, the aim was to maintain the same questions each year to
enable comparisons between, and pooling of, information over time.
Other features of the specific questions asked are described in the
relevant chapters. Answers to these questions were used as a key source
data for the separate chapters on class size and classroom processes
which constitute the core of this book.

For the analysis, all the responses were typed out verbatim and
about a 10 per cent sample of teacher responses used to carefully
devise a coding frame for application across the three years. All the
quotes were sorted into key categories (‘codes’), and the reliability of
this coding frame was tested by the extent to which separate coders
agreed on a subsection of TQ material. Agreement was higher than
80 per cent. These categories referred to the effects of both small and
large classes, for example, the effects of class size on the individual
attention received by a pupil were often increased in a small class but
decreased in a large class.
In the book we make extensive use of verbatim quotes from teachers to support and illuminate the conclusions drawn from the data.

2. Systematic observations

The second form of data collection was from systematic observation studies of classroom interactions and behaviour, in the CSPAR, DISS, MAST and SENSE studies. Systematic observation allows researchers to take representative observations of the classroom at regular intervals. Subsequent statistical analyses conducted on the large datasets provide an objective description of the main features of everyday classroom life, often unavailable to received opinion.

Systematic observation is a technique that is not without its critics (Delamont and Hamilton 1986) or defenders (Croll 1986; McIntyre and Macleod 1986). A main feature of the systematic observation method is the use of a category system determined prior to data collection, with explicit and rigorous definitions, and criteria for classifying behaviour and contexts. The strength of the method is its scale and limited susceptibility to inflection and interpretation by individual observers. It provides a relatively straightforward though labour-intensive means of obtaining descriptive quantitative data; if done well, the category system is reliable, which is usually assessed by the degree to which independent observers code behaviour in the same way. The trade-off, however, is that the pictures of classroom life painted using these data are typically achieved using broad strokes.

In the CSPAR, DISS and MAST/SENSE studies, categories were developed to code pupil behaviour and interactions in classrooms. As we shall see in Chapters 3 and 4, these were typically high frequency and easily recorded behaviours. The category system was based on extensive pilot work to ensure they were suited for the purpose, and accurately described teacher and pupil behaviour – both important indices of ‘validity’, an important facet of research.

More details on the CSPAR coding system are given in Chapters 3 and 4. We adapted a systematic observation schedule that had been developed by PB in previous research (Tizard et al. 1988). This involved a time sampling technique (see Croll 1986) in which 5-minute bursts of observation per pupil were divided into 10-second time units, followed by a short interval for coding. The schedule had categories describing how children behaved in three ‘social modes’: when with their teachers, when with other children, and when not interacting. Subcategories
within each of these three modes covered work, procedural, social and
off-task activities.

Systematic observations were conducted during the first year of
school (the reception year, 4–5 years old – these results are reported in
Blatchford et al. 2003b, and Blatchford, Moriarty et al. 2002b), but in
this book we report results from the second round of systematic observa-
tions when the pupils were in Year 6 (10–11 years). Classes were selected
on a random basis from class size information supplied by the school. It
was decided to select schools with classes in two bands: large (31 and
over) and small (25 or under). These bands were chosen to be repre-
sentative of the two ends of the class size distribution, and were slightly
different to those used in the reception year (20 or under versus 30 and
over), when class sizes are typically smaller. There were 42 classes in all,
16 small and 26 large.

Teachers were asked to select nine pupils, three from each ability
range – low, medium and high. Six of these ‘target’ children were then
chosen by the researcher, two from each ability band, one girl and one
boy. If a child was absent for more than a day they were replaced by a
‘reserve’ drawn from the nine. In some cases there were observations on
more than six children (for example, the originally selected child and
the reserve). Of the 257 children in all, 128 girls and 129 boys, 83 low
ability, 87 medium ability and 87 high ability.

The basic principle was to observe when classroom-based activities
could have taken place. Observations were not conducted during parts
of the day when all the pupils went out of the classroom. The aim was
to observe each child over two days. There were 22,312 observations in
total, with an average of 87 observations per child.

In the DISS study, systematic observations were carried out over 2005/6
in 49 mainstream schools. Details of the pupils observed were given
above. These schools were chosen at random from a national survey as
part of the DISS project and they then agreed to fieldwork by researchers.
There were 27 primary schools and 22 secondary schools. Two year
groups were generally observed in each school, either Year 1 and Year 3
(5–6 and 7–8 years) or Year 7 and Year 10 (11–12 and 14–15 years). Obser-
vations were conducted in 88 year groups.

The observations were on a sub-sample of eight pupils per class.
Information on the level of pupils’ special needs status (taken from the
forms completed by school staff during observation visits) was used to
classify the sample into three groups for the purposes of analysis: (1) no
special needs, (2) School Action and (3) School Action Plus/Statement.
The last two categories were combined to help balance numbers of pupils in groups and also because by definition they were the highest level of special need. This classification into three groups was used because of the obvious way in which it affected the amount of support received, as well as its likely effect on learning and attainment. For convenience, the three groups will be called ‘no SEN’, ‘School Action’ and ‘SEN’. To be representative of the attainment levels of each class, pupils were also classified into three attainment groups – low, medium or high – based on a classification made by the teacher. There were 686 pupils observed in total. There were 67,928 observations (data points) in total, collected over 1,132 hours of observation.

Further details of the two studies can be found in following chapters, and also in Blatchford and Webster (2018).

While the CSPAR study classified class sizes into large and small, a distinctive feature of the DISS study was that for each 10-second time interval the observers noted which of the observation categories occurred, and they also noted the class size at that time. Class size was therefore treated as a continuous rather than a dichotomous variable. Though time consuming to prepare, this allowed a more powerful statistical analysis. Another advantage is that class size is not measured in terms of some general figure on a class register, but rather in terms of the exact number of pupils in the classroom at the time of a given observation. This is much more accurate than the more obvious and easier method of examining associations between an average class size and totals of behaviours across all observations for each pupil.

As with the CSPAR study, observations were conducted on each child in turn in blocks of 10×10-second duration time intervals, with short gaps between observations to allow recording of what took place in the observation 10-second period. Visits lasted between two and four days per school and observations were made in maths, English and science.

The coding system used in DISS is described in Chapters 3 and 4. Each observer was carefully trained in the use of the categories so that observation data is reliable, and this is addressed by the extent to which independent observers are in agreement about the codes for the same behaviour, with a minimum criterion of 80 per cent usually seen as needed. This was achieved in the DISS study; full details can be found in Blatchford et al. (2011b).

The statistical analysis was based on the 10-second observation interval as the unit of analysis. This enabled a powerful and sophisticated analysis of the co-occurrence of behaviours and class size and
allowed us to establish the extent to which a given behaviour occurred in a 10-second time interval with a particular class size. We say more about this in Chapters 3 and 4.

We now turn to the MAST and SENSE studies. Details of pupils observed were given above. The observation categories were similar to those used in CSPAR and DISS, though modified. Details of those used in this book are given in Chapter 9.

Researchers observed for the first 10 seconds of each minute, then for the rest of the minute they coded the interactions, activities and contextual information in operation during those 10 seconds. In primary schools, where pupils with and without SEND tended to be taught in the same class, researchers ensured that every fifth minute of each observation, the focus moved to the comparison pupil. The procedure was similar for the secondary sample, but observations on comparison pupils were made in classes the schools defined as ‘average-attaining’.

Analyses are based on a large dataset, totalling 67,928 observations (data points), collected over 1,132 hours of observation. Researchers collected 57,467 observations (958 hours) of pupils with SEND: 30,782 (513 hours) in primary; 26,685 (445 hours) in secondary. As the main focus of each study was the pupils with Statements/EHCPs, fewer observations were collected on comparison pupils. In total, 10,461 observations (175 hours) were collected on pupils without SEND: 4,233 (71 hours) in primary; 6,228 (104 hours) in secondary.

Inter-rater reliability checks were calculated for the main mutually exclusive categories and showed a consistently high or very high agreement (see Blatchford and Webster 2018).

When totalled across the three studies, that is, CSPAR, DISS and MAST/SENSE, we had over 100,000 separate observation data points in which each had a measure of class size and for pupil behaviour in classrooms. The scale of the data collection makes this unique as a source of information on class size effects; we are not aware of a more substantial observation programme specifically directed at class size.

3. Case studies

The third main method of data collection used in this book stems from the CSPAR project case studies comprising interviews with practitioners (teachers, headteachers and TAs) and pupils, semi-structured observations and documentary evidence.
As we see in more detail in later chapters, case studies were conducted in 10 schools in Year 5 classes (9–10-year-olds) and 10 schools in Year 6 classes (10–11-year-olds). Of the 10 classes, and in order to represent the situation in large and small classes, five small (25 pupils or less, average 20) and five large (31 pupils or more, average 33) classes were chosen at random from the list of classes for those years. Each case study involved semi-structured observations and interviews with 30 pupils, 10 teachers, and 10 TAs each year, overall. Nine pupils were selected by the teacher in each class, three above average, three average and three below average in attainment levels, and then one child from each group was observed and interviewed by the researchers.

The aim of the case studies was to provide a complementary and detailed portrayal of individual classes, which would provide the basis for a more interpretive and grounded analysis of factors relating to class size differences. Selected aspects of classroom learning and experience, expected on the basis of findings from other methods of data collection to be connected to class size differences, were defined in advance, and were then refined on the basis of pilot field visits into headings (which included grouping practices, tasks and curriculum, and teacher–pupil interactions), which structured data collection. In this way whole class and selected child observations in terms of event sampling of significant events; semi-structured interviews with teachers, teaching assistants and pupils; end of session/day comments and judgements by the field worker, were all organised in terms of the main headings. Case studies were led by AR and conducted by members of the research team who were also experienced teachers as well as field workers. Quite deliberately, the aim was to marry aspects of systematic observation (which emphasises the objectivity of data), with professional and interpretative judgements by experienced teachers.

Further features of the methods used

A multimethod study

The approach adopted by the study and this book is therefore multimethod. This kind of approach has become commonplace these days in educational research, but one needs to be careful (Symonds and Gorard 2010). In general terms, we have sought to strive for ‘methodological integrity’ in mixed method research as set out in a recent paper for the American Psychological Society Publications and Communications...
Board Task Force (Levitt et al. 2018). They argue that qualitative and mixed method research should be judged in terms of, for example, the selection of procedures that usefully answer their research questions and address their aims and considering findings in their context – for instance, their location, time and cultural situation.

The fact that more than one method is used does not of course guarantee any special advantage to a study. Although it is sometimes assumed that a combination of methods makes up for limitations in individual methods of data collection, we think this is misguided. Combining methods successfully depends fundamentally on the strength of the individual methods of data collection and how well they are used together.

As we say in the course of the book it was not possible to use data from all three forms of data collection for all classroom processes. This is because we did not have, for example, systematic observation data to complement results from the TQs or the case studies. In Chapters 6 to 8 we have relied primarily on evidence from the TQs and the case studies. The findings and conclusions for these chapters are therefore more limited and will need to be treated more cautiously as a result.

In this book we have for the most part first presented results from each method of data collection separately. For example, in Chapter 4, where we could draw on all three forms of data, we present results on class size and teaching in the sequence: teacher questionnaire (TQ), systematic observation and case studies. But the three methods of data collection were also deliberately designed to mutually inform each other and be conceptually integrated. One way this integration works is in terms of the categories stemming from results from one method of data collection being used to extend and inform the methods and themes used to analyse another method of data collection. An example of this, as we have seen, is the way the structuring themes in the CSPAR KS2 case studies were informed by emerging results from the TQ and systematic observations. The long-term nature of the study also meant that key structuring categories for the interviews and semi-structured observations in the KS2 case studies were constructed around results from the earlier results from the study.

The long-term nature of the research also helped the development and piloting of methods of data collection. Methods of data collection first developed in the KS1 stage were extensively used and evaluated and were revised where appropriate. It was also possible to adapt methods in the light of emerging results. Another connected feature was
the presentation of early results to teachers and our advisory groups. We were able to discuss provisional findings with them and together work through some possible explanations and possible changes to data collection techniques, and identify additional information needed in future data collection.

The integration of the three methods of data collection is most obviously seen in the way that overarching themes are developed from careful cross examination of each method of data collection. To give an example, in Chapter 4 a section of results from the TQs was mutually informed by results from the systematic observations to show how, from different points of view, class size had important consequences for the balance between three interactive contexts for teaching – individual, group and whole class. We shall see how the TQ results were then analysed to extend these results with a broader and more qualitative analysis of ways teaching was affected by class size. This, in turn, was further supplemented by results from the case studies of large and small classes in schools. This shows the progressive and iterative approach that was used to ensure that the full analysis made a whole that was more complete than the sum of its parts.

Another feature of the method of data collection has been the way we deliberately collected data on the practitioner experience. The teacher-completed questionnaires and interviews with teachers conducted as part of case studies in schools provide a valuable insider view on classroom life. We argue that these are valuable in their own right. It is important, we believe, to systematically address practitioner experiences, because they have privileged access to their own teaching and classroom management, and that of colleagues. This enables them to draw on moment by moment experiences of the role of class size, as well as reflections over a block of time. We also recognise, however, that there can be issues of validity and reliability connected to the use of teachers’ perspectives and reported experiences, and that ideally these need to be combined with other forms of data collection. One of the strengths of the multimethod approach used in this book is that findings and themes can be cross checked against each form of data collection. It is important to cross check and, if necessary, seek to reconcile practitioner experience and results from the systematic observations. This is not least because of findings from, for example, Shapson et al. (1980) who showed that the teachers’ views about the effects of class size were not always supported by the results from their systematic observation analysis.
Interpretation of findings

The interpretation of the data is no less important than the methods of data collection. A strength of this study, we believe, is that the care that went into collecting and analysing data is matched by that taken in its interpretation. The long-term nature of the research has again helped. There have been a number of peer reviewed articles stemming from the KS1 stage of the research, as we have seen, which provided feedback and comments. This, in turn, helped provide the basis for the conceptual and empirical work over the later KS2 stage, described in this book. Moreover, in regular meetings of the large team on the project we carefully examined findings over time to ensure that the emerging conclusions were fully grounded, tested and coherent. There were also many presentations of results to various audiences, both academic and practitioner, and much was learned about the strengths and weaknesses of our interpretations and suggestions, along with possible alternative or additional ways of thinking about them. It is this interpretation that we present in Chapter 10.

Leverhulme International Network

The research programme was further supplemented by a Leverhulme-funded ‘Class Size and Effective Teaching’ International Network, led by Peter Blatchford (2015–18). This involved a group of ‘network partners’ and stemmed from a belief that such an international network would help advance understanding of educational effects of class size and how teachers can make the most of smaller (and larger) classes. It brought together prominent researchers from universities in the UK, mainland Europe, East Asia and the United States, representing countries of strategic importance in this field. It also addressed the neglected topic of effective pedagogies in different sizes of class, something which has helped inform the discussion in this book. All those involved in the network contributed to a recent edited book Class Size: Eastern and Western Perspectives (Blatchford et al. 2016b) and a Special Edition of the International Journal of Educational Research (Blatchford and Russell 2019). The project involved three workshops and the final one of these – in London in 2016 – was extended to include local authority officials, headteachers, teachers, teacher union representatives and academics. There were a number of roundtable discussions, the conclusions of which were written up and have also helped inform this book.
As described above, in the following chapters we present our findings on class size and classroom processes. In Chapter 10, we return to the more general perspective on class size and provide a summary description of our findings.

First, in the next chapter, we will look more closely at the evidence on class size effects on pupil outcomes.