REVIEW

October 2004

Secondary school size: a systematic review

The EPPI-Centre is part of the Social Science Research Unit, Institute of Education, University of London

AUTHOR AND INSTITUTIONAL BASE

This work is a review completed by Zoe Garrett, Mark Newman and Diana Elbourne of the EPPI-Centre, which is part of the Social Science Research Unit at the Institute of Education, University of London; Steve Bradley and Jim Taylor of the Department of Economics at Lancaster University; and Anne West and Philip Noden of the Centre for Educational Research, Department of Social Policy, London School of Economics and Political Science (LSE). Additional support in completing the review was provided by Lorraine Fincham, Veena Meetoo, Jennifer Sinclair and Tom Korolowicz.

Institutional base:
EPPI-Centre
Social Science Research Unit
Insititute of Education, University of London
18 Woburn Square
London, WC1H 0NR

REVIEW TEAM MEMBERSHIP

Zoe Garrett, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Mark Newman, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Diana Elbourne, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Steve Bradley, Department of Economics, Lancaster University
Jim Taylor, Department of Economics, Lancaster University
Anne West, Centre for Educational Research, Department of Social Policy, LSE
Philip Noden, Centre for Educational Research, Department of Social Policy, LSE
Jennifer Sinclair, Research Associate, Institute of Education, University of London
Lorraine Fincham, Research Associate, Insititute of Education, University of London
Veena Meetoo, Research Associate, Institute of Education, University of London
Tom Korolewicz, Research Associate, Institute of Education, University of London

ADVISORY GROUP MEMBERSHIP

Caroline Macready, Department for Education and Skills
Audrey Brown, Department for Education and Skills
Judy Sebba, Department for Education and Skills (until 2004)
Anthony Zacharzewski, HM Treasury
Will Straw, HM Treasury

ACKNOWLEDGEMENTS

We thank the DfES for their financial support for this review and for the education stream of the EPPI-Centre. We would also like to thank Lorraine Fincham, Tom Korolewicz, Veena Meetoo and Jennifer Sinclair for help with keywording and data-extraction; Harvey Goldstein for statistical advice; Ann Oakley, David Gough, Jo Garcia, Judy Sebba, Victoria White, Caroline Macready and Anthony Zacharzewski for comments on earlier drafts; and Ripa Haque for administrative support. Two members of the review group (SB, JT) declare their interest as studies authored by them were in the in-depth review, but assessment of these studies was carried out by other members of the group.

LIST OF ABBREVIATIONS

Add Health	National Study of Adolescent Health (USA)	
CCD	Common Core of Data	
DfES	Department for Education and Skills (UK)	
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre	
GCSE	General Certificate of Secondary Education (aged 15-16)	(UK, Not Scotland)
GNVQ	General National Vocational Qualification	
KS2	Key Stage 2: 7-11 years old	
KS3	Key Stage 3: 11-13 years old	
KS4	Key Stage 4: 14-16 years old	
LEA	Local education authority	
LOLSO	Leadership for Organisational Learning and Student Outcomes (Aus)	
NAEP	National Assessment of Educational Progress (USA)	
NCES	National Centre for Education Statistics (USA)	
NELS	National Educational Longitudinal Study (USA)	
OLS	Ordinary least squares	
OECD	Organisation for Economic Cooperation and Development	
PISA	Programme for International Student Assessment	
SASS	Schools and Staffing Survey (USA)	
SEA	State Education Agency (USA)	
SEN	Special educational needs	
SES	Socio-economic status	
STAR	Standardised Testing and Reporting (USA)	
WASL	Washington Assessment of Student Learning (USA)	
WoE	Weight of evidence	

This report should be cited as: Garrett Z, Newman M, Elbourne D, Bradley S, Noden P, Taylor J, West A (2004) Secondary School Size: A Systematic Review. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

TABLE OF CONTENTS

SUMMARY

Background

The question of the optimum size of school has received considerable attention in recent years. In England, the introduction of quasi-market conditions in English secondary education theoretically allows schools to expand or contract in size in accordance with parental preference. In the USA, there is a growing 'small schools advocacy' movement which has a high media and political profile. Whilst the research evidence base appears to be quite large, it is disorganised, often partial and characterised by conceptual, practical and methodological differences.

Aims

- To produce a systematic map describing the range of research investigating the impact of school size on a range of student, teacher and school outcomes
- To produce an in-depth review focusing on comparing outcomes between schools of different sizes*
- To consider implications from the review in terms of research, policy and practice

Methods

Due to the restrictive timeframe available for the project, the majority of studies were identified through searching bibliographic databases. There was no systematic use of personal contacts, websites, journal handsearching, or citation-checking. Criteria were used to restrict the included studies to those which contained empirical data, investigated outcomes which included school size or schools-within-schools; included a variable for school size; were written in English; were conducted in an OECD (Organisation for Economic Cooperation and Development) country; and were published post-1980. Included studies were keyworded, using both generic and review-specific keywords to create a 'map' of the research literature. For the in-depth review, a further set of criteria was applied to the studies in the map.

Studies were excluded from the in-depth review if the focus was schools-within-schools; the number of schools in the sample could not be ascertained; data were collected before 1990 (except where they were collected over a time span that included 1990), the analysis did not control for socio-economic status (SES), or the study did not focus on one or more of the following outcomes: (i) student attainment and progress, attitudes, behaviour (ii) teacher morale and experience, (iii) school organisation, management and costs, or the sample comprised only higher attaining or advanced students. The studies in the in-depth review were subjected to generic and review-specific data-extraction, including assessments of

* Such comparison may imply causality. The research is, however, based on study designs which are appropriate for considering associations rather than effects. Hence, where we have used the word effect, we have put it in inverted commas ('effect') to indicate the problematic nature of concept in this context.

the weight of evidence (WoE) each study lent to the review. Quality-assurance was carried out at the screening, keywording and data-extraction stages.

Results

Nearly 4,000 reports were identified as potentially relevant to the review. These were screened against the inclusion criteria. 3,503 reports were excluded and 252 were not available in time for the review. The remaining 134 reports of 119 different studies were then keyworded. The final map included only nine studies from the UK and showed a scarcity of relevant studies looking at the range of outcomes of interest for the review. In particular, there were too few for a separate analysis of the schools-within-schools literature.

Thirty-one studies were included in the in-depth review. Two-thirds of the 31 studies were from the USA and one-fifth from England. Nine of the 31 studies were judged to give high/medium WoE to answering the review questions, and five were judged to be low or low/medium. The majority of studies examined the 'effects' of school size on achievement without controlling for prior attainment (N=15); four studies examined achievement whilst controlling for prior attainment; 13 studies examined student attitudes and behaviour; five examined economic outcomes; two examined school organisation outcomes and two examined the perceptions of teachers.

Relationship between school size and achievement without prior attainment*

- Of the 15 studies that do not take into account prior attainment when considering the 'effects' of school size on achievement, approximately half show a positive relationship and half show a negative relationship with school size.
- The only English study in this category found that achievement increases as school size increases up to approximately 1,200 (for 11-16 schools) or 1,500 (for 11-18 schools) students (a quadratic relationship). After this point, achievement decreases as size increases.
- The majority of these studies do not report any statistically significant† association between school size and achievement.

Student achievement whilst controlling for prior attainment*

- Four studies in this section found that student achievement increases as school size increases up to a particular point (or range). After this point, student achievement decreases as school size increases.
- The point estimate or range of school size at which achievement is maximised varied within and between these studies. The optimum school size estimate ranged from 600 to 2155 and the optimal year group size ranged from 150 to 225.

Student behaviour and attitudes

- One study with a high/medium weight of evidence found that overall absence was lower in schools with up to approximately 1,400 students. After this point, overall absence appeared to increase as school size increased.

---

* Student achievement in the studies in this review is confined to exam performance.

† The concept of statistical significance in this context is not unproblematic (see sections 1.4, 4.5 and 5.2.2).

- Studies with a lower weight of evidence tended to find a negative association with school size: that is, an increase in the dropout rate and a reduction in attendance as schools get larger.
- All the studies considered found a negative association between students' feelings of engagement, connectedness and participation, and increased school size.
- Studies investigating the relationship between school size and violent student behaviour had somewhat contradictory findings. Some types of violent behavior increased as school size increased, whereas other types of violent behavior decreased as school size increased.
- The results suggest many of these relationships are comparatively weak, and are difficult to quantify and conceptualise.

Teacher perceptions of school climate and organisation

- Comparatively few studies included in the in-depth review included measures of the impact of school size on teachers.
- Results suggest that teachers in smaller schools tend to have more positive perceptions of school climate, of their abilities to influence school policies and control their classrooms, of school norms; teachers also perceive greater co-operation and more resource availability.

School organisation and structure

- Comparatively few studies in the in-depth review included measures of the impact of school size on elements of school structure and organisation. The two studies that did include such outcomes used very different measures and are not comparable.
- One study found that, as school size increases, so too does the *construct* of class size based on average class size, student teacher ratios and teachers' perceptions of satisfaction with their class size.
- The second study found that students in smaller schools may be more likely to be entered into higher tiers for mathematics, but not for science, and that some students in smaller schools may be less likely to be entered for some GCSE subjects. However, this pattern was not consistent across different subject areas.

Economic outcomes

- The studies in this category show a consistent negative relationship between average secondary school size and costs defined as direct public expenditure on schools.
- One study, despite finding the same negative relationship with average secondary school size, found that the relationship between size and costs was different between schools of different types. Relationships based on 'average' cost figures for all schools may not apply to particular types of school.
- The size of the relationship between average secondary school size and costs differs slightly between studies. An increase in school size of 10% is estimated to reduce costs per student by between 1% and 4%, depending on the definition of cost used.
- Studies of economic outcomes have considered only a limited range of costs (direct public expenditure on schools per student) and a limited range of outputs or benefits (cost per graduate, inefficiency).

Conclusions

The review question is concerned with the overall relationship between secondary school size and outcomes. At this macro- or global level, the review findings suggest that there is no overall consistent relationship between secondary school size and outcomes. However, at the level of the individual outcomes, the pattern of findings which emerged suggests that we can be reasonably confident that exam attainment is maximised and absence is minimised at a certain point in the range of secondary school size. Further, in the limited terms of expenditure per student, costs decline as schools get larger. However, they also suggest that teacher and student perceptions of school climate decline and some kinds of violent behaviour may increase. The design of the studies included in this review cannot definitively establish causal relationships and thus the direction of causal relationships is a problem for all the outcomes reported. Does the number of students determine cost or does cost determine the number of students? Does school size determine attainment or does attainment determine school size?

There are three key issues which remain more unclear than the directions of 'effects' results suggest. Firstly, even if the interpretation given above is accepted, to be of practical use we would need to know at what size attainment was maximised and/or absence minimised. The studies here do not provide a clear answer to this as the range reported is quite wide, especially in relation to the actual size of secondary schools in England.

Secondly, does the 'average' direction of 'effect' apply to all school/student types? There is some suggestion from the studies in the review that it may not, although there may be differences between the USA and England on this point.

Thirdly, we would want to know the 'effect' of planned or unplanned changes in the size of an individual school. Such an analysis would need to include not only the 'effects' on the school that had changed size, but also 'effects' in neighbouring schools.

This review would seem to refute some of the more prevalent myths regarding the advantages and disadvantages of smaller and larger schools. For example, that student achievement is universally higher in smaller schools and that student behaviour is universally worse in larger schools have been shown to be inconsistent with the current evidence. The relationship appears to be much more complex than such simple arguments suggest.

Strengths and limitations

Strengths

The main strength of the review lies in its systematic and comprehensive nature. The process of systematically identifying, screening and critically appraising the studies helps to ensure that the review process is transparent, replicable and updateable. Another strength is the presentation of the review results in terms of directions of effect, which facilitates direct comparison across studies with similar outcome measures for perhaps the first time in this topic area. Another important strength is the involvement of the commissioners of the review, especially at the point of moving from the map to the in-depth review. This helped to make the review more policy-relevant.

Limitations

The remit of this review extended only to a consideration of studies that investigated empirically the association between an outcome variable and school size. These were all quantitative studies. This meant that qualitative studies that investigated in more depth the *processes* whereby school size might be related to differing socio-cultural and organisational climates, or staff, student and community relationships were not included. This is a limitation imposed by the agreed focus of the review question rather than the review process itself, but means that little contribution is made to discovering why school size might affect outcomes.

The review process itself had a number of limitations. The truncated form of searching that was carried out because of the restricted timescale for conducting the review (with the cut-off date for retrieval of reports), may have resulted in missing some relevant studies, although it is difficult to estimate the extent of this problem. Since the application of inclusion criteria, keywording and data-extraction were carried out by two reviewers independently in only a sample of cases, the possibility of reviewer error was greater than if all these procedures had been carried out independently for all studies. However, the information extracted from the papers was continually being re-examined by different members of the review group during the process of analysis and synthesis, thereby minimising the risk of error and improving the data quality.

Most of the studies identified for inclusion in this review were taken from USA state data. Within the USA, there is much wider variation in the size of school, and differences in the socio-economic and cultural contexts of schooling. Taken together, these differences may limit the generalisability of conclusions to the UK context. The meaning and use of statistical significance is also difficult to interpret in this review because many of the study findings included all schools in a population as their 'sample'. Another important limitation of the findings is that the individual studies in the review only measured a limited range of outcomes. Attainment, cost and benefit in particular were conceptualised and measured in a limited way.

Implications

This review does not provide evidence to support policy initiatives that solely aim either to increase or to decrease the size of schools and/or to close or change the structure of schools below or above a certain size. Where policy options could have an impact on school size (e.g. through the expansion or retraction of school size through the option of parental choice), it would seem reasonable to make all stakeholders aware that, at some point, the characteristics which make a school appear desirable may be lost if the school's size changes dramatically.

Further research on the relationship between school size and a broad range of educational outcomes is required, using both quantitative and more in-depth qualitative analyses. It is important that future research builds on existing research both substantively and methodologically.

Schools-within-schools may have the potential to offer the benefits of both small and large schools by maintaining several 'small' schools within the same school site.

However, there appear to be few rigorous evaluations of such initiatives. Future schools-within-schools initiatives should be accompanied by rigorous evaluation.

1. BACKGROUND

This chapter describes the aims and rationale of the review, considers the definitional and conceptual issues surrounding the area of school size and examines the policy, practice and research background. Finally it lists the questions addressed by this review.

1.1 Aims and rationale for current review

There is considerable variation between schools in terms of their organisation, social structures, philosophies and pedagogical approaches. Education outcomes also vary considerably within and between students, schools and countries. In trying to understand the relationship between what might be called 'educational inputs' and 'educational outcomes', one focus of educational theorists, activists, researchers and policy-makers has been the role of school size. Such questions have achieved greater contemporary relevance for a number of reasons: the introduction of quasi-market conditions in English secondary education theoretically allows schools to contract and decline in accordance with parental preference; the growing evidence of systematic differences in educational attainment and the growing visibility of a 'small schools advocacy' movement buoyed by the apparent success of a range of school 'downsizing' initiatives in the USA. The evidence base appears to be quite large. However it is disorganised, often partial and furthermore characterised by conceptual, practical and methodological differences.

The review aims to:

- produce a systematic map describing the range of research investigating the impact of school size on a range of student, teacher and school outcomes
- produce an in-depth review focusing on a particular aspect of school size
- consider implications from the review in terms of research, policy and practice

1.2 Definitional and conceptual issues

In this section, we provide a brief overview of policy in relation to school size and the conceptual and theoretical frameworks employed by researchers investigating school size.

1.2.1 What is meant by the 'size' of schools?

There are broadly two approaches to the concept of 'size' adopted in the literature. Most policy-making and much research in the 'school effectiveness' tradition appears to consider 'size' as simply the total number of students attending a school or different sites of a school at a given date. Even within this approach there is variation in the definition of 'size' used: for example, either the number of students on a school's roll or actual attendance on site(s) at any given point in time. Such differences may be important if, for example, different school rolls are associated with different truancy rates. Sometimes the average size of year group (span size) may be used as a proxy for school size and this may have different implications, depending on whether students past the age of compulsory schooling

are included since the presence of this voluntary group of students may be associated with variations both in the size of schools and the quality of the 'outputs' of the schools.

The second approach, which is largely found amongst small school advocates and researchers supporting this position, appears to view 'size' as a combination of the number of students, the physical size of school buildings, a particular set of values and/or decision-making practices, and a particular culture (Muir 2000, Tasker 2003). For this review, we will largely be using the first approach.

1.2.2 Why the interest in 'size' of schools?

A striking feature of educational data and research findings across subjects and disciplines, across age groups, and within and between countries, districts and schools is the great variation in both educational inputs and outcomes (Sherman *et al*., 2003). Education-users, practitioners, policy-makers and researchers are therefore interested in understanding more about the 'causes' of these variations in order to produce policy interventions that reduce these variations. The conceptual and theoretical arguments made for the importance of school size are discussed in more detail later. The issue has achieved greater prominence in recent years for a number of reasons. First, the changes in school enrolment policies in the UK, introduced in the 1998 Education Act, mean that it is theoretically possible that schools could expand or contract in size according to 'market demand'. Second, there is growing evidence of systematic differences in educational attainment. Third, there is greater public visibility of small schools advocacy groups buoyed by the apparent success of a range of school downsizing projects in the USA in particular (see, for example, Cotton, 1996a; Darling-Hammond *et al.*, 2002; Tasker, 2003).

School size is considered as an issue on both the inputs and outcomes (or quality) side of the relationship. There are broadly two questions relating to school size, about which there is considerable controversy. First, does school size make a measurable difference in the quality of education and student achievement? Second, do costs per student vary with school size (McGuire 1989)? In practice, the questions are rarely considered separately. For the sake of clarity, however, it is useful to attempt to consider the conceptual or theoretical premises underlying each question separately.

These questions imply causality: that is, what the 'effect' of the size of schools is on a variety of outcomes. The research is, however, based on study designs which are appropriate for considering associations rather than effects. Hence, where we have used the word 'effect', we have put it in inverted commas ('effect') to indicate the problematic nature of concept in this context, unless we are quoting or paraphrasing from the work of others, or the word is used in a particular technical terms (such as main effects, interaction effects or effect size).

1.2.3 Why might school size affect the quality of education and student achievement?

Many different arguments have been offered as to why school size might affect student achievement. Overall, these arguments suggest that school size affects school organisation and culture in ways that may be more, or less, harmful/beneficial to teachers, students and in some cases the community of which the school is presumed to be a part.

One group of arguments is about how the size of school affects the number and calibre of teachers employed and what the teachers employed are required to do. It is argued that in smaller schools teachers are more likely to be required to teach across diverse subjects, and are thus less likely to be subject specialists and more likely to have to carry out administrative as well as teaching roles. By contrast, it is argued that teachers in larger schools are more likely to teach their specialist subject areas and greater ability grouping of students ('streaming) is possible. These arguments are also employed in the opposite direction. It is argued that the so-called 'efficiency gains' in larger schools, brought about by specialisation and ability grouping, change the nature of the interaction between students, and between staff and students, in ways that are less conducive to learning. Darling-Hammond *et al.* (2002), for example, cite evidence to suggest that, in successful small schools, 'personalisation' of education is important, as are collaborative learning structures, the formation of teaching teams and so on.

A similar set of arguments is made in relation to student participation in school life. It is argued that because larger schools can offer a wider range of extra-curricular activities more students can get involved in such activities. However, the argument is also made that because there are fewer students in smaller schools everybody has to get involved in extra-curricular activities and students cannot avoid being involved in the social community of the school.

Another related group of arguments concerns the ways in which changes in school size brought about through planning or the operation of quasi-market forces affect the quality of education provision. It is argued, for example, that schools that are recognised as 'successful' and growing are able to attract better quality teachers and students, which in turn will further increase the quality of the education provision in those schools. By contrast, schools that decline substantially in size due to losing students to more popular schools may suffer from low morale, especially among the teaching staff who may treat such information as an indication of their esteem in the local community. This is an important issue as it points to the fact that school size may be an 'effect' of education success, rather than a cause. For example, it may be that increased school size is a consequence of being a 'value-adding' school – or perceived by parents as such – rather than enhanced value added, arising from (among other things) school size. (Hence it is important to identify any studies that have differentiated between schools that are full to capacity and those with spare places.) There are also arguments about the value of the wider functions of schools as centres of community networks and the consequences for local communities of school closure. This is where the school size debate intersects with general issues about public provision in rural areas and is a prominent feature of USA school size research.

There is widespread evidence of the importance of family and social factors as predictors of educational attainment (McGuire, 1989) and it is likely that there may be differences in students who attend schools of different sizes. For instance, as size is partly a function of local population density (and therefore travel distances), the intakes to large and small schools may differ substantially in family background and socio-economic status (SES). There is therefore particular interest in investigating the potential impact of school size on the value added to students during their years in secondary education. Since direct measures of value added have only become available very recently in England, there is still very little evidence available about the relationship between school size and value added. Although much of the focus is on the impact of school size on educational attainment, it is also important that the relationship between school size and other 'quality' related outcomes, such as the social and personal development of

students, as well as more visible and more measurable outcomes such as truancy, are investigated.

1.2.4 Why might school size affect the cost of education?

The basic framework for arguments about costs of education in the context of school size is provided by the notion of economies of scale. Scale economies occur when the cost of enrolling an additional student (referred to as the marginal cost) is lower than the average cost at that point, thereby resulting in the average cost declining as enrolments expand. Scale diseconomies occur when the marginal cost of enrolling an additional student exceeds the average cost at that point, thereby leading to an increase in the average cost after that point. It is argued that if it can be shown that after controlling for the quality of educational output costs per student are lower in 'large' schools, there is a case for setting a mandatory 'minimum' school size. Similarly, if average costs are found to rise once a certain level of enrolment is exceeded, then there is a case for limiting the size of schools (McKenzie, 1995). However, this simplified version of the argument belies the complexity of the issues examined. There are many difficulties in conceptualising and specifying production models in education and different models produce different results, making the interpretation of size economies difficult.

A number of approaches have been used to investigate economies of scale in education. Different approaches have different strengths and weaknesses, McGuire (1989) argues that understanding size economies depends a great deal on how the problem is defined as well as the specific technique used. Problems include distinguishing supply from demand-side factors, deciding which 'costs' to include and arranging how fixed costs should be spread.

The educational production function (Hanushek, 1979) has been extensively employed in empirical analyses of the relationship between educational outcomes and a range of inputs into the educational process, including work by Bradley and Taylor in England (Bradley and Taylor, 1998, 2003; Taylor and Bradley, 2000). The fundamental hypothesis is that the value added to the human capital acquired by students due to schooling is determined by four main factors: the initial level of attainment, family background, peer group effects, and school inputs. Implicit in the production function approach to explaining educational outcomes is the view that the primary objective of schools is to maximise positive educational outcomes (such as the academic achievement of students) and minimise the negative outcomes (such as truancy), both subject to available inputs. Focussing on academic achievement for simplicity, the school is therefore treated as if it were a production unit, which uses inputs of teachers and other resources to add value to its annual intake of students (the raw material inputs). Empirical versions of this education production function are almost always estimated assuming that inputs are additive and have a linear 'effect' on educational outcomes. Hanushek (2004) has recently questioned this approach, suggesting that school inputs in particular may have a multiplicative and non-linear 'effect' on educational outcomes.

1.3 Policy and practice background

1.3.1 Policy and practice

McGuire (1989) argues that educational policy in the USA throughout the twentieth century was influenced by the idea that larger schools can offer more comprehensive instructional programmes of greater quality at lower cost than smaller schools. This idea is reciprocally linked with industrialisation and urbanisation. Cited as evidence of the powerful effect of such forces is the dramatic decline in the number of USA schools during the 20th century, relative to the dramatic increase in population (Cotton, 1996a).

Many authors cite the influence of Harvard educationalist James Conant on increasing school sizes in the USA from the 1960s. In his 1959 book, *The American High School Today*, he argues that 'the first priority of many states should be the elimination of the small high school by district reorganization'. However, it should be pointed out that Conant was arguing for an increase in the size of high schools to between 400 and 600 students, not schools of several thousand students, which are common in the USA today (Cotton, 1996a; Cutshall, 2003; McGuire, 1989; Muir, 2000; Tasker, 2003).

The challenge to the virtues of the large school began in the late 1960s and is credited by some authors to Barker and Gump's 1964 book *Big School, Small School: High School Size and Student Behavior*, which concludes that the supposed superiorities of large schools are illusions (Cotton, 1996a). Since then, there has been a growing 'small schools movement' in the United States. The conclusion that 'small is in vogue' (Muir, 2000) is supported by the endorsement of small learning communities by policy-makers and private philanthropists. The 'No Child Left Behind Act ' of 2001 reauthorised funding for the USA Department of Education's small learning communities programme (Cutshall, 2003), and many USA states have legislated to promote the development of smaller learning communities and/or limit the maximum size of schools (Tasker, 2003). A widely publicised programme, run by the Bill and Melinda Gates Foundation, has spent more than $600 million funding the development of smaller learning communities (Cutshall, 2003; Muir, 2000; Tasker, 2003).

The issue of school size appears to have excited less interest amongst policy-makers and educational researchers in the UK. For example, there appear to have been few major studies of school size in England and only a limited 'small school' lobby of comparatively recent origin (Human Scale Education - see Tasker, 2003). This may have been because, until recently, deliberate planning was largely limited to consideration of the requirements for 'viability', local geographical circumstances (such as population sparsity) seemingly being the major determinant of school size. It is also the case that England does not contain any schools on the scale of the 'large' American high school (see section 1.3.2). However, as a result of the establishment of a quasi-market in England's education sector by the 1988 Education Reform Act, the size of schools is determined by a combination of parental choice and the ability or willingness of schools to expand their student capacity. The greater freedom of the schools themselves to determine their size raises additional questions about the 'effects' of changes in school size.

1.3.2 What size are secondary schools?

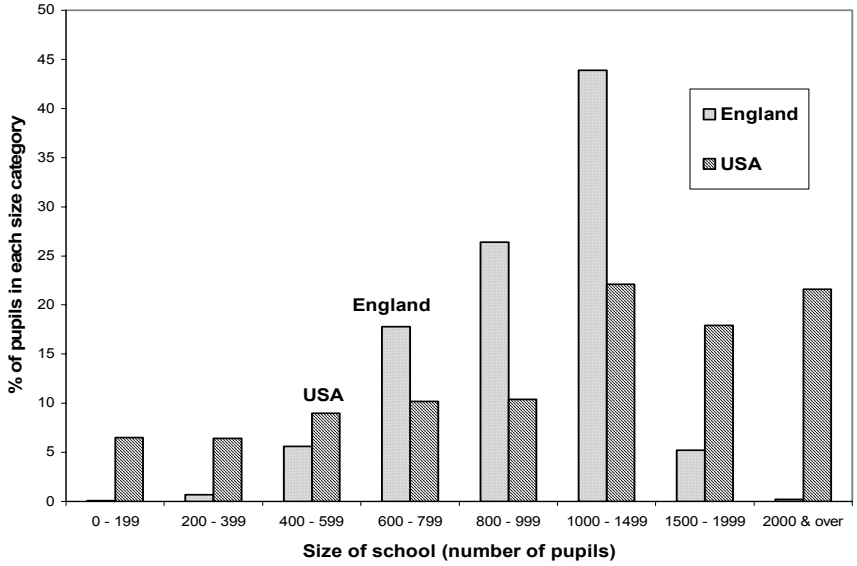
One crucial contextual issue to consider at the outset is what is meant by 'large' and 'small' in the context of schools. Clearly there will be differences, depending on the phase of schooling under consideration: for instance, primary schools are usually much smaller than secondary schools. Different studies use different definitions of 'small' and 'large', which is particularly problematic when comparing studies from different countries.

For publicly funded secondary schools in England (excluding schools admitting only students with special needs), average school size increased from 820 to 1000 between 1992 and 2002. These averages, however, conceal wide variation between schools. In 2002, for example, 29% of schools had fewer than 800 students and 27% of schools had over 1200 students. Eight per cent of secondary schools had 1,500 or more students on their roll in 2003 (DfES, 2003).

The average size of regular USA secondary schools (excluding alternative, vocational and special education schools) has also increased over the past decade, from 684 in 1990/91 to 795 in 2000/01 (NCES, 2003). However, this hides wide variations in school size. In 2000/01 15% of regular secondary schools enrolled more 1,500 students (enrolling 39% of the total student population), whilst 43% of regular secondary schools enrolled fewer than 500 students (enrolling 13% of the student population; NCES, 2002). Muir (2000) reports that in 1998 the largest high school in the United States enrolled over 5,000 students and that there were 274 high schools with enrolments of over 2,750 that had a combined enrolment of 900,000 students.

McKenzie (1995) gives the average size of Australian government 'secondary only grade' schools in 1993 as 710, with 20% having enrolments of fewer than 400 and 7% enrolments of 1,200+.

Figure 1 Size distribution of secondary schools in England (2002) and the USA (2001)



Sources: Schools' Census 2002 for secondary schools in England (excluding schools admitting only students with special needs); USA Department of Education, National Center for Education Statistics, Table 94 for USA 'regular' secondary schools.

1.4 Research background

The issue of school size has been investigated in a number of ways. In quantitative approaches, the relationship between school size as an 'independent variable' and a range of 'dependent' outcome variables is investigated, using different types of multivariate analysis. The complexity of the educational relationship is modelled by including other 'independent' variables that research findings suggest affect educational outcomes. Data are usually obtained from school census returns to government and/or large-scale school surveys.

It is controversial to claim that such studies establish causal relationships, given the necessarily non-experimental nature of the data. Nevertheless, depending on the type and quality of data, statistically significant associations in models are often over-interpreted as suggesting a causal link between the independent and dependent variables, especially if there is plausible theoretical support for such a link.

Both the questions about the 'effects' on quality of education (including student achievement) and costs are tackled using variations on this approach. Based on relevant theories and on pragmatic considerations such as availability of data researchers decide which variables to include in their models, make assumptions about the nature of the relationship between these variables (e.g. linear, quadratic, multiplicative) and select an appropriate method of statistical analysis. As well as the methods of economic analysis discussed in section 1.2.4, methods such as partial correlation analysis, linear regression, multiple regression, logistic regression and multi-level modelling are commonly used. Multi-level modelling recognises the nested nature of the educational process and allows for the possibility that factors come into play at each level of the hierarchy. In education research, the three levels usually employed are as follows: firstly, the individual student; secondly, the grouping of students into classes; and thirdly, the school (Aitken and Longford 1986). In addition, levels 'higher' than the school, such as district/local education authority (LEA) or state are sometimes also used. Interpretation of the results of such studies is complex. The use of more sophisticated models and statistical approaches may help to control some of the error arising from random variation but issues of non-random sampling and measurement error may also be problematic. In addition, results are rarely reported in the form of standardized effect sizes which makes interpretation of the size and thus potential importance of any 'effect' problematic.

Whilst the large-scale quantitative approach may offer the promise of generalizability, one disadvantage of the approach is that it often fails to provide sufficient detail of the mechanism or processes by which school size affects outcomes. The advantage of qualitative approaches is that it is possible through case studies of particular schools or LEAs to provide a very detailed description of the relationships between school size and other factors on educational processes. The downside of adopting this approach is that it is difficult to generalize the findings of the research to other schools or LEAs. It is reasonable to suppose that there may be different organizational, curricular and pastoral implications associated with different school sizes and also with changes in school size. Qualitative research may be particularly helpful in identifying contextual factors affecting school size, the preoccupations of key actors (such as headteachers, teachers and students) in schools of different sizes and, through observational studies, qualitative differences between schools of different size.

There appears to be a large body of published work that considers the issue of school size. However, much of this work is not empirical, and both the empirical

and non-empirical work often appears to adopt a particular standpoint in its consideration of the issues. Much of the literature is from the USA where a large number of what, by English standards, would be called very small and very large schools are comparatively common (see Figure 1), and where the issue of school size appears heavily entwined with issues of rurality, urban decline, school-funding systems and discrimination.

There do not appear to have been any previous *systematic* reviews of the issue. The research study on school size in England by Spielhofer *et al.* (2002) does not include a systematic review and contains only a limited review of previous studies. The reviews by Cotton (1996a) and Tasker (2003) are not systematic, appear to consider only USA-based studies and literature, and explicitly take a partial view. Two overview articles by McGuire (1989) and Muir (2000), whilst not systematic, take a less partial approach to the issue.

McGuire (1989) argues that, generally there is agreement that unit costs are higher in the smallest and largest schools: that is, there is a U-shaped average cost curve. However, moving beyond this to attribute this difference solely to the size of the school, or to applying cost-benefit analysis, is fraught with complexity. Different results appear to be obtained, depending on which variables and assumptions are built into the model used. A study of high school costs in New York demonstrates the potential importance of the selection of outcome measures. The difference in the budget per student between 'small academic' schools and 'large' high schools was estimated to be $15,000. However the estimated difference in budget per graduate was estimated to be $1,500 because smaller schools had higher graduation rates (Stiefel *et al.,* 2000).

McGuire (1989) argues that, in the literature on the relationship between school size and quality one finds reports of negative, positive and negligible relationships. In his view, these differences stem from differences between studies in the way the questions are asked and units of analysis chosen. More recently, Muir (2000) concurred with this view, pointing out that studies which have attempted to compare student attainment in schools of different size have come to different conclusions: some proclaiming to demonstrate higher attainment in 'large' schools, some in 'small' schools and others no difference. However, Muir (2000) does conclude that there is evidence that demonstrates that 'smaller' schools may provide more conducive learning environments – in particular, for students from disadvantaged groups. However, he appears to have based his conclusions solely on studies identified in the review by Cotton (1996a) which, as noted above, was not systematic and started from a 'small school' advocacy position.

1.5 Authors, funders and other users of the review

This chapter has described the conceptual, policy, practice and research background of school size issues. The review was funded by the Department for Education and Skills (DfES) and by the Treasury, both of which have an interest in developing appropriate education policies for schools in England. The research team at the EPPI-Centre had expertise in systematic reviews in education and worked with others who had previously carried out research on school size issues. The initial scope of the review was set by the funders and refined in collaboration with the research team. A complete list of the funders' representatives and of the research team can be found in Appendix 1.1.

1.6 Review questions

The issues raised in the previous sections have been conceptualised as the following questions to be answered by this review:

What are the characteristics of the empirical studies that have investigated the relationship between secondary school size and various 'outcome' variables?

What are the results of empirical research conducted in OECD countries since 1990 that compares outcomes between secondary schools of different sizes?

What are the implications from the review in terms of research, policy and practice?

These are explored further in Chapters 3 and 4.

2. METHODS USED IN THE REVIEW

This chapter outlines the methods used in the review, including the methods of user- involvement. Initially, it describes the identification of relevant studies, including the searching and screening processes. It then describes the methods of creating a systematic map of the research activity and the process of going from the map to the in-depth review, including appraisal of study quality and methods of synthesis.

2.1 User-involvement

2.1.1 Approach and rationale

Policy-makers at the Department for Education and Skills (DfES) and the Treasury were considered to be the primary users of this review. The findings may, however, be of interest to other policy-makers, such as local education authorities (LEAs). In addition, students, their parents and teachers will also be affected by policies about school size.

2.1.2 Methods used

Given the restricted timescale for this review, only representatives of the primary user group were involved in the review. The DfES commissioned the review and provided an outline commissioning brief. Members of the both government departments were asked to comment on the initial map. They came to a meeting in October 2003 with the full review group and helped to guide decisions about moving to the in-depth review. They were also asked to comment on emerging drafts in December 2003 and January 2004 (see Appendix 1.1 for details of the review and advisory group.)

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

Studies were included in the review if they met the following inclusion criteria:

1. Focus on schools or schools-within-schools.
2. Focus on secondary education (11-18 years)
3. Include a variable for school size
4. Written in English
5. Published post-1980
6. Contain empirical data and outcomes
7. From an OECD country

A school-within-school was defined as a separate and autonomous unit (as defined by the presence of its own head teacher) and run within a larger school

structure. The focus on secondary education was in accordance with the research brief. Middle schools and sixth forms were included in the search. To be included, the focus of the study did not have to be school size, but school size had to be included as a possible explanatory factor or comparison group in any qualitative or quantitative analyses. An arbitrary date of 1980 was chosen for feasibility within the timescale of the research. Empirical studies with outcomes were needed to answer the review question. While recognising that there are differences within OECD countries, it was considered that the education systems in countries outside the OECD would be too different from those in the UK, and inclusion of non-OECD countries in the synthesis would therefore be inappropriate. Further details of the inclusion and exclusion criteria are presented in Appendix 2.1.

2.2.2 Identification of potential studies: search strategy

Studies were identified through systematic searches of bibliographic databases (Appendix 2.2). A database system was set up to keep track of, and to code, studies found during the review. Titles and abstracts were imported and entered manually into the first of these databases.

To facilitate the completion of the review within the specified three-month timescale, little use was made of personal contacts to identify studies, no systematic handsearching of journals was undertaken and only limited searches of relevant websites and of citations from reference lists were possible.

2.2.3 Screening studies: applying inclusion and exclusion criteria

Inclusion and exclusion criteria were applied successively to (i) titles and abstracts, and (ii) full reports. Unless excluded at the first stage, the full text articles were ordered and those retrieved before the cut-off date of 27th October 2003 were screened again. If they did not meet the inclusion criteria, they were excluded. The remaining reports were entered into a second database. Three members of the EPPI-Centre team were involved in this screening.

2.2.4 Characterising included studies

Those that remained were characterised by assigning generic keywords (EPPI-Centre 2003a; Appendix 2.3) and review-specific keywords (Appendix 2.3) in order to produce a systematic 'map' of the research literature. Seven members of the EPPI-Centre team were involved in this keywording. Review-specific keywords were designed to consider some variables in more detail:

- whether the study focuses on school consolidation, schools-within-schools, or school size
- the outcomes measured, for example student achievement, teacher morale and school governance
- the construction of the school size variable as whole school size or using a proxy measure
- whether the size variable was measured categorically or continuously

The keywords were used to generate a map of the literature. All studies included in the systematic map have been added to the larger EPPI-Centre's Research

Evidence in Education Library (REEL) for others to access the keywords via the website.

2.2.5 Identifying and describing studies: quality-assurance process

In a moderation exercise, a random sample of 40 titles and abstracts was reviewed against the inclusion criteria. This was undertaken by three EPPI-Centre team members, working first independently and then comparing their decisions before reaching a consensus. Based on this consensus, the titles and abstracts for all remaining potentially relevant studies were then screened against the inclusion criteria by members of the review group, again working independently.

A moderation exercise was undertaken where the same two papers were keyworded independently by seven members of the EPPI-Centre who undertook keywording. The results were discussed by the group and any discrepancies clarified. Twenty papers were keyworded by two people from this group working independently, who then compared their decisions before reaching a consensus. The remaining papers in the map stage of the review were keyworded by one member of the group.

2.3 In-depth review

2.3.1 Moving from broad characterisation (mapping) to in-depth review

For the in-depth review, a second set of criteria was added to the first set of seven criteria (see section 2.2.1). These were developed in conjunction with the review group and review commissioners who wished to analyse the schools-within-schools literature separately and wished to narrow the focus of the synthesis to the studies likely to be of greatest relevance to answering the review question, taking account of the time constraints. The following types of studies were therefore excluded from further analysis:

8. The focus was schools-within-schools.
9. The number of schools in the sample could not be ascertained.
10. Data were collected before 1990 (except where they were collected over a time span that included 1990).
11. The analysis did not control for SES (socioeconomic status).
12. The study did not focus on one or more of the following outcomes: (i) student attainment and progress, attitudes, behaviour (ii) teacher morale and experience, and (iii) school organisation, management and costs.
13. The sample comprised only higher attaining or advanced students: for example, studies focusing only on sixth-form students were excluded.

2.3.2 Detailed description of studies in the in-depth review

For the in-depth review, detailed reading of the included studies was carried out and a range of information was extracted, using the standard EPPI-Centre guidelines and software (EPPI-Centre, 2003b). This included study aims and rationale; research questions and focus; research methods, including design,

sample and sampling strategy, data-collection, data-analysis, results and conclusions. Additional review-specific questions were identified through discussions with the broader review and advisory group and were applied alongside, and in the same manner as, the generic data-extraction questions. The topics addressed in the review-specific questions focused, for example, on the nature of the school size variable, number of grades in the participating schools, organizational structure of the schools, age at which outcomes were assessed, size of classes, student teacher ratios, possible confounding variables controlled for in the analysis, and use of outcome data from more than one year. Where relevant, authors were contacted for clarification.

Data from, and authors' narrative reports of, the results and conclusions for each study were extracted and considered by the review group members. Only those results that estimated the impact of school size, whilst controlling for an indicator of SES (e.g. percentage of students known to be eligible for free school meals) are reported. Statistically significant associations between the outcome (dependent) variable and a range of explanatory variables other than school size, were reported in some papers. These have not been reproduced in this report but the complexity of the model of analysis used was included in consideration of the weight of evidence each study provided for the review (see below).

The results as reported here represent a critical interpretation of those reported in the original studies. All relevant results are shown, but only those reaching a level of statistical significance of 5% (or a more stringent level), have been reported as being statistically significant in this review.

2.3.3 Assessing quality of studies and weight of evidence (WoE) for the review question

EPPI-Centre weight of evidence (WoE) judgments were applied to studies included in the in-depth review. Three components were identified to help in making explicit the apportioning of different weights to the findings and conclusions of different studies. Such weights of evidence are based on the following:

A. The overall quality of the study (internal methodological coherence). This component asks to what extent the study findings can be trusted in answering the study question. Studies can then be categorised as providing WoE on a range from high to low on the basis of answers to the EPPI-Centre generic questions on areas such as reporting, context, sample, design, reliability and validity of data-collection and analysis (including appropriate number and range of explanatory variables in the statistical models), ethics, sample size, risk of bias resulting from selection and maintenance of sample, and generalisability.

B. The appropriateness of the research design and analysis used for answering the review question. As the review question, in asking about the *impact* of school size, implies causality, no study using a non-experimental design can be ranked high in terms of research design. Studies using appropriate designs within that constraint and appropriate analyses for the review question (such as multiple regressions, taking account of multiple levels (class/school/district, etc.)) can be judged medium WoE B. Studies only using correlations or that do not use multivariate analysis would be low, as would studies in which results for secondary schools cannot be separated from other school types and studies with small sample sizes at their level of analysis.

C. The relevance of the study topic focus (from the sample, measures, scenario or other indicator of the focus of the study) to the review question. For some studies, the impact of school size was not central to their concerns, but it is unlikely that studies which do not provide any information for the review will have been included in the in-depth stage. Hence most of the studies will have either a high or a medium WoE C, although a classification of medium/low or low is possible (e.g. if results for secondary schools are not reported separately).

D. An overall weight taking into account A, B and C. Studies classified as low overall were still included in the synthesis as they met the inclusion criteria for the review, but less reliance was placed on their results.

To provide an additional framework for making judgments about the design and analysis of these studies, we considered what threats might exist to the validity of the studies, adapted from the work of Cook and Campbell (1979). The overall number of such 'threats' as well as the specific threats related to each study were compared with the original assessment on WoE A and also used as possible explanatory factors in consideration of any heterogeneity in the results of the studies. (Further details are provided in Appendix 2.4.)

The following algorithm for deriving the overall WoE D was applied:

- If WoE A is low, low/medium or medium/low then WoE D will be low, low/medium or medium/low respectively.

- In all other instances, an 'average' is taken. For example, a study receiving high for WoE A and medium for WoE B and C would receive medium/high for WoE D, whilst a study receiving medium for WoE A, low for WoE B and medium/low for WoE C would receive medium/low for WoE D.

- Given that no studies could be classified as high on WoE B, it was felt that no studies could therefore be classified as high overall (i.e. the highest categorisation was high-medium).

Preliminary judgments were discussed by the full review group and discrepancies resolved.

2.3.4 Synthesis of evidence

The data were synthesised to bring together the studies which answer the review questions and which meet the quality criteria relating to appropriateness and methodology. As very few studies provided data suitable for a statistical meta-analysis (see Chapter 4) and a meta-analysis on such a selective subset might be biased, studies were synthesised narratively, using as a framework the outcome (dependent) variable(s) for a study, and taking WoE judgments into account.

2.3.5 In-depth review: quality-assurance process

Five members of the EPPI-Centre undertook the in-depth data-extraction. As a moderation exercise, each person independently completed a data-extraction for comparison with at least one other group member. Confirmatory data-extraction by a second group member was undertaken on all the remaining studies.

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

Section 3.1 describes the derivation of the studies in the map; section 3.2 characterises the studies in the map in terms of the generic and the review-specific keywords and section 3.3 reports the results of the quality assurance process.

3.1 Studies included from searching and screening

A total of 3,874 citations were identified through systematic searches of six databases. The number of citations identified in each database is shown in Table 3.1. The 3,874 citations include 121 duplicate reports which were excluded when titles and abstracts were screened.

Table 3.1: Citations identified from electronic databases

Database	Number of citations
ASSIA	20
Australian Educational Index	332
British Educational Index	210
ERIC	2,563
Social Science Citation Index	126
Psycinfo	623
Total	***3,874***

A further 15 papers were identified through one-stage screening processes, including specialist websites, citation lists and personal contacts. After exclusions, 481 reports remained as potential 'includes', but 252 were not obtained in time for the cut-off date of 27th October 2003. A total of 229 full text documents were obtained and screened for inclusion. Following full text screening, a total of 119 studies were considered to meet the inclusion criteria. The filtering of the papers through the review process is shown in Figure 3.1.

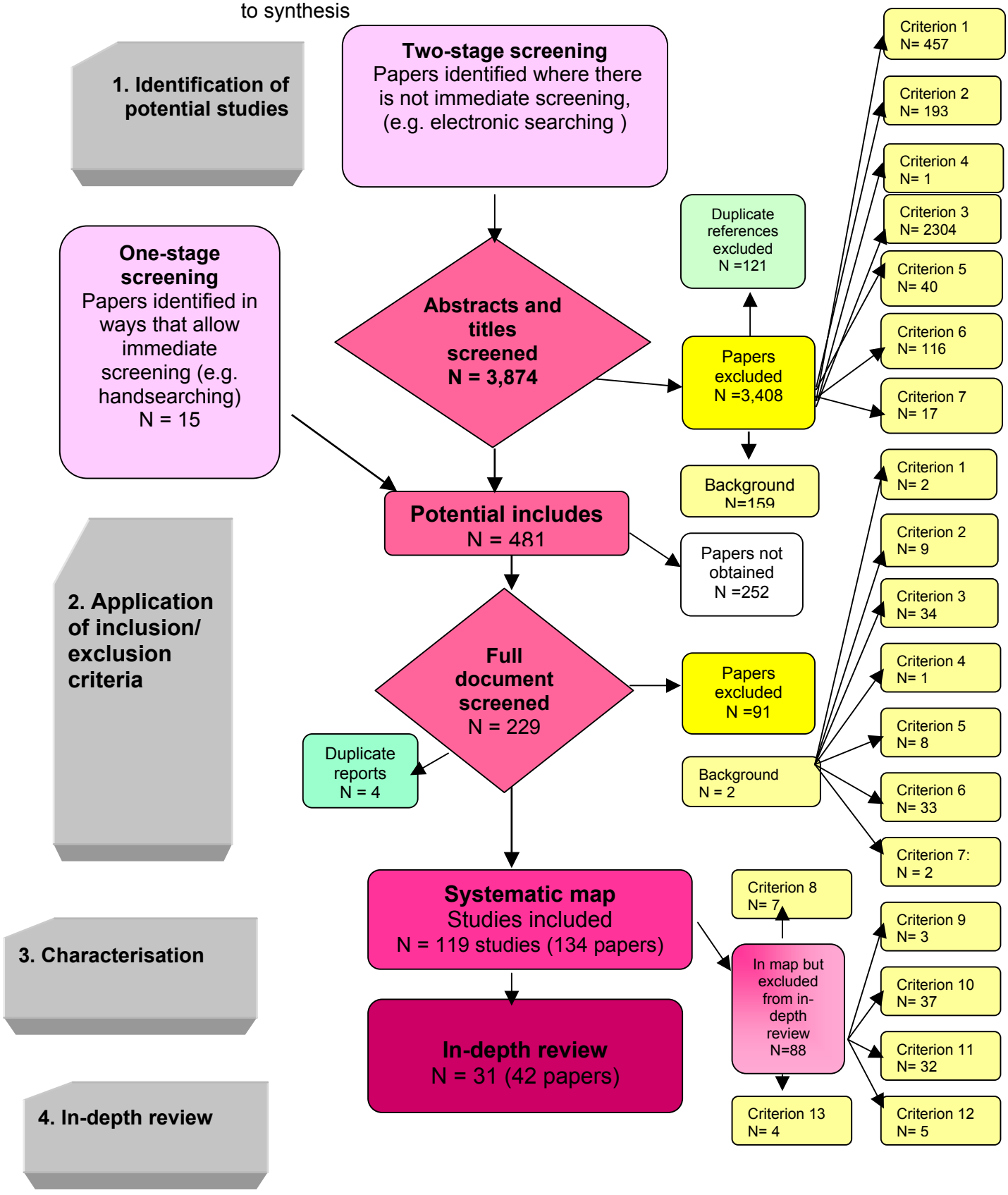
The majority of included studies were identified through searching of electronic databases (Table 3.2).

Table 3.2: Source of identification of 119 studies in the map* (N=119)

Sources	Number
Contact	2
Handsearching, including websites and citation lists	6
Electronic database	111

*Codes mutually exclusive

Figure 3.1: School size review: Filtering of papers from searching to map to synthesis



3.2 Characteristics of studies from searching and screening

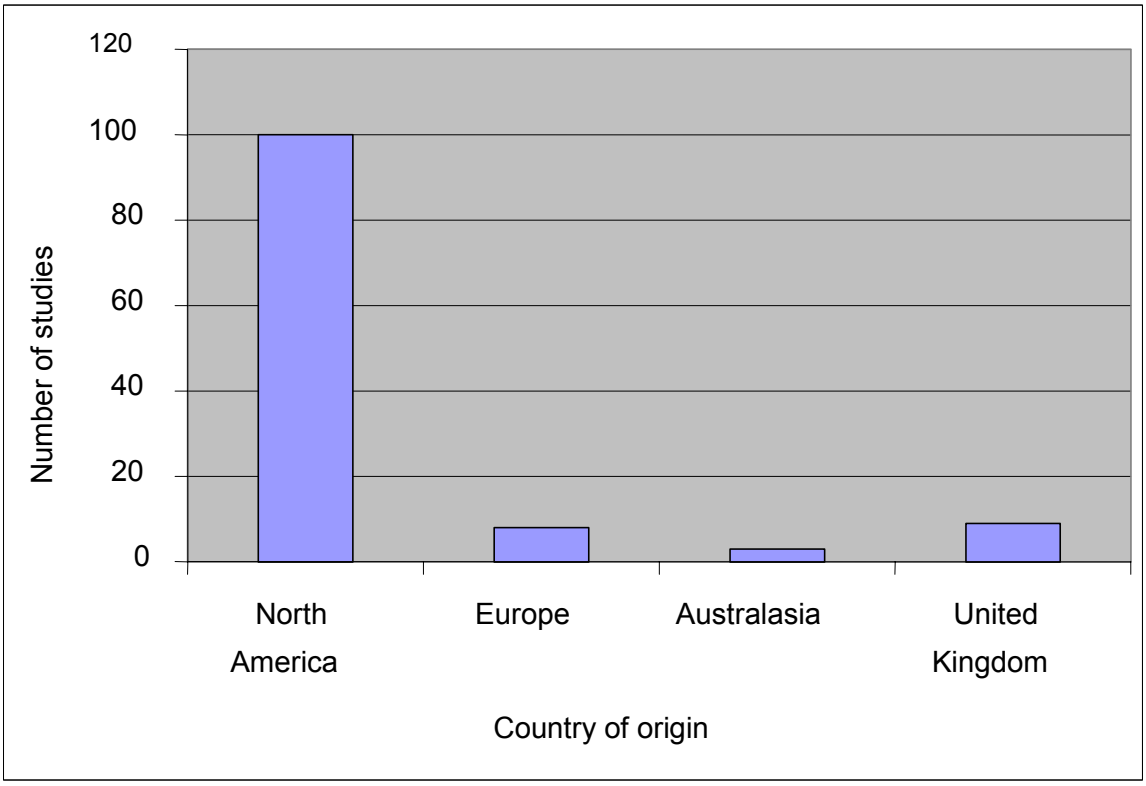
Following application of the inclusion criteria, the 119 studies included, were characterised to create a systematic map. The map in section 3.2.1 is based on all studies; section 3.2.2 is based on all studies analysing schools-within-schools; section 3.2.2 is based on studies from the UK only and section 3.2.4 is based on all studies after removing those only concerned with schools-within-schools.

3.2.1 All studies

Generic EPPI-Centre keywords

The majority of the studies (N=100) were carried out in North America. Of the other studies, eight were from European countries (other than the UK), three were from Australasia, and nine were from the UK (Figure 3.2). Of the nine studies from the UK, one was from Northern Ireland, one was based on data from England and Wales, and the others were based on data from England only. One study included in the systematic map analysed data from more than one country comparing national datasets obtained from The Netherlands, Sweden and the USA (Luyten, 1994).

Figure 3.2: Country of origin* (N=119 studies in the map)



*Single code for each study except one study using data from both America and Europe

All studies were categorised as 'organisation and management' due to their focus on school size (Table 3.3). Just over half the studies were also coded as 'teaching and learning' and 'curriculum', reflecting the dominant focus in the identified studies on the 'effects' of school size on student achievement. Smaller numbers of

studies were identified that considered other aspects such as teaching and school organisation. Those studies coded 'other' topic focus generally focused on an aspect of student behaviour such as participation, engagement, truancy, dropout or absence.

Table 3.3: Topic focus* (N=119 studies in the map)

Topic focus	Number
Classroom management	3
Curriculum	71
Equal opportunities	5
Organisation and management	119
Policy	8
Teacher careers	9
Teaching and learning	66
Other topic focus	41

*Codes not mutually exclusive

Studies in Table 3.3 were coded as being about curriculum in two instances: first, where a study measured achievement within a particular subject or across subjects; and second, where a study considered the range or breadth of curricular offerings.

Table 3.4: Curriculum focus* (N=71 studies with curriculum as a topic focus)

Curriculum focus	Number
Art	1
Business studies	3
Citizenship	3
Cross-curricular	20
Design and Technology	1
Environment	1
General	1
History	4
Information and Communication Technology (ICT)	2
Literacy - first language	28
Literacy further languages	7
Mathematics	36
PSE	1
Phys. Ed.	4
Science	22
Vocational	5
Other curriculum	18

*Codes not mutually exclusive

Studies in Table 3.4 coded as 'cross-curricular' considered the 'effects' of school size on exam marks as a whole (e.g. total GCSE points scores, overall HSC exam scores), whilst those coded as 'other curriculum' considered the 'effects' of school size on either the range of curricular offerings (N=8 studies), extra-curricular offerings (N=3 studies) or other subjects not coded in previous categories such as social science and agriculture (N=7 studies). Overall, those studies considering achievement in specific curricular areas were most likely to measure aspects of mathematics, science or literacy. This may reflect the fact that these subjects are generally completed by the majority of students and more likely to be tested at regular intervals throughout compulsory education.

The majority of studies were coded as focusing on learners (N=103 studies); far fewer studies focus specifically on other categories, such as teaching staff (N=23 studies) or senior management (N=7 studies), reflecting the focus in the research literature on student achievement and behaviour. Studies coded with the population focus 'other' in Table 3.5 include those studies that focus either on economic outcomes (including direct public expenditure on schools per student/per graduate, and inefficiency) or curricular diversity, where there was no explicit focus on either students or educational staff.

Table 3.5: Population focus* (N=119 studies in the map)

Population focus	Number
Learners	103
Senior management	7
Teaching staff	23
Non-teaching staff	2
Government	4
Local education authority officers	2
Parents	3
Governors	1
Other population focus	5

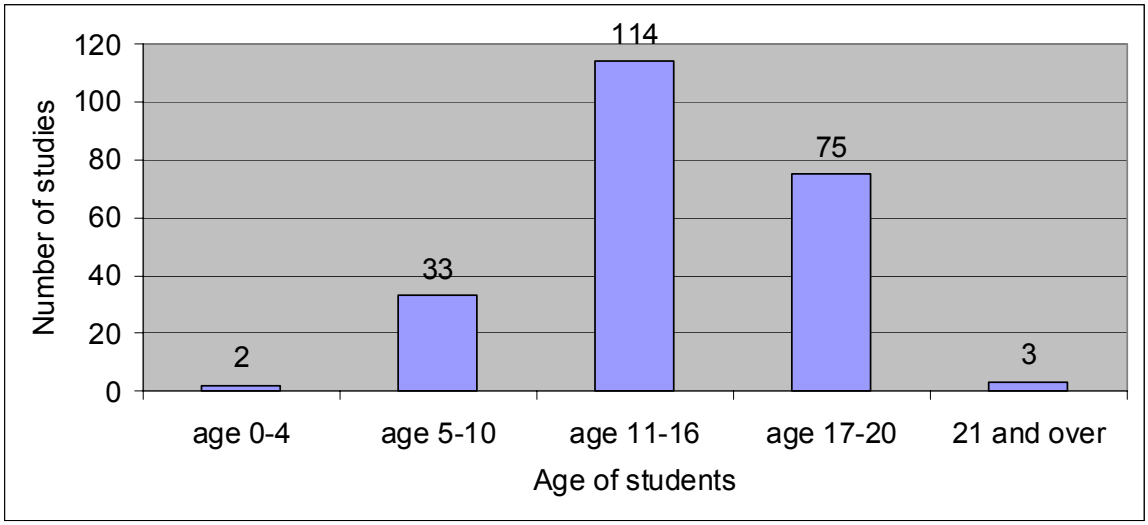
*Codes not mutually exclusive

Studies were only included in this review if they report data on secondary school size; therefore the predominant focus of most studies is on the 11-16 age range (N=114 studies), with a smaller number also including the 17-20 age range (N=75 studies). A number of studies investigated the issue of school size across a range of school types/age groups, reflected in the patterns seen in Figure 3.3 and Table 3.6. In some studies, the analysis was based on a composite across age groups (e.g. school level average across all grades in a school), whilst in others individual grades were taken as the sample for data-collection (e.g. all students in grade 10). In a small number of studies, average school scores were collapsed across school types to provide a single measure of 'effect'. It is important to remember when interpreting this map that it does not systematically investigate the research literature which analyses data about school size for pre-school and primary school students, or for students in higher education.

Study locations also reflect the focus on secondary age students, with all but one of the studies using data from secondary schools (N=118 studies). None of the studies specifically focuses on independent schools, although some studies from the USA include private or Catholic schools as well as public schools. No studies were identified that focus on schools for children with special educational needs. Those

studies focusing on higher education establishments (N=4 studies) or post-compulsory education establishments (N=1 study) typically measured student behaviour in university and then related this back to the size of high school the student attended. Only one of the studies from the UK included independent schools in the analyses (Gill *et al.,* 2002), and only two included analyses of primary school data (Spielhofer *et al.,* 2002; Thomas and Bullock, 1992).

Figure 3.3: Age of students* (N=119 studies in the map)



*Codes not mutually exclusive

Table 3.6: Study location* (N=119 studies in the map)

Study location	Number
Higher education institution	4
Independent school	12
Nursery school	1
Post-compulsory education institution	1
Primary school	32
Secondary school	118
Other educational setting	1

*Codes not mutually exclusive

The majority of studies were coded as exploration of relationships (N=108 studies; Table 3.7). This reflects the predominance of cross-sectional analyses of secondary data sources, such as national or state statistics of education. Of the studies coded as evaluations, eight studies evaluated schools-within-school programmes.

Table 3.7: Study type* (N=119 studies in the map)

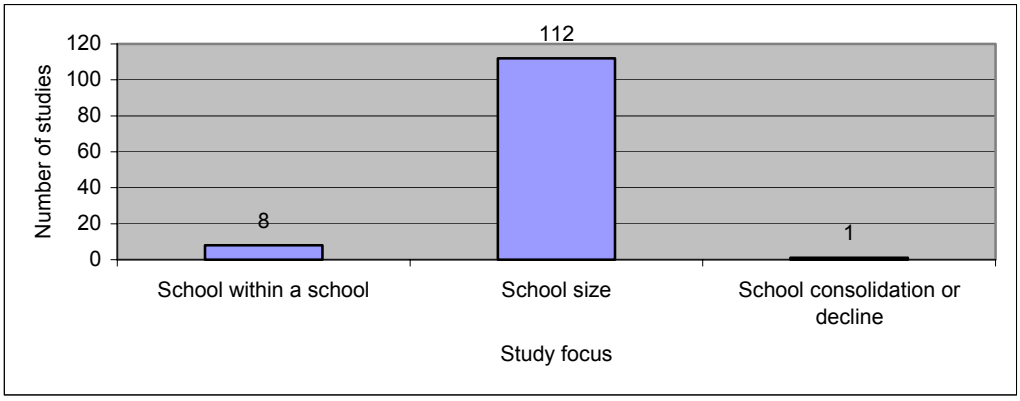
Study type	Number
Exploration of relationships	108
Evaluation: naturally occurring	24
Evaluation: researcher-manipulated	1

*Codes not mutually exclusive

Review-specific keywords

The majority of studies identified considered the 'effects' of school size (N=112). Relatively few schools-within-schools citations were identified (N=8 studies, Figure 3.4). This may reflect that schools-within-schools are a comparatively recent phenomenon and therefore the literature is at a fairly emergent stage. Searches identified 70 studies investigating schools-within-schools, of which only 17 were considered to meet the inclusion criteria, and only eight could be obtained within the time limits. The majority of the studies in this category were excluded because they were either descriptions of schools-within-schools without any measured outcomes, reviews of the literature, or opinion pieces.

Figure 3.4: Focus of the study* (N=119 studies in the map)



*One study was coded as both school size and school consolidation, whilst another was coded as both school size and schools-within-schools

Very few studies focus on subgroups of students (Table 3.8); all studies in the map were coded as studying mainstream students, with a small number of studies also focusing on students from ethnic minority backgrounds (N=8 studies). A greater number of papers include ethnicity as an explanatory variable in regression models, but, for the purposes of this map, studies were only coded as being about students of ethnic minority if they have a predominant focus on ethnicity. No papers were coded as focusing on either students who were gifted and talented or those with other special educational needs (SEN). The studies investigating schools-within-schools tend to focus on developing smaller schools for those considered 'at risk' of dropping out or failing examinations. For the purposes of the systematic map, these were subsumed under the mainstream category, rather than SEN.

Table 3.8: Student focus* (N=119 studies in the map)

Student focus	Number
Mainstream	119
Ethnic minority	8

*Codes not mutually exclusive

In the studies identified, school size was presented both as a categorical variable (e.g. schools of different sizes were divided into groups of similar sizes; N=55 studies) and as a continuous variable (e.g. school size was presented on a continuous scale where it was assumed that size could take any value; N=70

studies; see Table 3.9). In a small number of studies, multiple analyses were completed considering school size both as a continuous and as a categorical variable (N=6). Studies focusing on schools-within-schools were coded as categorical as they compared differences either between schools with and without the schools-within-school structure, or within schools following implementation of the schools-within-school structure.

Table 3.9: Presentation of the school size variable* (N=119 studies in the map)

Presentation	Number
Categorical	55
Continuous	70

*Codes not mutually exclusive

Studies either measured the school size variable based on the whole school enrolment (whole school size; N=90 studies), or some form of proxy measure such as average grade size (referred to as span size), final year size, or end of year membership (N=33 studies; Table 3.10). The use of proxies attempts to address the situation that a school with 800 students in three grades could be considered to be a larger school than one with 800 students across seven grades; a study measuring the whole school enrolment fails to reflect this distinction. No studies were found that distinguished between enrolment figures and attendance figures. Schools-within-school studies could be coded as either a whole school or proxy measure, depending on whether they evaluated outcomes for all children in the schools-within-school, or evaluated the outcomes of a sample of students within the school.

Table 3.10: Nature of the school size variable* (N=119 studies in the map)

Nature of the school size variable	Number
Whole school size studied	90
Proxy measure	33

*Codes not mutually exclusive

A total of 93 studies analysed student level outcomes (Table 3.11). The most common outcome is student attainment (N=54 studies), of which a small number of studies include a control for prior attainment (N=12 studies): that is, typically through using a gain score (i.e. the change in performance of an individual student at a particular subject between two points in time). Other relatively common student outcomes include student attitudes (N=23 studies) and student behaviour (N=36 studies). Student attitudes include measures of student engagement, self-esteem and loneliness, whilst those studying behaviour include those that are both negative (e.g. bullying, student disorder) and positive (e.g. participation in school activities). Studies considering absence and truancy were for the purposes of keywording all subsumed under the category ‘attendance’ (N=21). Very few studies investigated post-school destination (N=8 studies). Those that did considered both whether students entered higher education institutions and what type of higher education institutions they entered. The studies coded ‘other’ include a diverse range of outcomes, including course-taking patterns, number of

students repeating a year, opportunities available for students and physical wellbeing.

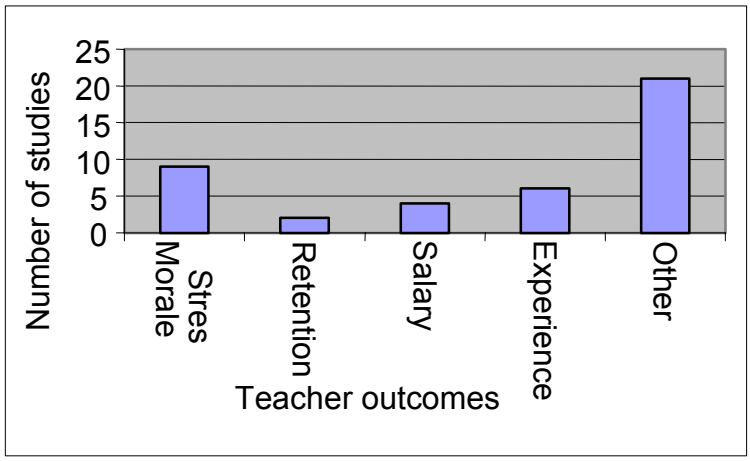
Table 3.11: Student level outcomes* (N=93 studies with student level outcomes)

Outcomes	Number
Performance measured with prior attainment	12
Performance measured without prior attainment	47
Performance measured with or without prior attainment	54
Post-school destination	8
Student attitudes	23
Student behaviour	36
Attendance	21
Other	6

*Codes not mutually exclusive

Of the 119 studies included in the systematic map, only 35 studies measured teacher outcomes (Figure 3.5). Nine of these studies measured teacher outcomes focusing on teachers' morale and stress; two studies focus on teacher retention (e.g. whether the retention of teachers differs across large and small schools); four studies focus on whether salaries differ across large and small schools; and six studies focus on 'experience' (e.g. the differences amongst teachers in different sized schools in terms of number of years teaching, or qualifications).

Figure 3.5: Teacher outcomes* (N=35 studies with teacher outcomes)



*Codes not mutually exclusive

The majority of studies that included teacher outcomes were coded as 'other' (N=21 studies). In these studies, the range of outcomes studied is diverse but a predominant focus is on teachers' or students' perceptions of teaching efficacy and quality (N=4 studies), school culture and climate (N=5 studies) and teachers' perceptions of their role and duties (N=7 studies). Other study outcomes included computer use, social skills, number and distribution of teaching staff, and coursework grading behaviour.

Just under half of the studies in the systematic map include a school level outcome (N=53 studies; Table 3.12). Of these, the most common outcomes are range of curricular provision (N=19 studies) and school accountability and governance (N=15 studies). Studies coded as school accountability and governance reported outcomes that were usually principals', teachers' and/or students' perceptions of leadership and/or decision-making 'cultures' within schools. Studies reporting the outcomes 'student–teacher relationship' (N=6 studies) and 'communication within the school' (N=5 studies) tend to be a subset of the studies coded as reporting 'school accountability and governance'. The category of education economics (N=8) include studies that considered the costs of education and the efficiency of schools of different sizes. Studies coded 'other' (N=8) include a wide range of outcomes, including resource availability, school enrolment characteristics and measures of school climate.

Table 3.12: School level outcomes* (N=53 studies with school level outcomes)

Outcomes	Number
Class size	5
Grouping arrangements	8
School accountability and governance	15
Student-teacher relationships	6
Communication within the school	5
Parental involvement	3
Relationships between school and wider community	3
Range of curricular provision	19
Range of extra-curricular provision	5
Economics of education	8
Other	8

*Codes not mutually exclusive

To explore the nature of the outcomes further, a series of cross tabulations was constructed, examining the nature of the variable (categorical versus continuous) and the nature of the outcomes (student, teacher or school level) (Tables 3.13 and 3.14, and Figure 3.6).

Table 3.13: Student level outcomes by presentation of the school size variable* (N=93 studies with student level outcomes, 44 studies using categorical variables, 60 studies using continuous variables)

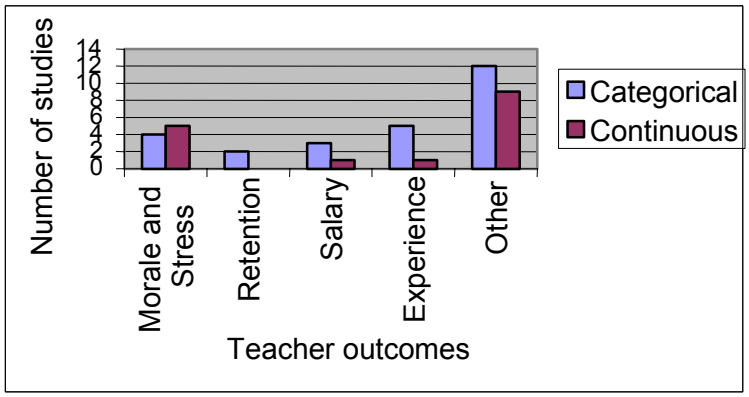
Outcome	Categorical N=44*	Continuous N=60*
Performance measured with prior attainment	5	11
Performance measured without prior attainment	22	26
Performance overall measured with or without prior attainment	25	33
Post-school destination	5	3
Student attitudes	17	8
Student behaviour	23	16
Attendance	12	10

Long-term economic outcomes	0	0
Other	1	5

*Codes not mutually exclusive

These analyses reveal that studies investigating student level outcomes were more likely to measure school size as a continuous variable; studies investigating teacher or school level outcomes were more likely to measure the variable categorically. This may reflect the fact that the studies analysing teacher and school level outcomes often used less sophisticated analyses and were more likely to be chronologically older than those analysing student outcomes.

Figure 3.6: Teacher outcomes by presentation of the school size variable* (N=35 studies with teacher outcomes, 21 studies using categorical variables, 14 studies using continuous variables)



*Codes not mutually exclusive

Table 3.14: School level outcomes by presentation of the school size variable* (N=53 studies with school level outcomes, 31 studies using categorical variables, 25 studies using continuous variables)

Outcome	**Categorical N= 31***	**Continuous**
Class size	4	1
Grouping arrangements	4	4
School accountability and governance	10	5
Student-teacher relationships	5	1
Communication within the school	3	2
Parental involvement	3	0
Relationships between school and wider community	2	1
Range of curricular provision	13	6
Range of extra-curricular provision	5	0
Economies of scale	3	6
Other	4	4

*Codes not mutually exclusive

3.2.2 Studies carried out in the UK

Nine studies were carried out in the UK. Many of these studies were coded as both explorations of relationships and naturally occurring evaluations (Table 3.15), reflecting the fact that many used national datasets and considered changes over time, or implicitly evaluated policy changes, such as the introduction of market forces in education or changes in education funding.

Table 3.15 Study type* (N=9 UK studies in the map)

Study type	Number
Exploration of relationships	8
Evaluation: Naturally occurring	7

*Codes not mutually exclusive

The dominant focus across all the UK studies is on student attainment and progress (Table 3.16). Very few studies focus on other outcomes (Tables 3.17 and 3.18). The study that includes the broadest range of outcomes is a qualitative evaluative study from 1987-89, comparing six small secondary schools and including county averages for some measures (Tomlinson and Mortimore, 1990).

Table 3.16: Student outcomes* (N=7 UK studies with student outcomes)

Outcomes	Number
Performance measured with prior attainment	3
Performance measured without prior attainment	6
Performance measured with or without prior attainment	7
Post-school destination	1
Student behaviour	1
Attendance	1
Other	1

*Codes not mutually exclusive, only seven of the nine studies focus on student outcomes

Table 3.17: Teacher outcomes* (N=2 UK studies with teacher outcomes)

Outcomes	Number
Salary	1
Experience	1

*Codes not mutually exclusive; only two studies of the nine UK studies in the map focus on teacher outcomes.

Table 3.18: School outcomes* (N=4 UK studies with school outcomes)

Outcomes	Number
Student-teacher relationships	1
Range of curricular provision	2
Economies of scale	2

*Codes not mutually exclusive, only four of the nine UK studies in the map focused on school outcomes.

3.2.3 Schools-within-schools studies

A second set of sub-analyses was completed to investigate further the literature evaluating schools-within-schools. All the studies identified were from the USA and evaluated specific schools-within-schools programmes. In all but one of the cases, the studies were unpublished. Seven of the studies were coded as naturally-occurring evaluations, reflecting the fact that the schools-within-schools structure was often used to engage specific 'at risk' students (Table 3.19). A single study randomly assigned students to enrol in the schools-within-schools structure or in a non schools-within-schools structure, and then completed separate sub-analyses on 'at risk' groups.

Table 3.19: Study type* (N=8 studies of schools-within-schools in the map)

Study type	Number
Evaluation: naturally occurring	7
Evaluation: researcher-manipulated	1

*Codes mutually exclusive

Studies evaluating schools-within-schools investigate a broad range of outcomes reflecting the fact that the studies were often more qualitative. However, as was seen in the school size studies, student attainment is the most common outcome variable (Table 3.20). Within this specific subset of studies, there was also a focus on student attitudes and attendance, reflecting the fact that students in these studies were often identified as being most likely to drop out and become disengaged with the educational process.

Table 3.20: Student outcomes* (N=8 studies of schools-within-schools with student outcomes)

Outcome	Number
Performance measured with prior attainment	1
Performance measured without prior attainment	7
Performance measured with and without prior attainment	7
Post-school destination	3
Student attitudes	3
Student behaviour	5
Attendance	7

*Codes not mutually exclusive

The majority of studies also investigated an aspect of teacher (Table 3.21; N=5) or school organisation (Table 3.22; N=5), although studies typically investigated fewer outcomes at these levels than at the student level. There are no dominant trends in the study foci, although relationships between teachers and students, and with parents and the wider community were more frequently investigated than they were in the other school size studies.

Table 3.21: Teacher outcomes* (N=5 studies of schools-within-schools with teacher outcomes)

Outcomes	Number
Morale and stress	3
Other	3

*Codes not mutually exclusive

Table 3.22: School outcomes* (N=5 studies of schools-within-schools with school outcomes)

Outcomes	Number
Class size	1
Grouping arrangements	2
School accountability and governance	3
Student-teacher relationships	2
Communication within the school	1
Parental involvement	2
Relationships between school and wider community	2
Range of curricular provision	2

*Codes not mutually exclusive

3.2.4 Summary from the systematic map

The systematic map shows that the literature considering the 'effects' of secondary school size is predominantly from the USA and focuses mainly on the differences in outcomes of schools of different sizes. This is as opposed to the impact of school-within-school initiatives (creating smaller schools within the larger school site) or the impact of consolidation or growth within a single school. The studies in the review were generally coded as explorations of relationship, reflecting the tendency for the studies to use cross-sectional data from nationally published datasets. The focus of the outcomes was predominantly concerned with achievement and behaviour or attitudes of mainstream students. Very few studies included in the review focused on longer-term student outcomes, such as post-school destination, or considered the possible differential impact on students from different socio-economic groups, ethnic minorities, or on students with special educational needs. In addition, only a minority of studies considered the impact of school size on teachers or on aspects of school organisation.

3.3 Quality-assurance

A moderation exercise was undertaken for which the same two papers were keyworded independently by seven members of the EPPI centre who undertook keywording. The results were discussed by the group and any discrepancies clarified. Twenty papers were keyworded by two people from this group working independently, who then compared their decisions before reaching a consensus. The remaining papers in the map stage of the review were keyworded by one member of the group.

4. IN-DEPTH REVIEW RESULTS

This chapter outlines the studies selected for in-depth review (section 4.1) and compares the studies in the systematic map with those in the in-depth review (section 4.2). It goes on to describe the characteristics of the studies in the in-depth review in terms of EPPI-Centre keywords (section 4.2.1), review specific keywords (section 4.2.2) and review-specific questions (section 4.3). It outlines the quality judgments assigned to each of the studies (weights of evidence in section 4.4) before presenting the synthesis of evidence (section 4.5) and quality-assurance results (section 4.6). The final section provides an overview of the user involvement in the review process (section 4.7).

4.1 Selecting studies for in-depth review

In determining which studies would best answer the review question, it was decided that studies required a control in their analyses for the 'effects' of SES. Additional limits were also set to ensure relevance to the review commissioners (e.g. data collection following 1990 and the focus on outcomes). The process of determining these second-stage criteria has already been outlined in Chapter 2. A total of 31 studies were identified for the in-depth review.

4.2 Comparing the studies selected for in-depth review with the total studies in the systematic map

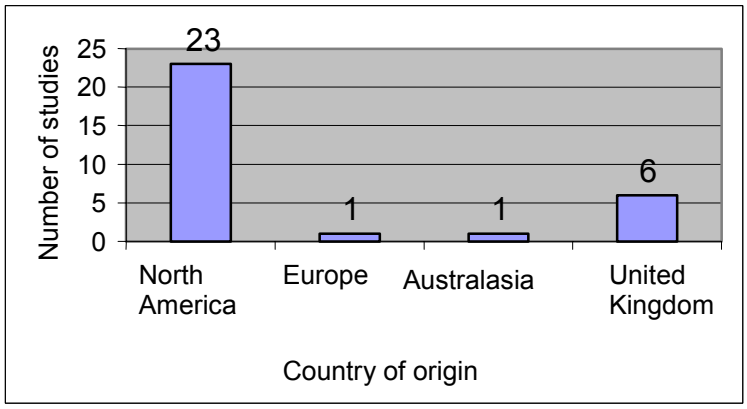
The sections below describe the studies in the in-depth review in relation to the EPPI-Centre generic (section 4.2.1) and review specific keywords (section 4.2.2). The tables show data for the studies in the in-depth review only, whilst Appendix 4.1 provides tables of data comparing the studies in the systematic map with those in the in-depth review.

4.2.1. Keywords of studies in the in-depth review

The majority of the studies were identified through electronic databases (N=25), with an additional four studies identified through handsearches and two studies through personal contacts.

As in the systematic map, the majority of studies in the in-depth review were carried out in North America, with 21 from the USA and two from Canada. Fewer studies came from Europe, Australasia and the United Kingdom. Proportionally, the number of studies from the UK is greater in the in-depth review than in the systematic map (6/31 (19%) versus 9/119 (8%)). All studies coded as being from the UK used data from England only.

Figure 4.1: Country of origin of studies * (N=31 studies in in-depth review)



*Categories mutually exclusive

The topic foci of the studies in the in-depth review (Table 4.1) show similar patterns to those in the systematic map. All studies in the in-depth review were coded as 'organisation and management' due to their focus on school size. Other common categories are curriculum (N=18 studies), and teaching and learning (N=21 studies), reflecting the trend in the literature towards measuring student attainment and progress. A large number of studies coded 'other' remained in the in-depth review, reflecting interest in the 'effect' of school size on students' attitudes and behaviour.

Table 4.1: Topic focus* (N=31 studies in the in-depth review)

Focus	Number
Curriculum	18
Equal opportunities	2
Organisation and management	31
Policy	2
Teacher careers	1
Teaching and learning	21
Other topic focus	15

*Categories not mutually exclusive

The curricular foci of the studies in the in-depth review reveal that the studies were most likely to measure literacy (N=13 studies), mathematics (N=13 studies) and science (N=8 studies) (Table 4.2). This is a similar pattern to that of the studies in the systematic map. Overall there is a higher proportion of studies in the in-depth review coded as cross-curricular (7/31 (23%) versus 20/119 (17%)), reflecting the fact that many studies in the in-depth review focus on exam marks as a whole, instead of individual subject scores. A similar proportion of studies were coded as 'other curriculum' (4/31 (13%) versus 18/119 (15%)). All studies coded 'other' focus on a single subject that had not been previously coded, rather than the depth and range of curricular or extra-curricular courses.

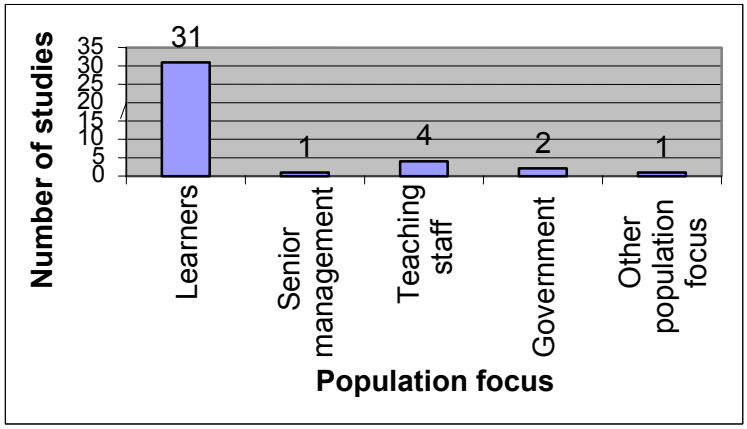
Table 4.2: Curriculum focus* (N=18 studies with curriculum focus in the in-depth review)

Curriculum focus*	Number
Citizenship	1
Cross-curricular	7
Design and Technology	1
History	3
Literacy - first language	13
Literacy further languages	2
Mathematics	13
PSE	1
Science	8
Other curriculum	4

*Categories not mutually exclusive

The studies in the in-depth review all focus on learners (Figure 4.2). When compared with the studies in the systematic map, proportionally fewer studies focus on other populations, such as teaching staff or senior management.

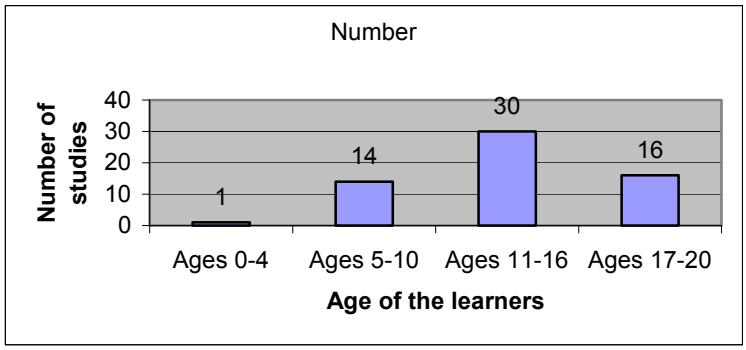
Figure 4.2: Population focus of the studies* (N=31 studies in the in-depth review)



*Categories not mutually exclusive

In all but one of the studies in the in-depth review, the age of the learners focuses on the core age group 11-16, reflecting the in-depth review criterion excluding studies that only focus on higher attaining or advanced students (Figure 4.3). Comparing the studies in the in-depth review with those in the systematic map shows that studies in the in-depth review are more likely also to include additional age groups in their analyses. This is particularly seen in the 5-10 category (14/31 (45%) versus 33/119 (28%)).

Figure 4.3: Age of the learners* (N=31 studies in the in-depth review)



*Categories not mutually exclusive

Due to the nature of the review question, all studies in the in-depth review were based in secondary schools. Within the broad category of 'secondary schools', one study used data from senior secondary schools, and four studies used data from middle schools only. It was not clear in a large number of studies (N=14) whether independent schools were excluded from the sample. Of the studies that explicitly stated whether they included independent schools, five studies included private schools and twelve studies did not (Table 4.3).

Table 4.3: Educational setting of the studies in the in-depth review* (N=31 studies in the in-depth review)

Setting	Number
Independent school	5
Primary school	14
Secondary school	31

*Categories not mutually exclusive

All studies in the in-depth review were coded as exploration of relationships. Five of the 31 included studies were also coded as evaluations of naturally-occurring interventions.

4.2.2 Characteristics of the studies in the in-depth review: review-specific keywords

As in the systematic map, all the studies in the in-depth review focus on mainstream education populations. Four studies also had a focus on ethnic minority groups. The majority of studies in the in-depth review present school size as a continuous variable (N=29/31 studies). However, two studies present school size as a categorical variable, whilst four studies present school size both categorically and continuously.

Twenty-two studies in the in-depth review measured school size as whole school enrolment, while twelve studies used a proxy measure. Of the studies that used a proxy measure, it was most likely to be span size (e.g. total enrolment divided by the number of grades) (N=8 studies) so as to control for variations in grade configuration across schools. Twenty-five of the studies stated their school size figures were based on enrolment and the other six studies provided no information.

Twenty-eight studies in the in-depth review considered student outcomes. Comparing the student outcomes measured in the studies in the in-depth review

with those in the systematic map shows that a greater proportion of the studies in the in-depth review considered the 'effects' of school size on achievement (19/28 (68%) versus 54/93 (58%)). Four studies in the in-depth review considered student attitudes, seven considered student behaviour and five studies considered attendance (Table 4.4). No studies in the in-depth review measured the 'effects' of school size on post-school destinations, compared with 8/93 (9%) in the systematic map. There are a number of possible explanations for this phenomenon. It may reflect the inclusion criteria for the in-depth review since studies measuring these outcomes tended to be older and to use less sophisticated methods of analysis. However, it may also reflect a movement in research activity towards a focus on student attainment at the expense of other potentially important student outcomes.

Table 4.4: Student outcomes* (N=28 studies measuring student outcomes in the in-depth review)

Outcomes	Number
Performance measured with prior attainment	5
Performance measured without prior attainment	17
Performance measured total	19
Student attitudes	4
Student behaviour	7
Attendance	5
Other	1

*Categories not mutually exclusive

Three studies in the in-depth review measured teacher outcomes. Of these, one (Fetler, 1997) reports student outcomes using regression analysis whilst controlling for socio-economic status (SES), whilst teacher outcomes are presented in an initial correlation matrix which looked at strength of association between school size and teacher variables but did not control for SES (one of our criteria for inclusion in the synthesis). Therefore, although the student outcomes for this study were included in the synthesis, the teacher outcomes were excluded. Proportionally, there are fewer studies reporting teacher outcomes in the in-depth review than there are in the systematic map (2/31 (6%) versus 35/119 (29%)).

Nine studies in the in-depth review include school level outcomes (Table 4.5). The majority of these investigated economic outcomes, such as costs or efficiency (N=5 studies). Five studies include other school level outcomes (one study has both an economic and an additional school level outcome), of which one did not include controls for SES in the analysis and so was not used in the synthesis. Comparing these studies with those in the systematic map, proportionally more studies in the in-depth review considered issues in education economics: 5/9 (56%) versus 8/53 (15%) studies considering school level outcomes in the map.

Table 4.5: School level outcomes* (N=9 studies with school level outcomes in the in-depth review)

Outcomes	Number
Class size	1
Grouping arrangements	1
School accountability and governance	2
Student-teacher relationships	1
Communication within the school	1
Relationships between school and wider community	1
Range of curricular provision	1
Education economics	5

*Categories not mutually exclusive

4.3 Further details of studies included in the in-depth review

As part of the in-depth data-extraction, additional information was extracted from the studies about the nature of the schools and grade level when student outcomes were measured.

For 14 of the 31 studies in the in-depth review, it was not possible to identify the grade configurations (e.g. number of years taught in each school) of the schools in the sample (Table 4.6). This was most common for the studies from the USA where all schools with a particular grade was sampled covering a wide range of grade configurations. The six studies from England all include grade spans of 5 years (years 7-11) and 7 years (years 7-13). A large proportion of the other studies also include middle schools with grade spans covering either two years (grades 7-8 / years 8-9), or three years (grades 6-8/years 7-9). In the USA high school studies, the majority of grade spans are four years (grades 9-12/years 10-13), although one study has seven years (grades 7-12/years 8-13). A single study includes a group of sub-analyses on single unit schools covering grade spans of kindergarten to grade 12 (K-12).

Table 4.6: Number of years/grades in participating schools* (N=31 studies in the in-depth review)

Number of years/grades	Number
Two	2
Three	8
Four	5
Five	6
Six	1
Seven	6
More than seven (e.g. K-12)	1
Not given/Not possible to deduce	14

*Categories not mutually exclusive

All studies in the in-depth review controlled for a measure of SES in their analyses (Table 4.7). The nature of the measure varies, but is most frequently the proportion of children eligible for free or reduced-price school meals. In some of the studies from the USA, the measure is based on a category called 'aid for dependent children' (AFDC) figures. Eight of the studies also controlled for prior attainment in the analyses; of these, five primarily examined the 'effects' of school size on student attainment and progress, whilst the other three studies include prior attainment as an explanatory variable when considering other outcomes (e.g. student behaviour, attitudes).

Table 4.7: Controlling factors in the analysis* (N=31 studies in the in-depth review)

Factor	Number
Socio-economic status (SES)	23
Both prior attainment and SES	8

*Categories mutually exclusive

The studies in the in-depth review measured the 'effects' of school size within a range of different school establishments (Table 4.8). Ten studies only sampled secondary or high schools, four studies only sampled middle schools and a further four studies sampled both middle and high schools. In a number of cases (N=9 studies), data came from a range of institutions, reflecting the trend towards analysing data from any school with a particular grade (e.g. the universe of schools with a grade 8).

Table 4.8: Nature of the schools* (N=31 studies in the in-depth review)

Nature of school	Number
High/secondary school only (use for all UK secondary where age not specified)	6
Middle schools only	4
Mixture of high/secondary and middle schools	4
USA: Mixture of high/secondary, middle & K-12 schools	9
Sample consists of mixture of school types but results not reported separately	2
Not reported	2
UK: 11-16 schools only	4
UK: 11-18 schools only	4

*Categories not mutually exclusive

Eight of the 31 studies in the in-depth review measured outcomes based on a school level average across all grades in the schools (Table 4.9). Of the other studies, three included no measures of student attainment, and the others took data from specific grade levels, either at a single point in time, or over a defined period of time to obtain a gain score. Studies examining attainment in particular grades were most likely to do so at grade 8/year 9 (N=12 studies) and grade 10 /year 11 (N=8 studies). This is likely to reflect common test years. For example, in England and Wales, these years correspond to the ends of Key Stage 3 (age 13-

14) and Key Stage 4 (age 15-16), whilst in the USA grade 8 is the final year of middle school.

Table 4.9: Age at which outcome is assessed* (N=31 studies in the in-depth review)

Age	Number
Age 11-12, Grade 6, Year 7	3
Age 12-13, Grade 7, Year 8	2
Age 13-14, Grade 8, Year 9	12
Age 14-15, Grade 9, Year 10	3
Age 15-16, Grade 10, Year 11	8
Age 16-17, Grade 11, Year 12	7
Age 17-18, Grade 12, Year 13	4
Average score across all grades in school	8
Outcome is gain score	2
No student outcomes assessed	3

*Categories not mutually exclusive

Further details of the studies are provided in Appendices 4.2 (Aims and overviews of the studies), 4.3 (Sampling, student and school characteristics), and 4.4 (Summaries of the results and interpretations).

4.4 Weight of evidence (WoE)

During the data-extraction process, each study was assigned a weight of evidence based on four criteria: the overall methodological quality of the study within its own terms (A), appropriateness of the study design for answering the specific review question (B), the relevance of the study focus to answering the specific review question (C), and an overall weight of evidence (D) based on A, B and C. Details of the methods are given in section 2.3.3 and Appendix 2.4. The assessments of threats to validity are shown for particular outcomes in Appendix 4.5. The WoE judgments for all the studies are shown in Table 4.10.

Overall, nine studies were judged to provide a high/medium weight of evidence in answering this specific review question; nine studies were judged to provide medium (N=8) or medium/high (N=1) weight of evidence in answering the review question; and thirteen were judged to provide either medium/low (N=8), low/medium (N=3) or low (N=2) weight of evidence.

Table 4.10: Weight of evidence judgments for studies in the in-depth review

Study	A Internal coherence	B Appropriate design/analysis	C Relevance	D Overall
Abbott *et al. (*2002)	Medium	Medium	Medium	Medium
Atkinson and Wilson (2003)	High	Medium	High	High/Medium
Bedard *et al. (*1999)	Medium	Medium	Medium	Medium
Bickel and Howley (2000)	Medium	Medium	Medium	Medium
Bickel *et al. (*2001)	Medium	Medium	Medium	Medium
Bowen *et al. (*2000)	Medium	Low	Low	Low/Medium
Bowles and Bosworth (2002)	Low	Low	Low	Low
Bradley and Taylor (1998)	High	Medium	High	High/Medium
Bradley and Taylor (2003)	High	Medium	High	High/Medium
Driscoll *et al. (*2003)	High	Medium	Medium	Medium/High
Fetler (1997)	Medium	Medium	Medium	Medium
Gill *et al. (*2002)	High	Low	Low	Medium/Low
Heck (1993)	Medium	Low	Low	Low/Medium
Howley (1996a)	Medium/Low	Medium	Medium	Medium/Low
Howley (1999a)	Medium/Low	Medium	Medium	Medium/Low
Howley (1999b)	Medium/Low	Medium	Medium	Medium/Low
Johnson *et al.* (2002)	Medium/Low	Medium	Medium	Medium/Low
Kirjavainen and Loikkanen (1998)	High/Medium	Medium	Medium	Medium
Lee and Smith (1997)	High	Medium	High	High/Medium
Lee and Burkam (2001)	High	Medium	High	High/Medium
Leung and Ferris (2002)	Medium/Low	Medium	Medium	Medium/Low
Ma (2001)	High	Medium	High	High/Medium
McLaughlin *et al. (*2000)	High	Medium	High	High/Medium
McMillen *et al. (*2000)	Low	Low	Medium	Low
McNeely *et al. (*2002)	Medium/High	Medium	Medium	Medium
Silins and Mulford (2000)	Low/Medium	Medium	Medium	Low/Medium
Spielhofer *et al.* (2002)	High	Medium	High	High/Medium
Stiefel *et al. (*2000)	Medium	Medium	Medium	Medium
Taylor and Bradley (2000)	High	Medium	High	High/Medium
Welsh *et al. (*1999)	Medium	Low	Medium/Low	Medium/Low
Welsh *et al. (*2000)	Medium	Low	Medium	Medium/Low

4.5 Synthesis of evidence

The synthesis of evidence is based on the study outcomes and is divided into six categories: student achievement (sub-divided into studies which did not and did control for prior attainment), student behaviour and attitudes, teacher perceptions of school climate and organisation, school organisation and economic outcomes. In categories for which there are a large number of studies (e.g. student achievement), the synthesis is further broken down into sub-categories based on

weight of evidence judgments. Study details of the participants are given in Appendices 4.2 and 4.3, and the results in Appendix 4.4.

The data used in all the studies were obtained from either official data sets or author-designed questionnaire surveys. Official datasets are records collected on school performance, collected by either central or local government. These are usually based on annual returns, completed by the schools. In their analyses, the studies used either regression analysis, or partial correlation analysis, or analyses that considered interaction effects between at least two variables. The different methods of analysis, combined with a wide range of model specifications and a broad range of outcome variables, make numerical synthesis difficult and narrative synthesis complex.

The synthesis of study results draws both on the authors' interpretation of their study results and our (the review group's) interpretation. These are based on unstandardized coefficients, unless otherwise stated. In this section of the report, no actual coefficients are presented, as we felt that it was misleading to think that coefficients are directly comparable across different regression models, different populations and different outcome measures. The relevant coefficients for each study can be found in Appendix 4.4 alongside interpretations of the statistical results. Some authors also present standardized coefficients alongside the unstandardized coefficients. Standardizing a coefficient allows comparison of the relative 'effects' of different independent variables within the *same* regression model, as the variables are presented on the same measurement scale (standard deviation units). Comparing standardized coefficients *between* regression models can be misleading since the coefficients need to be interpreted relative to the estimated 'effect' on the outcome variable of the other variables in the model. No attempt to compute standardized coefficients has therefore been made in this study.

The majority of studies report a figure for statistical significance alongside their estimated 'effect' and some authors refer to these in their interpretation of the data. There are a number of limitations of this approach, both generally and specifically, to its use in this review. Firstly, statistical significance is an estimate of the likelihood of a particular result occurring by chance. Statistical significance is linked to sample size and the probability of a study being able to detect an 'effect' of a given magnitude. The absence of a statistically significant 'effect' does not necessarily indicate that there is no 'effect', but may simply be the result of the sample being too small to detect an 'effect' at the given significance level (by convention $p = 0.05$). None of the studies in the review supply information about prior sample size calculations. Secondly, inferential statistics are used for generalizing from a sample estimate to a target population. Many of the studies in the review use all schools in a state or country as their 'sample'. It is not clear, therefore, how estimates of 'effect' and their statistical significance should be interpreted.

This synthesis section is therefore focused on searching for patterns of similarity or differences in the direction of 'effects' (regardless of their statistical significance) across studies in each outcome category. Summary tables showing the direction of 'effect' as either increasing as school size increases (positive,+), decreasing as school size increases (negative, -), or increasing as school size increases and then decreasing (quadratic, ∩), or some combination of these, are therefore presented at the end of each outcome category, unless that study has not provided sufficient data for that outcome to be able to summarise the direction of 'effect'. Unless otherwise stated, directions of 'effect' that are statistically significant are indicated using the following notation in the summary tables: *=$p<0.05$, **=$p<0.01$, ***=$p<0.001$.

These summaries are based on the overall results (not sub-group analyses or interaction effects) of each of the studies in the most fully specified model presented.

The most obvious reasons for any differences in direction of 'effect' in all outcome categories are the different ways in which independent and dependent variables have been constructed, differences in the variables that have been entered into the regression models and the extent which the studies have dealt with threats to their validity (see Appendices 2.4 and 4.5). However, if any consistent patterns in direction of 'effect' emerge notwithstanding these differences, this would provide more confidence that such an 'effect' is 'real' and in the direction indicated.

In all except one of the studies considered, the approach used was to compare the outcome of one or more dependent variables across schools of different size (the independent variable). Where the method of analysis was some form of regression, a direction of 'effect' can be reported as outlined above. For example, a positive direction of 'effect' is interpreted by authors and in the synthesis below as meaning that, as schools get larger, then the result or score on the dependent variable also gets larger. However, it should be made clear that this does not mean that, as individual schools change size, the dependent variable scores change in a particular direction in that school; rather that the outcomes are on average different, in schools of different sizes.

4.5.1 Student outcomes: achievement

4.5.1.1 Achievement *without* control for prior attainment

Fifteen papers report on the 'effect' of school size without control for prior attainment. Two were judged to provide high/medium WoE (Bradley and Taylor, 1998; McLaughlin *et al.*, 2000); five were judged to provide either medium/high (Driscoll *et al.*, 2003) or medium (Abbott *et al.,* 2002; Bedard *et al.*, 1999; Bickel and Howley, 2000; Bickel *et al.,* 2001) WoE, five were judged to provide medium to low WoE (Gill *et al.*, 2002; Howley, 1996a; 1999a; 1999b; Johnson *et al.*, 2002) and three were judged to provide low/medium (Heck, 1993) or low WOE (Bowles and Bosworth, 2002; McMillen *et al.*, 2000).

High/Medium WoE

Bradley and Taylor (1998) use 1992-96 data taken from the Schools' Census and School Performance Tables for 2,864 secondary schools in England to examine the impact of a range of school level variables on exam performance as measured by the proportion of students obtaining A*-C grades in the GCSE. Using regression methods, the authors find a statistically significant non-linear relationship between school size and exam performance, both cross-sectionally (e.g. for each individual year) and over time (e.g. for the years 1992-96). They show exam performance increases with school size, but at a decreasing rate. Further analysis by the study authors shows that exam performance is maximised at a school size of around 1,200 for 11-16 schools and 1,500 for 11-18 schools. The authors report that the relationship is 'flat topped' for both groups of schools. Thus, for 11-16 schools, an increase in school size above 900 (but under 1,500) has very little 'effect' on exam performance; the same is true for 11-18 schools with above 1,200 (but under 1,800) students. Examining the change in exam performance the authors go on to state that an increase in school size of 100 students is associated with an increase of 0.7 percentage points in exam performance.

McLaughlin *et al.* (2000) use 1993-94 data from 20 states across the USA to investigate the potential value of linking routinely collected national and state level datasets. Using regression analyses and partial correlations, the study examines the inter-relationships between school size, achievement, school climate, teachers' perceptions of self-influence, normative cohesion and class size. Student test

score data for 496 schools with a grade 8, and 595 schools with a grade 11, are used to examine the possible relationship between school size and achievement. Results from partial correlations show a statistically significant negative relationship between size and achievement for middle schools, but not for secondary schools; whilst results of ordinary least squares (OLS) regression and simultaneous equation models show a statistically significant positive relationship between school size and test scores at the secondary level, but not at the middle school level. The results are inconsistent across analyses but suggest that at grade 11 achievement increases as school size increases, whilst for grade 8 the relationship is generally negative, suggesting that as school size increases achievement scores go down for middle schools.

Table 4.11: Direction of 'effect' and student achievement without control for prior attainment (High/Medium WoE)

Study	Direction of 'effect'		
	+	-	∩
Bradley and Taylor, 1998			∩***
McLaughlin *et al.*, 2000[1]	+*	-	

[1] *Results based on OLS regression model*

Medium/High WoE

Driscoll *et al.* (2003) use 1999 data from 5,525 schools in 755 districts in California to examine the impact of district size on test scores in middle and high school, whilst controlling for school level variables (e.g. school size, class size) and student variables (e.g. SES, parental income). The authors report no statistically significant relationships at middle school or at high school for academic achievement. Our interpretation of the results is that an increase in school size of 100 was associated with a decrease in average middle school API (academic performance index) of 0.21, and a decrease in high school API of 0.27.

Medium WoE

Abbott *et al.* (2002) use 2001 data from the Washington Assessment of Student Learning (WASL) from 417 schools, including 7th grade in Washington State. Using hierarchical linear modelling, they nest school level data in district level data to investigate the 'effects' of school size, district size and SES on achievement in mathematics and reading. The authors report no statistically significant 'effects' of school size on either mathematics or reading achievement after controlling for SES. Our interpretation of the results is that an increase in span size of 100 was associated with an increase in grade 7 mathematics achievement of four marks, and an increase in grade 7 English achievement of one mark.

Bedard *et al.* (1999) use 1998 Standardised Testing and Reporting (STAR) data from 801 middle schools and 618 high schools in California. The study examines the 'effects' of school size and other student level variables (e.g. ethnicity, SES, English proficiency) on the *distribution* of test scores at grades 8 and 10. The authors use both standard and modified ordered probit models (a form of regression modelling). Results for the standard ordered probit do not show a statistically significant relationship between school size and mathematics scores for either middle or high schools. However, the modified ordered probit model shows some evidence that small schools are less likely to produce disproportionately 'bad' outcomes (reported for grade 10 reading and math scores). In clarifying the findings, the authors were contacted and provided an updated unpublished paper (Bedard *et al.,* 2001).

Bickel (1999b) and Bickel *et al.* (2001) use a 1996-97 state dataset for Texas high schools. In the first paper (Bickel, 1999b), the author includes 1,441-1,448 grade 8 schools and 1,190-1,197 grade 10 schools. The study examines the relationship between school size, SES and achievement. In the second paper (Bickel *et al.,* 2001) the authors extend the research by including further explanatory variables and focusing on grade 10 results from 1,001 schools. The 1999 paper shows statistically significant positive relationships between school size and achievement across all grade levels and across each of the subject areas studied, suggesting that as school size increases achievement also increases. The paper also reports statistically significant negative interaction effects, showing that achievement in students from lower socio-economic backgrounds is negatively affected by larger school size. In the extension of this paper, when further explanatory variables are included in the model, the 'effect' of school size on achievement disappears at grade 10, although the negative interaction effect with size and SES remains statistically significant. Our interpretation of the these results is that an increase in span size of 1,000 was associated with increases in reading, mathematics and writing scores at grade 10 of 0.2, 0.02 and 0.05 respectively.

Bickel and Howley (2000) (see also Bickel, 1999a) use a 1996-97 Georgia State dataset to examine the relationship between school size, district size, SES and achievement. The authors present two papers: one (Bickel, 1999a) seeks to replicate other studies (see for example Bickel *et al.,* 2001) and the other uses hierarchical linear modelling (Bickel and Howley, 2000). Using data from more than 300 grade 8 and grade 11 schools, the 1999 paper shows a statistically significant positive relationship between school size and achievement across seven curricular areas for both grades 8 and 11. This result suggests that larger school size is associated with higher achievement within grades 8 and 11 across the curriculum. The 'effects' considering the interaction between SES, school size and achievement show a statistically significant negative relationship across curricular areas and grades, suggesting that achievement in students from lower socio-economic backgrounds is negatively affected by larger school size. The second paper (Bickel and Howley, 2000), using hierarchical linear modelling and specifying a greater number of explanatory variables, shows no statistically significant relationships between school size and achievement at either grade 8 or grade 11. A comparison carried out by the review authors of the results between the two models shows that, in the model of lower specification (1999), the results suggest that an increase in span size of 1,000 was associated with an *increase* in mean grade 8 composite achievement of 24 percentile points. In the more fully specified model (2000) an increase in span size of 1,000 was associated with a *decrease* in mean grade 8 composite achievement of six percentile points. This difference illustrates the importance of model specification in regression analysis.

Table 4.12: Direction of 'effect' and student achievement without control for prior attainment (Medium/High or Medium WoE)

Study	Direction of 'effect'		
	+	-	∩
Medium/High			
Driscoll *et al.*, 2003		-	
Medium			
Abbott *et al.*, 2002	+		
Bedard *et al.*, 1999[1]		-	

Bickel and Howley, 2000[2]	+	-	
Bickel *et al.*, 2001[2]	+		

[1]*This study is not about student attainment per se, but the distribution of test score results. The results in this table are taken from the standard ordered probit model.*
[2]*The results in the table are taken from the most fully specified models.*

Medium to low WoE

Gill *et al.* (2002) analysed data from 4,120 young people aged 15 in 155 schools that were collected as part of the OECD Programme for International Student Assessment (PISA). The study examines the 'effects' of school size and a range of other school (e.g. school sector, class size) and student level variables (e.g. SES, gender, first language) on literacy achievement in reading. The authors report that school size was initially considered in the regression model but that it was not found to be statistically associated with reading literacy, because another factor (school sector) with which it was highly correlated had greater statistical significance. The study reports that private schools made up the majority of the smaller schools in the sample and that this factor was more closely associated with achievement than size of school *per se*.

Howley (1996a) uses a 1990 West Virginia State dataset to examine the 'effect' of school size, district size and SES on achievement. Using data from 508 schools with a grade 6, 196 schools with a grade 9 and 106 schools with a grade 11, the study shows a statistically significant positive relationship between school size and achievement at grade 11, suggesting that, as school size increases, achievement in grade 11 also rises. This relationship was not found across other grades. Analysis of the interaction effects between size, SES and achievement show statistically significant negative relationships across the three grades, suggesting that students from lower socio-economic backgrounds are adversely affected by larger school size. Our interpretation of the grade 11 results is that, at grade 11, an increase in span size of 100 is associated with an increase in achievement of 2.6 marks in the comprehensive test of basic skills.

Howley (1999a) uses a 1998 Montana State dataset for 220 schools at grade 8 and 168 schools at grade 11 to analyse the 'effects' of school size and SES on achievement. Using regression analysis, the author reports no statistically significant relationship between school size and achievement at either grade. Analysis of the interaction between size, SES and achievement also shows no statistically significant 'effects'. Sub-analyses, based only on schools of popular grade configurations (grades 7-8 and 9-12), show some statistically significant relationships, but these are inconsistent across grades. Our interpretation is that the overall results show an increase in span size of 100%, which is associated with an increase in grade 8 achievement of 0.06 marks and a decrease in grade 11 achievement of 1.1 marks.

Howley (1999b) uses a 1995 state dataset for schools in Ohio (1,314 schools at grade 6, 811 schools at grade 9, and 650 schools at grade 12). The author investigated the 'effect' of school size and SES on the percentage of students passing the Ohio Proficiency Tests in reading, writing, mathematics, citizenship and science. Mixed results were found for the 'effects' of school size on percentage passing across grades 6, 9 and 12. For pass grades, no statistically significant relationship is found at grade 6, a negative statistically significant relationship is found at grade 9 and a positive statistically significant relationship is found at grade 12. For advanced pass grades, statistically significant positive relationships are found both at grade 6 and 12. The study shows a consistent negative interaction effect for size, SES and achievement across all three grades, suggesting that achievement in students from lower socio-economic backgrounds

is more negatively affected by larger school size. Our interpretation of these results is that an increase in span size of 100% is associated with a 0.87 decrease in percent passing at grade 6, 4.0 at grade 9 and a 2.1 increase in the percentage passing at grade 12.

Johnson *et al.* (2002) use 1998-2000 data from just over 300 schools in Arkansas. Using regression models similar to those of Howley (1999a, 1999b), the authors investigate the 'effects' of school size and SES on test scores at grades 7, 8 and 10. The authors report a statistically significant negative relationship between school size and grade 10 achievement scores, suggesting that in grade 10 larger school size is associated with lower achievement. This relationship was not found at either of the other grade levels. Statistically significant interaction effects were seen between size, SES and achievement across all three grades, suggesting that achievement in students from lower socio-economic backgrounds is negatively affected by larger school size. Our interpretation of the results is that an increase in span size of 100% is associated with a decrease of 0.44 marks in SAT tests at grade 7, and 2.26 marks in the SAT tests at grade 10. For the benchmark tests, an increase in span size of 100% is associated with a decrease of 1.5 (literature) and 1.1 (mathematics) marks.

Table 4.13: Direction of 'effect' and student achievement without controlling for prior attainment (Medium to Low WoE)

Study	Direction of 'effect'		
	+	-	∩
Gill *et al.*, 2002			
Howley, 1996a	+*	-	
Howley, 1999a	+	-	
Howley, 1999b	+***	-[1]	
Johnson *et al.*, 2002		-[2]	

[1] *Grade 6 p>0.05, grade 9 p<0.001*
[2] *Grade 10 p<0.001 other grades p>0.05*

Low/Medium WoE

Heck (1993) studies all elementary, intermediate and high schools in one state in western USA. Using school level data from 235 schools and survey responses from 176 principals, 3,976 teachers and 4,429 parents, the study examines how school characteristics and attitudes affect achievement, attendance and behaviour. The study reports a statistically significant negative relationship between school size and mathematics and reading achievement, suggesting that when school type (e.g. elementary, middle, high school) is controlled for, larger school size is associated with lower student achievement.

Low WoE

McMillen *et al.* (2000) use 1997 and 1998 data collected from 308 grades 6-8 schools and 292 grades 9-12 schools. The study uses partial correlation analysis to estimate the relationship between school size and achievement whilst controlling for other student level variables (i.e. ethnicity, eligibility for free school meals, percentage with parents with no formal education). No statistically significant relationship was found between school size and achievement at the high school level; however, the authors state that 'average achievement gains for the smallest middle schools were slightly higher than those of the other two groups of schools' (p 17).

Bowles and Bosworth (2002) use 1994-98 data from 17 Wyoming school districts, representing approximately 80 schools, to examine average costs per student across a four year time span. As part of their analysis of economies of scale in Wyoming schools, they also examine the impact of school size on test results. The study shows no statistically significant relationship between school size and test scores.

Table 4.14: Direction of 'effect' and student achievement without controlling for prior attainment (Low/Medium or Low WoE)

Study	Direction of 'effect'		
	+	-	∩
Low/Medium			
Heck, 1993		-[1]	
Low			
McMillen *et al.*, 2000		-	
Bowles and Bosworth, 2002		-	

[1] *Mathematics p<0.05, Reading p<0.01*

Summary

Relationship between school size and achievement without prior attainment

- The 15 studies that do not take into account prior attainment when considering the 'effects' of school size on achievement show inconsistent results. Approximately half of the studies show a positive relationship and half show a negative relationship. The reason for this inconsistency is not known, but it does not appear to be due wholly to grade level studied or subject area.
- The only English study in this category shows achievement maximized at a school size of about 1,200 (for 11-16 schools) and 1,500 (for 11-18 schools) students (a quadratic relationship) (Bradley and Taylor, 1998).
- The majority of studies in this category do not report any statistically significant association between school size and achievement. Our interpretation of the results is that comparatively large increases in school size were associated with comparatively small practical increases or decreases in test scores.
- The potential importance of model specification is clearly illustrated in this section. Two studies (Bickel and Howley, 2000; Bickel *et al*, 2001) found that that more detailed specification of their models produced very different results.

4.5.1.2 Student outcomes: achievement whilst controlling for prior attainment

Four studies analysed the relationship between school size and student achievement, whilst controlling for prior attainment. All these studies were judged to provide high/medium WoE (Atkinson and Wilson, 2003; Bradley and Taylor, 2003; Lee and Smith 1997; Spielhofer *et al.,* 2002). Three of these studies (Atkinson and Wilson, 2003, Bradley and Taylor, 2003 and Spielhofer *et al,* 2002) use the same nationally collected DfES datasets. This means that the data on school size used in the regression model in each study comes largely from the same schools. However, the student achievement outcomes in each study are based on different study subjects and/or examination years, so each individual student will only be 'counted' in one of the studies. The results of these three studies should therefore not be seen as entirely independent of one another.

Atkinson and Wilson (2003) use 1997 and 1999 academic test data taken from DfES matched exam datasets for 517,695 students in 3,129 secondary schools in England. This study uses exam results for a single cohort for which data were collected at Key Stage (KS) 3 and GCSE/GNVQ (General national vocational qualification) in order to examine patterns of attainment and possible factors that may affect student attainment. The results include both raw scores and value-added. The authors report a nonlinear (quadratic) relationship between school size and achievement for KS3 results and across a range of GCSE subjects, and for analyses both with, and without, value added. The relationship is shown to increase with increasing size up to a maximum of between 1,280 and 2,155 students, depending on the subject area. This general pattern of non-linearity is similar to that described previously in Bradley and Taylor (1998).

Bradley and Taylor (2003) use multiple regression analysis to estimate the relationship between school size and KS4 score (GCSE) and value added between KS3 and KS4 (GCSE). The results of this analysis found a statistically significant relationship between school size and both types of outcomes. In both analyses, achievement increased as school size increased up to a certain point. Beyond this point, achievement decreased as school size increased (a quadratic relationship). Uniquely amongst all studies in the review, this study also investigated the relationship between change in the size of individual schools and the change in GCSE exam results between 1993 and 2002 for 3,098 schools in England. After controlling for a range of variables relating to the characteristics of students and the characteristics of the school they attended, they find a significant positive relationship between the *change* in school size and the *change* in the proportion of students obtaining five or more A* to C grades in the GCSE exams. Specifically, they find that an increase of 100 in the number of students on the school roll is associated with an increase in the percentage of students obtaining 5 or more A* to C grades of around one percentage point. This compares with the average increase over all schools of 7.4 percentage points during 1993-99 (from 38.4 to 45.8). Average school enrolment increased by around 100 students (from 830 to 927) between 1993 and 1999.

Lee and Smith (1995, 1997; Lee *et al.,* 1997) use 1988-92 data collected as part of the USA National Educational Longitudinal Study (NELS) to investigate student and school level factors affecting achievement gains and engagement. The study draws on data for 11,794 students in 820 schools in grades 8 to 10 (Lee and Smith, 1995), and 9,812 students in 789 schools in grades 8 to 12 (Lee and Smith, 1997). The results in the 1995 paper report a statistically significant negative relationship between school size and achievement gains in mathematics, reading, history and science, suggesting that larger schools are associated with smaller achievement gains. Additional analyses show that, in terms of achievement, smaller schools provide more equitable learning environments for students from lower socio-economic backgrounds. Similar results are presented in the Lee *et al.* (1997) paper for achievement gains in grades 10-12, both in terms of the direct 'effect' on achievement, and the equity of the learning environment. In the Lee and Smith (1997) paper, the achievement gains over grades 8 to 12 are examined and show a non-linear (quadratic) 'effect', with achievement gains in mathematics and reading maximised in schools of size 601-900. Although this is a smaller figure than that reported in studies of schools in England, the enrolment is across four grades rather than five (as in English 11-16 schools) or seven (as in English 11-18 schools). If the average number of students per year is calculated, then it appears to be similar across the studies. Separate analyses, based on 'effects' in high and low socio-economic schools, and high and low minority enrolment, show similar results.

Spielhofer *et al.* (2002) use 1996-2001 data taken from national student level datasets collected by the DfES and covering 369,341 students in 2,954 schools in England to estimate the impact of school size on performance at KS4. Initial attainment is taken into account by including KS2 scores for each individual in the sample. The main finding with respect to school size is that attainment improves as year group size increases up to a certain point, and then declines. The authors report that attainment in most subjects is maximized in schools with between 175 and 200 students in year 11, and the greater the prior attainment of a child, the larger the optimum school size for that child. The optimum school size for mixed sex schools was the same for boys and girls, although lower for single sex schools (particularly girls' schools). Additional analyses found that the optimum size for schools with low numbers of children eligible for free school meals was much lower than that in schools with high numbers of students eligible for free school meals, suggesting that in schools where there was a small percentage of students eligible for free school meals, the performance of students tended to be better when the school was smaller.

Table 4.15: Direction of 'effect' and student achievement whilst controlling for prior attainment

Study	Direction of 'effect'		
	+	-	∩
Atkinson and Wilson, 2003			∩**
Bradley and Taylor, 2003	+[1]		∩[2]
Lee and Smith, 1997			∩[3]
Spielhofer *et al.*, 2002			∩*

[1] *Direction of 'effect' refers to change in school size and change in exam performance p<0.000*
[2] *Directions of 'effect' refers to change in school size and KS4 and value added KS3 to KS4.*
p<0.05 value added scores (controlling for prior attainment),
p<0.001 scores without controlling for prior attainment.
[3] *Categorical analysis; 'effects' for schools of different size categories when compared to the 'effect' of a school with 1200-1500 students, for mathematics and reading (for fuller details, see Appendix 4.4).*

Summary

Student achievement whilst controlling for prior attainment

- The four studies in this section are consistent in finding that achievement increased as school size increased, up to a certain point (or range). After this point, achievement decreased as school size increased. Three of these studies use a dataset which contains largely the same schools but the outcomes reported are for students in different years and/or curriculum subjects.
- Three of the studies provide an estimate of school size at which attainment is maximised. Atkinson and Wilson (2003) found that achievement is maximised in schools of between 1,150 and 2,155 students, depending on the subject; Spielhofer *et al.* (2002) found that achievement is maximised at 875-1,000 students (schools 11-16), and 1,225-1,400 students (schools 11-18) (175-200 students per year). The study by Lee and Smith (1997) found that achievement was maximized at 600-900 students (grades 9-12). This converts to year groups of between 150 and 225 students, which is similar to the study by Spielhofer *et al.* (2002).

4.5.1.3 Student outcomes: overall attainment

The previous sections have synthesised information about attainment within WoE categories, and reported separately studies which did and did not control for prior attainment. This categorisation was pre-specified in Chapter 2 as potentially likely to explain variations in the direction of the 'effect'. Table 4.16 presents information on attainment, regardless of whether the study design controlled for prior achievement. With one exception, studies with a high/medium WoE found a quadratic relationship: that is, attainment increased with school size up to a certain point, after which increases in size were associated with a decline in attainment. A positive 'effect' was more likely to be found in the studies classified as providing medium WoE, and the lower the WoE, the more likely studies were to report negative directions of effect. Finding this consistent pattern of direction of effect across a number of high quality studies provides more confidence that a quadratic relationship exists between school size and attainment, where student attainment first increases and then decreases, as schools get larger . We may, however, speculate about other factors which might explain some of the heterogeneity in the statistical results.

Table 4.16: Direction of 'effect' and student attainment

Study	Direction of 'effect'		
	+	-	∩
High/Medium WoE			
Atkinson and Wilson, 2003			∩**
Bradley and Taylor, 2003			∩[1]
Lee and Smith, 1997			∩[2]
Spielhofer *et al.*, 2002			∩*
Bradley & Taylor 1998			∩***
McLaughlin *et al.,* 2000	+*	-	
Medium/High WoE			
Driscoll *et al.,* 2003		-	
Medium WoE			
Abbott *et al.,* 2002	+		
Bedard *et al.,* 1999		-	
Bickel and Howley 2000	+	-	
Bickel *et al.,* 2001	+		
Medium/Low WoE			
Gill *et al.,* 2002			
Howley, 1996a	+*	-	
Howley, 1999a	+	-	
Howley, 1996b	+***	-[3]	
Johnson *et al.,* 2000		-[4]	
Low/Medium WoE			
Heck, 1993		-[5]	
Low WoE			
Bowles and Bosworth, 2000		-	
McMillen *et al.,* 2000			

[1] *p<0.05 value added scores (controlling for prior attainment), p<0.001 scores without controlling for prior attainment.*

[2] *Categorical analysis; 'effects' for schools of different size categories when compared to the 'effect' of a school with 1200-1500 students, for mathematics and reading (for fuller details, see Appendix 4.4).*

[3] *Grade 6 p>0.05, grade 9 p<0.001*

[4]*Grade 10 p<0.001 other grades p>0.05*
[5]*Mathematics p<0.05, reading p<0.01*

WoE categorisation does appear to explain some of the differences in the results obtained. The overall WoE (D) provides an indication of a study's methodological quality (A), the appropriateness of its design and analysis for the review (B) and the extent of its relevance for the review (C). Some of the methodological issues were explored in more explicit detail in the 'threats of validity' (see Appendices 2.4 and 4.5), and the information from this was closely correlated with WoE A. The few discrepancies were generally explained by lack of clarity in the reporting.

Other potentially relevant factors relate to the model used, the setting in which the studies were carried out, and the particular outcomes which they measure. A positive then negative direction of effect (i.e. quadratic) could only be detected if studies had used appropriate methods to do so. Most of the studies showing a negative direction of 'effect' were based on outcomes measured at younger ages. This may imply that older children do better in bigger schools where there is more likelihood of a broader range of specialist teaching, or that there is more dropping out amongst children who do badly in larger schools. However, the evidence about dropping out is equivocal (see section 4.5.2.1).

4.5.1.4 Interaction effects between SES and school size

Another possibility explored in some of the studies is that the association between school size and attainment is different for students at different levels of SES. A number of studies together form a series which have focused on the association between school size, SES and student attainment in seven different states in the USA, using similar methodology. In West Virginia (Howley, 1996a), Georgia (Bickel and Howley, 2000), Texas (Bickel *et al.*, 2001), Ohio (Howley, 1999b) and Arkansas (Johnson *et al.*, 2002), the interaction effect of school size and SES on student attainment was found to be negative and statistically significant. In Montana (Howley, 1999a) and Washington (Abbott *et al.*, 2002), the 'effect' was negative but not statistically significant. The authors' interpretation of these results is that the influence of school size on attainment is not constant, but varies with changes in the SES of the school's students, and that larger school size benefits more affluent students while smaller schools benefit more impoverished students. The authors also conclude that small schools may have what they term 'an equity effect': that is, may serve to reduce the differences in attainment that are associated with differences in SES.

The interaction between SES and school size was also investigated in two of the studies that did control for prior attainment (Lee and Smith, 1997, Spielhofer *et al.*, 2002). Whilst the study by Lee and Smith (1997) shows a similar pattern to the studies which did not control for prior attainment, the study by Spielhofer *et al.* (2002) shows the opposite pattern, suggesting that the optimum size of school was larger for schools with a large proportion of students eligible for free school meals.

It is unclear why studies which have investigated the interaction of school size and socio-economic status in the USA should have produced different results from the only study to have investigated the issue in England. Possible reasons may include the very marked difference in the range of school sizes in the USA and/or markedly different levels of socio-economic inequality in the two countries.

Unlike the pre-specified categorisations by WoE and by whether a control for prior achievement was included, exploration of possible explanations for the variations

in directions of 'effect' for 'attainment as a whole' and in interaction effects were carried out *post hoc* and must therefore be considered speculative.

4.5.2 Student behaviour and attitudes

Studies in this outcomes category were divided into three subgroups: (i) dropout behaviour and absence, (ii) attitudes and perceptions of school, and (iii) behaviour.

4.5.2.1 Dropout behaviour and absence

Five studies were identified that investigated the relationship between school size and dropout or absence: two were judged to provide high/medium WoE (Bradley and Taylor, 2003; Lee and Burkam, 2001), one was judged to provide medium WoE (Fetler, 1997) and two were judged to provide either low/medium (Heck, 1993) or low WoE (McMillen *et al.*, 2000).

High/Medium WoE
Lee and Burkam (2001) use 1990 and 1992 data obtained from the National Educational Longitudinal Survey (NELS), collected at grades 10 and 12 for 3,840 students in 190 high schools in the USA. The study uses multilevel modelling to examine the factors that are associated with dropping out at school level (e.g. size, sector) and at student level (e.g. gender, ethnicity, SES, academic background). The authors report that the statistical probability of dropping out of high school is not significantly different for those attending schools which they define as of medium size (601-1,500 students) compared with small schools (<600 students), but attending a large (1,501-2,500 students) or a very large school (>2,501 students) significantly increases the probability of dropping out. For a large school, the authors report that there was nearly a 300% increase in the odds of dropping out ($p<0.001$) compared with being in a medium school, whilst in small schools there was a more than 100% increase in the odds of dropping out compared with a medium school ($p<0.05$). In clarifying the findings, the authors were contacted and provided a helpful published version of this study (Lee and Burkam, 2003).

Bradley and Taylor (2003) also investigated the 'effect' of school size on truancy and absence. The study reports no relationship between a school's truancy rate and school size; however, the overall absence rate was found to be statistically significantly related to school size. This relationship is described as non-linear, with the absence rate first falling as school size increases and then increasing as school size increases. The authors report that the results suggest that the absence rate is at a minimum at approximately 1,400 students.

Medium WoE
Fetler (1997) uses a 1993-96 state dataset for 805 high schools in California. Using regression analysis, the study analyses the impact of school level factors (e.g. percentage of new teachers, annual faculty growth) and student level factors (e.g. percentage eligible for free school meals) on dropout rates. A statistically significant positive relationship between school size and dropping out is reported, suggesting that, as school size increases, the likelihood of dropping out also increases. The results indicate that an increase of 1,000 students was associated with an increase in the dropout rate (defined as percentage of students leaving the courses over a given year) by 1%.

Low/Medium WoE
Heck (1993) reports a statistically significant negative relationship between attendance and school size, after controlling for school type (e.g. elementary, middle or high). The study suggests that as school size increases attendance in school decreases.

Low WoE
McMillen *et al. (*2000) also considered the 'effects' of school size on dropping out. This study reports no statistically significant relationship between high school size and the dropout rate after controlling for SES.

Table 4.17: WoE and direction of 'effect' on dropping out and absence

Study	Direction of 'effect'		
	+	-	∩ / ∪
High/Medium WoE			
Bradley and Taylor, 2003			∪[1]
Lee and Burkam, 2001			∪[2]
Medium WoE			
Fetler, 1997	+***		
Low/Medium WoE			
Heck, 1993		-**	
Low WoE			
McMillen *et al.,* 2000			

[1] *Absence P<0.01, truancy p>0.05*
[2] *School size <600 p>0.05, school size 1,500-2,500 p<0.001, school size >2,500 p<0.01*

Table 4.17 shows some consistent patterns across the studies in this category. Studies graded high/medium weight of evidence show a relationship where truancy and absence first decrease and then increase as schools get larger (a quadratic relationship). The point at which dropping out and absence is minimised is broadly consistent between the two studies, once differences in grade span are taken into account. Although the studies coded either medium or low/medium weight of evidence show different directions of 'effect', the interpretation of the 'effect' is consistent. In the study by Fetler (1997), increases in school size are associated with an increase in dropping out whilst in Heck (1993) increases in school size are associated with decreases in the rate of attendance.

As in the results obtained for attainment, it does appear that weight of evidence explains at least some of the variation in the results. The studies with a higher weight of evidence suggest absence is minimised in schools with between 600 and 1,500 students. However, given that there are only two studies in this category, such a conclusion should be regarded as tentative. Further speculation about the differences in results across weight of evidence is difficult because the research methodologies used in the studies graded medium or low/medium weight of evidence would not have been able to pick up any quadratic relationships. Lee and Burkam (2001) focus on drop- out from grades 10-12, whilst Fetler (1997) measures dropout across high schools as a whole, including grade 9 when a significant number of students leave, and Heck (1993) presents an overall figure for attendance across elementary, middle and high schools. These differences in study samples may in part explain the differences in findings between the North American studies.

4.5.2.2 Attitudes towards school and perceptions of school

Four studies have examined the relationship between school size and attitudes towards, or perceptions of, school. One study was judged to provide high/medium WOE (Lee and Smith, 1997), one was judged to provide medium WoE (McNeely *et al.*, 2002), and two were judged to provide low/medium WoE (Bowen *et al.*, 2000; Silins and Mulford, 2000).

High/Medium WoE

As well as investigating student achievement in their 1997 paper (see section on achievement 4.5.1), Lee and Smith (1995) also consider the student and school level factors that affect academic engagement. Academic engagement is measured through a questionnaire to students, asking them to state whether they often work hard or feel challenged in a range of subject classes. The study shows a statistically significant negative relationship between school size and engagement. This suggests that, as schools get larger, levels of engagement as defined by self-report of working hard and feeling challenged decrease. Additional analyses do not, however, reveal a relationship between SES, school size and engagement, suggesting that there is equity amongst the levels of engagement (or disengagement) in students of higher and lower SES in larger and smaller schools.

Medium WoE

McNeely *et al.* (2002) use 1994 data collected as part of the United States adolescent health survey (Add Health, 2003) to examine the 'effects' of school structure and environment on school connectedness. School connectedness is measured using a student questionnaire, asking whether students are happy and feel safe in their school; whether they are close to people and feel part of the school; and whether they perceive teachers to be treating students fairly. The study uses hierarchical linear modelling and data from 71,515 students in 127 schools across America. The study reports a statistically significant negative relationship between school size and school connectedness, suggesting that, as school size increases, the degree of perceived school connectedness is reduced. The strength of this association is reported as being fairly weak: an increase of 500 students in school size is associated with a decline in connectedness of 0.4 units on a 1-4 ratings scale. The authors go on to report that school size appears to mediate the association between participation in extra-curricular activities and mean school connectedness.

Low/Medium WoE

Silins and Mulford (2000) use path analysis to consider the relationships between 12 potentially interrelated variables including school context variables (e.g. size and SES), internal school variables (e.g. availability of resources) and student outcome variables (e.g. student participation and engagement). Participation is defined as the extent of students' participation in curricular and extra-curricular activities, setting own learning goals and voicing opinions in class. Engagement is defined as the extent to which students feel that teachers related to them, perceptions of their relationships with peers, their perceptions of the usefulness of their schoolwork in later life, and the extent of their identification with their school. Both variables are measured through student questionnaires. The study investigates 3,500 Year 10 students drawn from high schools in Tasmania and South Australia, using data from a national longitudinal research study (Leadership for organisational learning and student outcomes (LOLSO)). The authors report that school size had a statistically significant negative 'effect' on student participation, suggesting that student participation was greater in smaller schools. The 'effect' of school size on engagement was indirect (e.g. having an 'effect' through another variable) and much smaller in magnitude than that of participation.

Bowen *et al.* (2000) use data from 945 students from a total of 39 schools across the USA to examine the 'effect' of school size, gender, ethnicity and SES on student perceptions of school satisfaction, teacher support and school safety. The study only reports where statistically significant relationships were observed. No results are reported for the relationship between school size and SES on any of the three outcome measures.

Table 4.18: Direction of 'effect' and student attitudes towards school and perceptions of school

Study	Direction of 'effect'		
	+	-	∩
High/Medium WoE			
Lee and Smith, 1997		-*	
Medium WoE			
McNeely *et al.,* 2002		-**	
Low/Medium WoE			
Bowen *et al.,* 2000			
Silins and Mulford, 2000		-[1]	

[1] *The authors report that the results were statistically significant without presenting p-values.*

Table 4.18 shows that the results for student attitudes are consistent in all the studies in this category across all weights of evidence. All studies show an association between increases in school size and decreases in student engagement, despite different conceptualisations of 'engagement' and using a variety of survey outcome measures.

4.5.2.3 Behaviour

Six studies examined the relationship between school size and student behaviour in terms either of violence, bullying, suspensions or incidents. One study was judged to provide high/medium WoE (Ma, 2001), three studies were judged to provide medium/low WoE (Leung and Ferris, 2002; Welsh *et al.*, 1999; Welsh *et al.*, 2000) and two studies were judged to provide either low/medium (Heck, 1993) or low WoE (McMillen *et al.*, 2000).

High/Medium WoE

Ma (2001) used the 1996 New Brunswick School Climate Study of 6,883 sixth grade students and 6,868 eighth grade students in middle schools to investigate bullying and victimisation. The study uses hierarchical linear modelling to consider the student and school level characteristics that are associated with being a bully and being a victim. Being a bully is assessed using a student questionnaire, asking whether the students had participated in bullying and teasing of others; whilst being a victim is assessed using a student questionnaire, asking whether they had been threatened or physically attacked by other students, or been afraid to go to school. The author reports a statistically significant negative 'effect' of school size on bullying at grade 8 and grade 6, suggesting eight and sixth grade students in smaller schools tend to bully more. School size was less likely to be associated with being a victim; a statistically significant relationship was reported only at grade 8, suggesting that grade 8 students were more likely to be victims of bullying in smaller schools.

Medium/Low WoE

Leung and Ferris (2002) use data from the 1995 Montreal Longitudinal Study collected for 17 year-old males. They analyse data for 616 students in 107 secondary schools in Montreal and use regression analysis to estimate the 'effect' of school size on self- reported involvement in youth violence, whilst controlling for other demographic variables. The authors report two models, one measuring school size as a continuous variable and the other as a categorical variable. Analyses of school size as a continuous variable show a statistically significant positive relationship, suggesting that as school size increases, so too does violence. The estimated marginal effects revealed that an increase of 1,000 students is associated with about a 10% increase in the probability of violent behaviour. Using the categorical school size variable, a statistically significant positive relationship was found between violence and schools of over 2,000 students, suggesting that students in schools enrolling over 2,000 students are more prone to violent behaviour. The estimated marginal effects suggested that students who attend schools with more than 2,000 students are about 22% more likely to engage in violent behaviour than those who attended schools with less than 1000 students.

Welsh *et al.* (1999) is a study related to that of Welsh *et al.* (2000) but uses different methods of data-collection and analysis over a different time period. The study has the same aims as Welsh *et al.* (2000) but analyses 1994 data collected from a survey of 6,693 students carried out in 11 middle schools in Philadelphia. The study employs hierarchical linear modelling to examine student and school variables, and reports no statistically significant relationship between school size and school disorder when measured as student report of self-misconduct (e.g. self-report of removal from class, detention, suspension or fighting to protect themselves).

Welsh *et al.* (2000) use 1990-93 secondary data collected from 43 Philadelphia middle schools, serving approximately 34,000 students. Using path analysis, the study assesses the direct and indirect factors affecting school disorder (measured by school incident and discipline data). The analysis sets up a hypothesised causal model in which school size is one factor which affects school disorder. However, it is hypothesised that any 'effect' of school size on school disorder is mediated by 'school stability'. The study uses two models. In one model, the data for the independent variables used to create the factors in path analysis are based on the community in which the school is located. In the second model, the data for the independent variables are based on the community in which the students actually live. The authors report that school size exerts a moderate indirect effect on school disorder through school stability. This relationship was only statistically significant in the model using data from the community in which the students lived. The coefficient of school size and school stability indicates a negative direction of 'effect' that suggests that school stability is lower in larger schools.

Low/Medium WoE

The study be Heck (1993) described in the section on achievement (section 4.5.1) considers the school and student level factors affecting the rate of suspensions. The study reports no statistically significant relationship between school size and the number of student suspensions for significant offences.

Low WoE

The study by McMillen *et al. (*2000) described in the section on achievement (section 4.5.1) reports no significant relationship between violence and the size of high schools, but does report a statistically significant positive relationship between the size of middle schools and violence rates. These results suggest that, as middle school size increases, the violence rate also increases.

Table 4.19: Direction of 'effect' and student behaviour

Study	Direction of 'effect'		
	+	-	∩
High WoE			
Ma, 2001		-[1]	
Medium/Low WoE			
Leung and Ferris, 2002	+[2]		
Welsh *et al.,* 2000	+[3]		
Welsh *et al.,* 1999		-[4]	
Low/Medium WoE			
Heck, 1993		-	
Low WoE			
McMillen *et al.,* 2000	+		

[1] *Grade 6 bullies p<0.05, grade 6 victims p>0.05, grade 8 victims and bullies p<0.01*
[2] *Continuous variable p<0.05, categorical variable p<0.014*
[3] *Negative correlation coefficient given in results indicates that school instability is greater in larger schools hence +ve direction of 'effect' shown in table. Direction of 'effect' on school stability as mediator of school disorder, p<0.05 in one version of model (see text).*
[4] *Lower scores indicate higher levels of misconduct, therefore the positive correlation coefficient shown in the study results means that there are lower levels of self-reported violence.*

The summary of directions of 'effect' for school size and student behaviour in Table 4.19 illustrates the inconsistency of results in the included studies. This inconsistency does not appear to be linked to weights of evidence.

Summary

Student behaviour and attitudes

- Studies providing the highest weight of evidence suggest that overall absence is minimised in schools with between 600 and 1,500 students. The English study (Bradley and Taylor, 2003) provided a point estimate for the minimisation of overall absence of 1,400 students. The small number of studies in this category suggests caution is required against being over-confident of this pattern of 'effect'.
- Studies with a lower weight of evidence tend to show a negative association with school size: that is, an increase in the dropout rate and a reduction in attendance as schools get larger.
- All the studies considered found a consistent negative association between students' feelings of engagement, connectedness and participation, and increased school size.
- Studies investigating the relationship between school size and violent student behaviour had somewhat contradictory findings: for some types of violent behaviour results suggest a positive association (i.e. violence increases as school size increases); for other types of violent behaviour, results suggest a negative association (i.e. violence increases as school size decreases).
- Many of these relationships are comparatively weak, and are difficult to quantify and conceptualise as they are based on multiple responses to questionnaires across a number of areas which are then factor-analysed to create constructs, such as 'connectedness' or 'engagement'.

4.5.3 Teacher perceptions of school climate and organisation

Two studies, one providing high/medium WoE (McLaughlin *et al.,* 2000) and one providing low/medium WoE (Silins and Mulford, 2000), consider the 'effect' of school size on teacher perceptions. Both studies have been described in previous sections, the former in the section on student achievement (section 4.5.1) and the latter in the section on student behaviour (section 4.5.2).

McLaughlin *et al.* (2000) also considered the influence of school size on the following constructs: school climate, perceptions of influence and normative cohesion. School climate is defined as teacher identification of problems, such as absenteeism, physical attacks and vandalism. Perceptions of influence are defined as teacher perceptions of their control over the class and control over school policies, and normative cohesion is defined as clarity of norms and co-operation amongst staff. The authors report statistically significant negative relationships between high school size and school climate. They also suggest that teachers in bigger schools tend to have more negative perceptions of their abilities to influence school policies and control their classrooms. However, although relationships between the above constructs and middle school size are consistently negative across two statistical models, they are not consistently statistically significant across the two models. The results suggest that teachers in smaller high schools perceive fewer problems with student behaviour, a greater degree of influence over school and classroom policies, a greater clarity of norms between staff and greater cooperation amongst staff.

The second study by Silins and Mulford (2000) reports that school size had a statistically significant negative relationship with teacher's perceptions of resource adequacy. This result suggests that teachers in smaller schools are more likely to perceive that they have adequate resources. No other statistically significant relationships were identified for other school level variables in the analysis.

Table 4.20: Direction of 'effect' and teachers' perceptions of school climate and organisation

Study	Direction of 'effect'		
	+	-	∩
High/Medium WoE			
McLaughlin *et al.,* 2000		-[1]	
Low /Medium WoE			
Silins and Mulford, 2000		-[2]	

[1] *p<0.05 for middle and high school climate, perceptions of influence and normative cohesion, exception middle school normative cohesion p>0.05. These results are based on the OLS models.*
[2] *Reported as statistically significant but authors do not present p-values.*

Table 4.20 shows that both studies investigating teacher perceptions show negative relationships. As school size increases, teachers tend to have fewer positive perceptions. This relationship holds across the different outcomes measured in the studies, including school climate, ability to influence, normative cohesion and resource adequacy.

Summary

Teacher perceptions of school climate and organisation

- Comparatively few studies included in the in-depth review included measures of the impact of school size on teachers.
- Evidence suggests that teachers in smaller schools tend to have more positive perceptions of school climate and of their abilities to influence school policies and control their classrooms. In addition, teachers in smaller schools also perceive clearer norms and greater levels of co-operation (normative cohesion) and resource availability.
- The result regarding school climate appears to be in contrast to the more mixed evidence from studies measuring actual or student perceptions of school disorder, bullying and suspensions which in some studies were less prevalent in larger schools
- As in the previous section, these relationships are difficult to quantify as they are constructs based on multiple responses across a number of areas brought together using factor analysis.

4.5.4 School organisation and structure

Two studies, both described previously and both providing high/medium WoE (McLaughlin *et al.,* 2000, Spielhofer *et al.,* 2002), consider very different aspects of school organisation.

McLaughlin *et al. (*2000) examine the relationship between school and class size. Class size is defined as a construct based on factor analysis of the average class size, the average student teacher ratio and teacher satisfaction with class size. The study shows a statistically significant positive relationship between school and class size for both middle and high schools, suggesting that, as school size increases, so does the *construct* class size.

Spielhofer *et al. (*2002) consider the differences in opportunities available to students in terms of entry to higher KS3 tiers and likelihood of taking popular GCSE subjects. The authors report that students in smaller comprehensive schools had a statistically significant greater chance of being entered for the higher tier in mathematics, but not for science, than those in medium schools. There were no differences in the likelihood of being entered into higher tiers for students in larger schools, compared with those in medium schools. In comparison with students in medium schools, students in smaller schools were less likely to be entered into either double science or single science (chemistry, physics and biology). Students in large schools were more likely to be entered into these subjects than students in medium-sized schools. Students in small schools were more likely to be entered for the food technology and resistant materials options of the design and technology GCSE than students in middle schools, but less likely to be entered into the graphics option. Students in large schools were less likely to be entered for both French and German GCSE than students in medium-sized schools.

Table 4.21: Direction of 'effect' and school organisational outcomes

Study	Direction of 'effect'		
	+	-	∩/ ∪
High/Medium WoE			
McLaughlin *et al.,* 2000	+*		
Spielhofer *et al.,* 2002	+[1]	-[1]	

[1] *p-values differ depending on the subject; however, all are significant to at least the p<0.05 level.*

Table 4.21 shows the directions of 'effect' for the two studies in this category. Both show different 'effects' and the studies measure very different dependent variables, and therefore cannot be synthesised together.

Summary

School organisation and structure

- Comparatively few studies in the in-depth review include measures of the impact of school size on elements of school structure and organisation. The two studies that include such outcome use very different measures and are not comparable.
- One study (McLaughlin *et al.,* 2000) provides evidence that, as school size increases, so too does the *construct* of class size based on average class size, student teacher ratios and teachers' perceptions of satisfaction with their class size.
- The second study (Spielhofer *et al.*, 2002) provides some evidence that students in smaller schools may be more likely to be entered into higher tiers for mathematics, but not for science.
- This study also provides some evidence that students in smaller schools may be less likely to be entered for some GCSE subjects. However, this pattern is not consistent across all subject areas.

4.5.5 Economic outcomes

Five studies investigated economic outcomes. In all the studies, the 'costs' measured refer to direct public expenditure on schools. Three studies investigated the variation in costs in schools of different sizes (Bickel *et al.*, 2001; Bowles and Bosworth, 2002; Taylor and Bradley, 2000). One investigated variations in costs and variations in costs per unit of output (Stiefel *et al.*, 2000) and one investigated variation in costs per unit of output alone (Kirjavainen and Loikkanen, 1998). One study was judged to provide high/medium WoE (Taylor and Bradley, 2000), three were judged to provide medium WoE (Bickel *et al.*, 2001; Kirjavainen and Loikkanen, 1998; Stiefel *et al.*, 2000), and one study was judged to provide low WoE (Bowles and Bosworth, 2002).

High/Medium WoE

Taylor and Bradley (2000) use data for between 2,034 and 3,087 secondary schools in England collected between 1993 and 1997 to examine the average cost per student in the secondary school sector. Since cost data were not available for all schools, teaching and support hours per student were used to approximate cost per student. The aim of the study is to estimate the separate 'effects' on costs per

student of two factors: (i) school size and (ii) the extent to which a school's capacity is being utilised. School size is measured in terms of each school's capacity and the utilisation rate is measured by the ratio of actual students to capacity. The school's capacity utilisation rate is included as an explanatory variable in order to distinguish between the short-run 'effects' of changes in student numbers on cost per student and the long-run 'effects' that arise as a result of changes in the school's scale of operation. The authors find that both teaching costs per student and support costs per student fall as school size increases. They also find that teaching costs per student and support costs per student fall as a school's capacity utilisation rate increases. In addition, the study also found that schools that increased their capacity utilisation rate and/or grew in size between 1993 and 1997 experienced a reduction in their costs per student. Their results indicate that an 'increase of 10% in student capacity is associated with a decline in support hours per student of 2%; and an increase of 10% in the capacity utilisation rate is associated with a decline in support hours per student of 3%' (p 137). The equivalent estimated reductions in teaching hours per student are approximately 1% and 1.5% respectively.

Medium WoE

Bickel *et al.* (2001) use 1996-97 Texas State data for 1,001 high schools to analyse the relationship between school size, SES and achievement. The study seeks to replicate a number of previous studies (Howley, 1996a; Howley, 1999a; 1999b) and to extend them by incorporating further explanatory variables into the regression model, and including measures of expenditure as a dependent variable. Using regression analysis, the authors report a statistically significant negative relationship between school size and expenditure per student, suggesting that as school size increases costs per student decrease. The estimated partial derivatives show diminishing costs with each increment in size, suggesting that as school size increases expenditure per student decreases, but at a decreasing rate as size increases. The authors estimate the 'effects' of school size using two different models. Our interpretation of the main findings is that the results for the first expenditure model suggest that an increase in span size of 100% is associated with a decrease in expenditure of $254.4 per student per year. The second expenditure model, using a different set of independent variables, shows that an increase in span size of 100% is associated with a decrease in expenditure per student of $290.5 per student per year. This difference in results highlights again the importance of model specification in regression analysis.

Kirjavainen and Loikkanen (1998) use data for 1988-91 to investigate school level factors that explain the inefficiency differences between 291 senior secondary schools in Finland. The authors report that, once other potentially explanatory factors are controlled for, no statistically significant relationship between school size and a measure of inefficiency emerges, suggesting that school size does not help to explain variations in inefficiency between senior secondary schools.

Stiefel *et al.* (2000) use 1995 data from 121 high schools in a single New York district to examine costs per student and costs per graduate whilst controlling for a range of school (e.g. size) and student level (e.g. poverty, mathematics achievement) variables. The authors report a statistically significant negative relationship between school size and budget per student and budget per graduate, indicating that as school size increases costs per graduate and costs per student both decrease. Our interpretation of the main results is that an increase in school size of 10% is associated with a decrease in budget per student of 0.96% per year and a decrease in budget per graduate of 1.4% per year.

The authors also report statistically significant interactions between school size and type of school. The relation between school size and cost is analysed for different school types, 'normal' academic high schools, vocational high schools, and transfer schools. 'Transfer' schools are schools of last resort enrolling students that may have dropped out or been suspended from other schools. The results are difficult to interpret as the authors found different results, depending on the regression model used. The results appear to suggest that amongst 'normal' academic high schools costs per graduate are similar in smaller and larger schools. According to the authors, this is due to the much lower dropout rate in smaller schools. For transfer and vocational schools, the budget per student and per graduate was twice as high in 'small' schools compared with medium-sized or large schools. The implication here is that the very large difference in cost per student in this group of schools was distorting the relationship between average secondary school size and cost.

Low WoE

Bowles and Bosworth (2002) use 1994-98 data from 17 Wyoming school districts, representing approximately 80 schools, to examine average costs per student across a four-year time span. The study uses four different regression models to estimate the average cost per student and consistently reports a significant negative relationship between school size and costs, suggesting that as schools become larger costs per students decrease. The authors state that 'an increase of 10 percent in school size decreases cost per student by approximately 2 percent' (p 299).

Table 4.22: Direction of 'effect' and economic outcomes[1]

Study	Direction of 'effect'		
	+	-	∩
High/Medium WoE			
Taylor and Bradley, 2000		-***	
Medium WoE			
Bickel *et al*., 2001		-***	
Kirjavainen and Loikkanen, 1998			∩[2]
Stiefel *et al.,* 2000		-[3]	
Low WoE			
Bowles and Bosworth, 2002		-[4]	

[1] *For economic outcomes negative direction of 'effect' for 'costs' implies average cost per student falls as size increases.*
[2] *Dependent variable is 'inefficiency' and thus is an exception to 1 above.*
[3] *Budget per graduate p<0.01, budget per student p<0.001*
[4] *Reported as statistically significant but no p-values presented*

Table 4.22 shows the directions of effects reported in the different studies for the main dependent variable. With the exception of the Kirjavainen study, the dependent variable in each study was some kind of 'input' measure (i.e. 'cost per student'). These studies show a consistent direction of effect across weights of evidence. These findings suggest that, as school size increases, the costs per student (defined as direct public expenditure on schools) decreases. The only study in which a quadratic term was included (Taylor and Bradley, 2000) found that this relationship was not quadratic but suggest that there are no further reductions in cost per student when school size expands beyond 1,600 students. However, the study by Stiefel and colleagues (Stiefel, 2000) found an interaction effect with school type: that is, the direction of effect was different for different types of school. This study also found that when a combined input and output measure was used as the dependent variable difference in budget per graduate

between 'small' (<600) and 'large' (2,000+) academic schools (i.e. excluding alternative and vocational schools) was very small, due to the vastly lower dropout rate in the smaller schools. These results would caution against being over-confident in concluding that negative directions of effect apply across all types of schools and all types of costs.

Summary

Economic outcomes

- The studies in this category show a consistent negative relationship between average secondary school size and costs defined as direct public expenditure on schools per students. Larger school are associated with lower costs per student, regardless of whether this is measured as expenditure per student or per graduate, or as teaching hours and support hours per student.
- However, one study (Stiefel *et al.*, 2000), despite finding the same negative relationship with average secondary school size, found a statistically significant interaction with school type, suggesting that costs per student differ markedly between different types of school. The sample of schools in this study was, however, comparatively small.
- The size of the relationship between average secondary school size and costs differs slightly between studies. An increase in school size of 10% is estimated to reduce costs per student by between 1% and 4%, depending on the definition of cost used
- Studies of economic outcomes have considered only a limited range of input measures (direct public expenditure on schools per student). Only two studies used a combined input and output measure ('inefficiency', expenditure per graduate), which provide only a limited measure of 'benefit'.

4.6 In-depth review: quality-assurance results

Due to the limited timescale for this review, only a small number of articles (N=4) were data-extracted independently by at least two of the review group and then compared. Other articles (N=27) were subjected to confirmatory data-extraction, where a second member of the review group checked the data-extraction of the first. There were very few differences between reviewers on the core questions on the data-extraction guidelines, other than discrepancies about the extent of the data from the studies needed to be recorded. There were greater differences about the interpretation of results and answers to the review-specific questions. Discussions between all the review groups members and with a second statistician (Harvey Goldstein) resolved these discrepancies. WoE judgements for all the studies were assessed by the wider review group.

4.7 User-involvement in the review process

The wider review group met initially in August 2003 to define the parameters of the review, including the inclusion and exclusion criteria, and sources for searching. An additional meeting took place in December 2003 prior to completion of the preliminary report to the DfES to discuss the emergent findings. Drafts of the reports were also sent out to the wider group for comments and clarifications.

The advisory group, including members from the DfES and the Treasury, met once following completion of the systematic map, in order to discuss and define the

focus of the in-depth review. Additional contact was carried out through email with the advisory review group members who commented on the preliminary and final reports sent to the DfES in December 2003 and January 2004.

5. FINDINGS AND IMPLICATIONS

This chapter outlines the principal findings of this review relating to the 'effects' of secondary school size on student, teacher and school level outcomes. It briefly outlines the results of the searches and systematic map, before presenting the main findings. It then goes on to consider the strengths and limitations of this review, before describing implications for policy, research and practice.

5.1 Summary of principal findings

5.1.1 Identification of studies

This report is based on evidence available to the review group within the restricted time frame of this project. The main method of searching was bibliographic databases, with little use made of personal contacts, websites, handsearching of journals and citations from reference lists. Nearly 4,000 reports were identified as potentially relevant to the review. These were screened against a series of inclusion criteria, to identify studies containing empirical data and outcomes that investigated school size or schools-within-schools, included a variable for school size, were written in English, came from an OECD country, and were published post-1980.

5.1.2 Mapping of all included studies

After application of rigorous criteria, 3,503 reports were excluded and 252 were not available in time for the review. The remaining 134 reports of 119 different studies were then keyworded using both generic and review specific keywords to create a 'map' of the research literature. The final map includes only nine studies from the UK and shows a scarcity of relevant studies looking at the range of outcomes of interest for the review. In particular, there are too few for a separate analysis of the schools-within-schools literature.

5.1.3 Nature of the studies selected for in-depth review

In discussions with the commissioners of the review, a further set of criteria was applied to the studies in the map, and 91 were excluded. For the present report, 31 studies were considered in detail and were subjected to generic and review-specific data extraction. The included studies were all published post-1990, investigated the 'effect' of school size whilst controlling for the 'effects' of SES and considered 'priority' outcomes, such as student attainment or behaviour, teacher morale or experience, and school organisation and costs.

Two-thirds of the 31 studies were from the USA and one fifth from England. In terms of quality, nine were judged to give high/medium WoE to answering the review questions, and five were judged to be low or low/medium. The majority of studies examined the 'effects' of school size on achievement without controlling for prior attainment (N=15); four studies examined achievement whilst controlling for prior attainment; 13 studies examined student attitudes and behaviour; five examined economic outcomes; two examined school organisation outcomes and two examined the perceptions of teachers.

5.1.4 Synthesis of findings from studies in in-depth review

The studies in the in-depth review were synthesized, based on the category of outcome investigated (e.g. student attainment, student behaviour, teacher perceptions, organisation of schools and costs per student) and the weight of evidence judgments (i.e. high, medium, low).

Relationship between school size and student achievement

Taken at face value, the results of studies of the association between school size and attainment appear to suggest that there is no consistent relationship. If, however, the quality of the study is taken into account (i.e. through controlling for prior attainment and the studies' weight of evidence rating), a more consistent picture begins to emerge. Studies that control for prior attainment suggest a quadratic relationship between school size and achievement where attainment is maximised in schools of between 600 and 2,000 students, and studies with a higher weight of evidence tend to show a positive direction of 'effect'.

However, this conclusion would appear to be undermined by the fact that many of the studies that did not control for prior attainment, but had a higher weight of evidence found both positive and negative directions of 'effect'. Closer examination of these studies found that, in at least some cases, positive directions of 'effect' were found for outcomes in children past the minimum school leaving age, whilst the negative direction of 'effects' were in younger age groups in the secondary range. Given the well-established relationship between SES and attainment, it could be that these differences in direction of 'effect' are linked to differences in SES. This argument would seem to be supported by the findings of the USA studies which investigated interaction effects and which suggest that children from lower socio-economic backgrounds do better in smaller schools.

However, this argument would seem to be contradicted by the findings from the only English study which analysed interaction effects and found that the performance of children from lower socio-economic backgrounds was maximised in larger schools. It is also the case that three English studies which found quadratic 'effects' were measuring attainment at GCSE level (i.e. at the point of leaving compulsory education). This suggests that perhaps the relationship between school size and attainment is different in the USA and the UK. This is plausible, given (a) the very much wider range of secondary school size found in the USA, and (b) the very different levels of socio-economic inequality found in the two countries.

Such conclusions are highly speculative at this stage, however, as it is clear that the relationship between school size and attainment is very sensitive to changes in the specification of the regression models. It is also not clear what the size of such 'effects' might be, although it is likely to be much smaller than the 'effect' of differences in SES. It is also the case that formal exams represent only a fairly narrow picture of the learning outcomes that education aims to develop. So any conclusions that increasing school size may have beneficial 'effects' on attainment should be interpreted very cautiously indeed.

Student behaviour and attitudes

Overall, it is difficult to draw firm conclusions about the relationship between school size and student behaviour and attitudes. Studies in the review measured student behaviour and attitude in a variety of ways which makes direct comparisons difficult. It would appear that the relationship between school size,

student attitudes and/or student behaviour is not as straightforward as advocates of particular school sizes (whether large or small) seem to propose. If school size does have an 'effect' on attitudes and behaviour, it seems likely that this is mediated in some way by other school and community related factors.

There is a consistent relationship between student engagement and participation in school and school size; student engagement and participation was greater in smaller schools. This is seen across all studies in this category, regardless of weight of evidence and the methods used to measure the outcome variable. Given these findings, it would seem reasonable to expect that the relationship between school size and absence would show a similar direction of 'effect': that is, more engaged students are less likely to be absent. However, the results of studies with a higher weight of evidence investigating various kinds of absence, indicated that absence from school decreased as schools got larger up to a particular size (which varied between studies), after which absence increased.

The relationship between school size and student violence and disorder is unclear. This may partly be because the studies in this section conceptualised violence in different ways. Whilst higher weight of evidence studies suggest a negative relationship between school size and 'violence' when measured as bullying and school disorder (e.g. as school size increases, disorder and bullying decrease), both studies that label their outcome as 'violence' suggest a positive relationship with greater levels of violence in larger schools. Other lower weight studies that do not label the outcome as violence (e.g. suspensions and student perceived disorder) also appear to support the negative relationship between school size and disorder.

As with attainment, the explanations for the inconsistency in the patterns for absence and violence seem to relate to weight of evidence, age of the students from which the measure was taken, and the specification and nature of the regression model and the variables included. One possible explanation for the inconsistency in the results between 'attitudes' and 'absence', and 'bullying' and other types of 'violent behaviour' may be the different kinds of design used in the studies. The studies of attitudes and of bullying used survey methods and were thus at risk of response biases. For example, it is possible that the characteristics of responders were different in smaller and larger schools. In the studies of 'absence' and other types of violence, the data came from official police reports and school returns. These may be prone to other types of response bias, such as under-reporting, but is seems less likely that this would be systematically linked to school size.

Another possible explanation for the different directions of 'effect' found in studies measuring violent behaviour is that the age groups involved were different between studies. Studies showing a negative relationship included students in grades 6-8, whilst those showing a positive relationship included students grades 9-12. Another possible explanation may relate to the way in which the variables were defined and set up. School disorder is measured as a variable that includes violence, absence and suspensions. It may be that the relationship overall is negative, but that, within this, there may be a range of 'effects' which are obscured by the use of such a wide ranging variable.

Teacher perceptions of school climate and organisation

Comparatively few studies in this category measured the impact of school size on teacher perceptions. Those studies that did showed a consistent negative relationship: teachers in smaller schools perceived their environments more

favourably than teachers in larger schools. This held across all outcome measures, including perceptions of school climate, ability to influence school policies, co-operation amongst staff and students, and resource adequacy. Explanations for this are speculative, but it may be that teachers in smaller schools perceive that they are more involved in all aspects of school life and and/or receive greater support, whether or not their student attitudes and behaviour are different from that of students in larger schools.

School organisation and structure

Only two studies in the in-depth review include measures of the impact of school size on elements of school structure and organization; these studies had very different aims and scope.

One study measured the construct of 'class size', including measures of class size, staff-student ratio and teacher perceptions of satisfaction with class size. This study shows a positive association between school size and 'class size': as schools get larger, so does the construct of class size. However, because this is a construct measure, it is difficult to explain this association and the implications of the association are difficult to interpret.

The second study shows inconsistent relationships between school size and the range of subjects offered and entry levels into higher tiers. Although in some instances students in smaller schools were more likely to be entered into higher tiers, this was not consistent across both science and mathematics. Similarly, the results show that students in smaller schools and larger schools were more likely or less likely to be entered into some subjects, but this was not consistent across subjects or subject types. This result would seem to raise questions about the assumption that larger schools always offer students a broader range of subjects.

Economic outcomes

Studies measuring economic inputs i.e. cost per student show a consistent negative direction of effect: that is, on average, as the size of school increases, cost per student reduces. This association remains consistent whether the outcome is measured either as expenditure per student, cost per graduate, or using proxy measures, such as teaching hours or support hours. The size of this effect varies between studies with estimates of the effect of a 10% increase in school size associated with a reduction in cost per student of between 1% and 4%. One study found that the relationship was linear (i.e. a downwards slope remaining constant) and not quadratic. However, one study also found that there were interaction effects with school type and when a combined input/output measure was used (budget per graduate), the difference in 'cost' between 'small' and 'large' academic schools (i.e. excluding schools that were alternative or vocational) was very small due to the vastly lower dropout rate in 'small' schools. This was a small study based in one USA state, but it does suggest focussing solely on average input costs may give a misleading impression.

It is also important to note that all these studies considered only one particular type of cost: that is, the money provided to schools for their students. Other forms of costs –such as costs to parents, students and other social costs (such as the costs associated with premature dropout) – were not considered. The only study which considered both costs and outputs used graduation rates as a measure of output. This is only a limited concept of educational output as the value of the education received by those that did not graduate was not taken into account.

It is difficult therefore to draw firm conclusions about the likely relationship between school size and economic outcomes. Further studies are needed which adopt more complex models of analysis including a wider variety of 'costs' and benefits'.

What is the relationship between secondary school size and outcomes?

The review question is concerned with the overall relationship between secondary school size and outcomes. The results of individual studies and the synthesis of these results have up to this point been reported within the mutually exclusive categories derived from the reporting of studies included in the review. This section attempts to explore what overall conclusions can be drawn about the relationship. The conclusions reached about the main directions of the relationships in each category are summarised in Figure 5.1. It should be emphasised that this figure is for illustrative purposes only and is not based on literal interpretations of quantitative coefficients reported. The caveats within each category have been discussed in the sections above and this overview should be read with these caveats in mind. Figure 5.1 suggests that, at a global macro-level, there is no simple consistent relationship between secondary school size and outcomes.

However, within particular categories of outcome, the results of higher weight of evidence studies do suggest patterns of directions of effect that may be worthy of further investigation. Exam attainment appears to be maximised and absence minimised at a certain point in the range of secondary school size. The results also suggest that direct public expenditure on schools per student declines as schools get larger. Results also suggest that teacher and student perceptions of schools' climate decline and some kinds of violent behaviour may increase, although, conversely, others may decrease.

These findings would seem inconsistent with a simple linear model of the impact of school size: such as school size 'determines' school climate which, in turn, influences student attainment. However, the relationship between inputs, perceptions and attitudes, and between perceptions, attitudes and attainment is far from straightforward. Negative perceptions of school climate may impact on other student outcomes not measured in the studies in this review. The negative perceptions of school climate found in larger schools may be a result of higher levels of certain types of student violence in these schools. The relationship between school size and perceptions of climate may be indirect in this fashion and/or may also be reciprocal rather than unidirectional. The design of the studies included in this review cannot definitively establish causal relationships and thus the direction of causal relationships is a problem for all the outcomes reported. Does the number of students determine cost or does cost determine the number of students? Does school size determine attainment or does attainment determine school size?

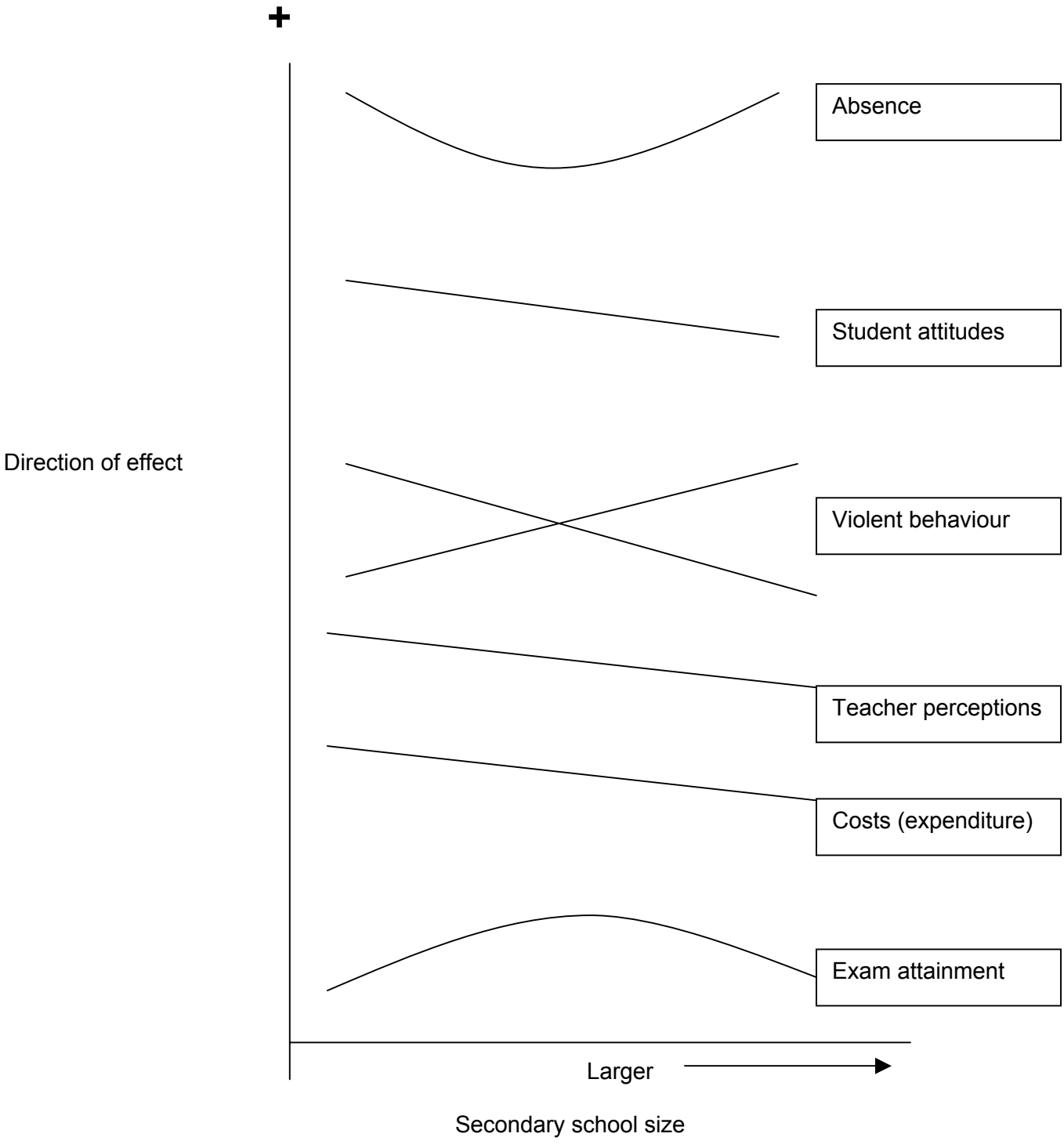
There are three key issues which remain less clear than the directions of 'effects' results suggest. Firstly, even if the interpretation given above is accepted, to be of practical use we would need to know at what size attainment was maximised and/or absence minimised. The studies here do not provide a clear answer to this as the range of school sizes within which attainment is maximised is wide in relation to the actual size of secondary schools in England. Secondly, does the 'average' direction of 'effect' apply to all school/student types? There is some suggestion from the studies in the review that it may not, although there may be differences between the USA and England on this point.

Thirdly, we would want to know the 'effect' of planned or unplanned changes in the size of an individual school. Such an analysis would need to include not only the 'effects' on the school that had changed size, but also 'effects' in neighbouring schools. Only one study in the review addressed this question (Bradley and Taylor, 2003). This study found that average GCSE attainment increased in all schools between 1993 and 2002, but that there was a positive association between increased enrolment and GCSE attainment. However, as the authors point out, it is not clear whether higher GCSE performance led to increased enrolment or higher enrolment led to increased GCSE performance.

This review would seem to refute some of the more prevalent myths about the advantages and disadvantages of smaller and larger schools. For example, that student achievement is universally higher in smaller schools and that student behaviour is universally worse in larger schools have been shown to be inconsistent with the current evidence. This review suggests that the relationship between school size and outcomes is more complex, less clear and less convincing than many of the arguments made in previous 'reviews', referred to in the Background would suggest (see, for example, Raywid, 1999; Gregory, 2001). This review does not, for example, 'overwhelmingly affirm the superiority of small schools' (Cotton, 1996b), nor does it present any 'devastating effects of larger schools' (Raywid, 1997). The differences in the conclusions between this systematic review, previous 'reviews' and the recently published National Research Council and Institute of Medicine report on High School Reform (National Research Council 2004) is likely to be in part at least because the other 'reviews' do not appear to have used systematic approaches to the identification and selection of studies for inclusion. The findings and conclusions from this review appear to support the conclusions drawn by McGuire that 'some research finds positive relationships, some finds negative associations' (1989, p 171). Our conclusions would also reaffirm the caveat drawn by McGuire that 'a good deal of the work to date cannot find relationships of much strength in either direction' (1989, p 171).

Recommendations that secondary school size should be deliberately made smaller, as, for example, in the National Research Council and Institute of Medicine Report on High School Reform (National Research Council, 2004) are not supported by the findings of this review. Whilst the complex patterns seen in the review do not necessarily support the continued expansion of schools, neither do they support school downsizing. 'Reviews' considering school reform practices have commented that school downsizing does not in itself necessarily lead to better outcomes (Irmsher, 1997; Gregory, 2001). It seems likely that these findings reflect in part findings found in this review; that the relationship between school size and outcomes is not as straightforward as has been presented in the literature and that most current research measures the relationships between schools of different sizes, rather than the impact of a single school changing its size.

Figure 5.1: Summary of directions of main effect for different outcome* categories
(Note: This figure illustrates direction not magnitude of 'effects'.)



* The directions of effect for exam attainment and absence are based on higher weight of evidence studies only.

5.2 Strengths and limitations of this systematic review

5.2.1 Strengths

The main strength of the review lies in its systematic and comprehensive nature. The process of systematically identifying, screening and critically appraising the studies helps to ensure that the review process is transparent, as well as replicable and updateable. Another strength is the presentation of the review results in terms of directions of effect, which facilitates direct comparison across studies with similar outcome measures for perhaps the first time in this topic area. Another important strength is the involvement of the commissioners of the review, especially at the point of moving from the map to the in-depth review. This helped to make the review more policy-relevant.

5.2.2 Limitations

The remit of this review extended only to a consideration of studies that investigated empirically the association between an outcome variable and school size. These were all quantitative studies. This meant that qualitative studies that investigated in more depth the *processes* whereby school size might be related to differing socio-cultural and organisational climates, or staff, student and community relationships were not included. This is a limitation imposed by the agreed focus of the review question, rather than the review process itself, but means that little contribution is made to discovering why school size might affect outcomes.

The review process itself had a number of limitations. The truncated form of searching that was carried out because of the restricted timescale for conducting the review (with the cut-off date for retrieval of reports), may have resulted in missing some relevant studies, although it is difficult to estimate the extent of this problem. A simple preliminary analysis of the sensitivity of the bibliographic database search suggests that the impact may have differed between subgroups of the relevant literature. For instance, the citation list from two of the more recently published papers included in the review (Bradley and Taylor, 2003, Bickel *et al.,* 2001) were compared with each other and with the list of citations generated by the review search strategy. Of papers published after 1979, in the Bradley and Taylor paper, 11 out of 95 citations used the words school size and/or organisation compared with 32 from 77 in the paper by Bickel *et al..* Of these, only four were cited in both papers. Eight out of the 11 papers (72%) in the Bradley and Taylor (2003) paper were identified in the bibliographic search carried out for the review and 20 from 32 (62%) in the Bickel *et al.* (2001) paper. It is likely that, where school size was a central feature and thus used in the title of the publication, the searching was more sensitive than for studies where school size was being considered as one variable in a broader conceptual framework (such as economic studies of competition or school choice, for example).

Since the application of inclusion criteria, keywording and data extraction were carried out by two reviewers independently in only a sample of cases, the possibility of reviewer error was greater than if all these procedures had been carried out independently for all studies. However, the information extracted from the papers was continually being re-examined by different members of the review group during the process of analysis and synthesis, thereby minimising the risk of error and improving the quality of the data.

The design of studies included in the in-depth review allow for sophisticated analysis of the degree of association between the dependent variable and school size. However, the studies did not assign students or teachers at random to different size schools (i.e. they did not use an experimental design) and therefore may be affected by selection bias. For example, there is an absence of studies that differentiate between schools that are small for reasons of limited capacity and those that are small for reasons of unpopularity. Therefore, conclusions about causality must be considered tentative.

Studies use different analytical models, different methods of analysis and different methods of constructing both the dependent and independent variables. Making comparisons across studies is therefore difficult, even when they measure the same dependent variable, as the regression coefficients are relative to other variables in the model and are affected by the methods of analysis (e.g. hierarchical or single level; school or student). For these reasons, the review has not standardized the measures of effect presented in the studies, or quantitatively synthesised the results to identify an overall measure of 'effect'. Because of this, the review findings are not able to clearly quantify the size of 'effect' or make any claims as to the statistical significance of such an 'effect'.

Most of the studies identified for inclusion in this review were from taken from USA state data. Within the USA, there is much wider variation in the size of school and in differences in the socio-economic and cultural contexts of schooling. Taken together, these differences may limit the generalisability of conclusions to the UK context. The meaning and use of statistical significance is also difficult to interpret in this review because many of the study findings included all schools in a population as their 'sample'.

Another important limitation of the findings is that the individual studies in the review only measured a limited range of outcomes. Attainment, cost and benefit in particular were conceptualised and measured in a limited way. Examinations only measure one aspect of achievement and direct public expenditure on schools is only one aspect of cost. School size may impact on other measures of cost and attainment in ways that are different from the results presented here.

Finally, partly because of the truncated timescale for conducting the review and the fact that the review commissioners were the DfES and Treasury, we did not involve parents or other users in the review. It is possible that this may have led to different or additional emphases in reporting the syntheses and their implications.

5.3 Implications

5.3.1 Policy

This review does not provide evidence to supports policy initiatives that solely aim either to increase or to decrease the size of schools and/or to close or change the structure of schools below or above a certain size. The review has shown that, for some constructs, there may be advantages of smaller schools. However for other areas there appear to be advantages of larger schools while in others the advantage appears to be for medium-sized schools. There are also a number of qualifications that need to be taken into consideration when considering the practical application of these results. Firstly, the studies in this review do not show the 'effects' on a school that changes its number of students over time. The 'effects' of changing school size within a single school and the upheaval associated with this are not identified or investigated in this review. Secondly, of

the associations reported in the studies, although many were statistically significant, they were often comparatively weak and other factors were shown to be as important or more important in predicting the outcome variable. In particular, in some studies, whilst there were associations between school size and attainment, these were considerably smaller than the associations between socio-economic status and attainment. Thirdly, the types of study design employed in the studies considered in this review cannot by themselves establish 'cause' and 'effect' relationships.

The issue of parental preference is important and becoming increasingly so due to market-based reforms and decisions to allow 'popular' schools to expand to take into account parental choice. There was no research in this review that studied parental preference. One study carried out in the USA, which did not meet the inclusion criteria for this review, found that parents associate more benefits with smaller schools than with larger schools (Johnson, 2002). A Norwegian study, published after the review cut-off date, found that there is a tendency for the parents of students in smaller schools to report lower levels of disillusionment (Westergard and Galloway, 2004).

The pattern of results seen in this review would also seem to suggest that either indefinite expansion or reduction in the numbers of students in any school has the potential to alter the educational environment, including student outcomes. This, in turn, would suggest parents, teachers and students should be made aware of the these possible consequences should policy initiatives be adopted that may lead to dramatic changes in the size of a school.

5.3.2 Research

Overall, we would strongly recommend that resources are devoted to further quantitative work on the relationship between school size and a broad range of educational outcomes. Furthermore, this should be supplemented by rigorous qualitative analysis to unpack some of the more subtle relationships that quantitative analysis cannot detect. However, it is important that future research builds on existing research, both substantively and methodologically. Simply conducting more small- scale evaluation studies with weak research designs will not add to, or strengthen, the evidence base in ways which will be helpful.

More specific research implications concern the outcome variables included in this review. Very few studies measured outcomes relating to teachers and to school organisation and structure. Whilst evidence of these studies was uncovered within the systematic map, a large number of these studies were excluded when more rigorous inclusion criteria were applied at the in-depth review stage. It is important that research considers these outcomes alongside more 'common' outcomes, such as student attainment and behaviour. Other potentially important student outcomes – such as post-school destination and achievement, and extra-curricular participation – were also under-represented in the systematic map and not present in the in-depth review.

Additional research implications arise from the study design and methodology employed in these studies. The review showed that different model specification could have an impact on the significance and direction of 'effect'. Very few of the studies used a similar set of independent variables in their models. Approximately 84 different independent variables had been used across the studies in this review. Whilst regression models should not be over-specified, it is important that models do reflect important variables therefore, research where authors showed the

'effects' of different model specifications would be of value. Likewise, using the same independent variables with different methodologies (e.g. hierarchical linear modelling versus ordinary least squares regression) might also help to elucidate some of the more complex differences in the analyses.

Research needs to consider the different populations of students within these studies. Some of the studies analysing different student age groups (e.g. grade 7 and grade 10) suggest differences in relationships between school size and outcomes across age groups, even when the same methodology and variables are used. Researchers should be cautious in taking a single age group and then extrapolating the results across all age groups present in the school. For example, the needs of GCSE students may be different from those of a year 7 student and those of an A-level student. Research on school size would benefit from understanding more about the needs of different types of students within secondary schools.

Finally, schools-within-schools were not investigated in this review. Schools-within-schools have the potential to offer the benefits of both small and large schools by maintaining several 'small' schools within the same school site. Whilst schools-within-schools are increasing in popularity in the United States and a small number exist in England, there has been very little rigorous evaluation of these types of school. During the literature search, the review identified very few school-within-school evaluations and even fewer that were published and that had been subjected to academic peer- review. We would recommend that evaluations of existing school-within-school initiatives are carried out and that systematic review methods are used to ascertain the full extent of the school-within-school evidence base.

6. REFERENCES

6.1 Studies included in map and synthesis

Studies in systematic map

119 studies written up in 134 papers; linked studies are shown at the end of each reference; schools-within-schools references are included in these figures but reported separately. Studies in the in-depth review are marked with an asterisk (*) and also provided in a separate list at the end of this section.

*Abbott ML, Joireman J, Stroh HR (2002) The influence of district size, school size and socioeconomic status on student achievement in Washington: a replication study using hierarchical linear modelling. Unpublished Technical Report #3. Washington: Washington School Research Center. Available from: http://www.spu.edu/orgs/research/WSRC%20HLM%20District%20Size%20Final%2010-2-02.pdf

Alexander NA (2002) Race, poverty, and the student curriculum: implications for standards policy. *American Educational Research Journal* **39**: 675-693.

Aptekar L (1983) Mexican-American high school students' perception of school. *Adolescence* **18**: 345-357.

*Atkinson A, Wilson D (2003) Measuring pupil attainment in English secondary schools: a preliminary analysis. Unpublished CMPO Working Paper Series No 03/063. Bristol: CMPO, University of Bristol.

Barker B (1985) Curricular offerings in small and large high schools: how broad is the disparity? *Research in Rural Education* **3**: 35-38.

Barker B (1986) The role of the microcomputer in rural schools. *Rural Educator* **7:** 1-3.

Barnett RR, Glass CJ, Snowdon RI, Stringer KS (2002) Size, performance and effectiveness: cost-constrained measures of best-practice performance and secondary-school size. *Education Economics* **10**: 291-311.

Barrow R, Van Zommeren W, Young C, Holtman P (2001) School counselors' and principals' perceptions of violence: guns, gangs and drugs in rural schools. *Rural Educator* **22**: 1-7.

*Bedard K, Brown Jr WO, Helland E (1999) School size and the distribution of test scores. Unpublished. Claremont College: Working Papers in Economics. Claremont CA: Claremont Colleges. Available from: http://econ.claremontmckenna.edu/papers/1999-11.pdf

Beldon, Russonello and Stewart, Research and Communications (2001) New York City's public high school students review their schools' performance: a report of a survey of 1,001 students. Unpublished. Washington DC: Beldon, Russonello and Stewart, Research and Communication. Available from: http://www.kfny.org/publications/hspreport.pdf

Berk LE, Goebel BL (1987a) High-school size and extracurricular participation: a study of a small college environment. *Environment and Behavior* **19**: 53-76.

Berk LE, Goebel BL (1987b) Patterns of extracurricular participation from high school to college. *American Journal of Education* **95**: 468-485.

*Bickel R (1999a) School size, socioeconomic status and achievement: a Georgia replication of inequity in education. Unpublished. West Virginia: College of Education and Human Services, Marshall University. (Linked to Bickel and Howley, 2000; Howley and Bickel, 1999)

*Bickel R (1999b) School size, socioeconomic status and achievement: a Texas replication of inequity in education with a single-unit school addendum. Unpublished. West Virginia: College of Education and Human Services, Marshall University. (Linked to Bickel *et al.*, 2001; Bickel *et al.,* 2000; Howley and Bickel, 1999)

*Bickel R, Howley C (2000) The influence of scale on school performance: a multi-level extension of the Matthew principle. *Education Policy Analysis Archives* **8**. Available from: http://epaa.asu.edu/epaa/v8n22/ (Linked to Bickel and Howley, 2000; Howley and Bickel, 1999)

*Bickel R, Howley C, Williams T, Glascock C (2000) Will the real 'Texas miracle in education' please stand up? Grade span configuration, achievement, and expenditure per pupil. Unpublished. Vermont: Rural School and Community Trust. (Linked to Bickel *et al.*, 2001; Bickel 1999b; Howley and Bickel, 1999)

*Bickel R, Howley C, Williams T, Glascock C (2001) High school size, achievement equity and cost: robust interaction effects and tentative results. *Education Policy Analysis Archives* **9**. Available from: http://epaa.asu.edu/epaa/v9n40.html (Linked to Bickel 1999b; Bickel *et al.,* 2000; Howley and Bickel, 1999)

Bonesronning H (1996a) School characteristics and student achievement: evidence from combined upper secondary schools in Norway. *Education Economics* **4**: 143-160.

Bonesronning H (1996b) Student body composition and school performance: evidence from Norway. *Education Economics* **4**: 11-31.

Bos KT, Ruijters AM, Visscher AJ (1990) Truancy, drop-out, class repeating and their relation with school characteristics. *Educational Research* **32**: 175-185.

Boswell D, Carr B (1988) Secondary school counselors' perceptions of their duties: a comparison of small, medium and large schools. *Contemporary Education* **59**: 227-229.

*Bowen GL, Bowen NK, Richman JM (2000) School size and middle school students' perceptions of the school environment. *Social Work in Education* **22**: 69-82.

*Bowles TJ, Bosworth R (2002) Scale economies in public education: evidence from school level data. *Journal of Education Finance* **28**: 285-300.

*Bradley S, Taylor J (2003) The economics of secondary schooling. In: Johnes G, Johnes J (eds) *International Handbook of Education Economics.* Cheltenham: Edward Elgar.

*Bradley S, Taylor J (1998) The effect of school size on exam performance in secondary schools. *Oxford Bulletin of Economics and Statistics* **60**: 291-324.

Bryk AS, Thum YM (1989) The effects of high-school organization on dropping out: an exploratory investigation. *American Educational Research Journal* **26**: 353-383.

Bulach CR, Williams R (2002) The impact of setting and size on a school's culture and climate. Presentation at the Eastern Educational Research Association at Sarasota, Florida, March 1. Available from: http://westga.edu/~sclimate/article%20setting%20and%20size.htm

Caldas SJ (1993) Re-examination of input and process factor effects on public school achievement. *Journal of Educational Research* **86**: 206-214.

Cicmanec KM, Johanson G, Howley A (2001) High school mathematics teachers: grading practice and pupil control ideology. Paper presented at the annual meeting of the American Educational Research Association, Seattle WA, April 10-14.

Coladarci T and Cobb CD (1996) Extracurricular participation, school size, and achievement and self-esteem among high school students: a national look. *Journal of Research in Rural Education* **12**: 92-103.

*Driscoll D, Halcoussis D, Svorny S (2003) School district size and student performance. *Economics of Education Review* **22**: 193-201.

Echternacht L (1981) Instructional problems of business teachers as perceived by first-year teachers and experienced teachers. *College Student Journal* **15**: 352-358.

Edge OP, Friedberg SH (1984) Factors affecting achievement in the first course in calculus. *Journal of Experimental Education* **52**: 136-140.

Edington ED, Gardener CE (1984a) The relationship of school size to scores concerning the individual and society from the Montana testing service examination. *Education* **105**: 180-184. (Linked to Edington and Gardener, 1984b)

Edington ED, Gardener CE (1984b) The relationship of school size to scores in the affective domain from the Montana testing service examination. *Education* **105**: 40-45. (Linked to Edington and Gardner, 1984a)

Egelund N, Hansen F (2000) Behavioural disorders in Danish schools: a qualitative survey. *European Journal of Special Educational Needs* **15**: 158-170.

Elsworth GR (1998) School size and diversity in the senior secondary curriculum: a generalisable relationship? *Australian Journal of Education* **42**: 183-203.

Epstein JL (1990) What matters in the middle grades: grade span or practices? *Phi Delta Kappan* **71**: 438-444.

*Fetler M (1997) Staffing up and dropping out: unintended consequences of high demand for teachers. *Education Policy Analysis Archives* **5**. Available from: http://epaa.asu.edu/epaa/v5n16.html

Forbes RH, Fortune JC, Packard AL (1993) North Carolina rural initiative study of secondary schools: funding effects on depth of the curriculum. Paper presented at the annual meeting of the Eastern Educational Research Association, Florida, February 18-20.

Foster CM, Martinez I (1985) The effects of school enrollment size in the middle and junior high school on teacher and student attitude and student self concept. *Research in Rural Education* **3**: 57-60.

Fowler WJ, Walberg HJ (1991) School size, characteristics and outcomes. *Educational Evaluation and Policy Analysis* **13**: 189-202.

Franklin BJ, Crone LJ (1992) School accountability: predictors and indicators of Louisiana school effectiveness. Paper presented at the annual meeting of the Mid-South Educational Association, Knoxville TN, November 11-13.

Franklin BJ, Glascock CH (1996) The relationship between grade configuration and student performance in rural schools. Paper presented at the annual meeting of the National Rural Education Association, San Antonio TX, October 11-14.

*Gill B, Dunn M, Goddard E (2002) *Student Achievement in England: Results in Reading Mathematical and Scientific Literacy among 15 year-olds for OECD PISA 2000 Study.* London: The Stationery Office.

Grabe M (1981) School size and the importance of school activities. *Adolescence* **16**: 21-31.

Green J, Barnes D (1993) Discipline in secondary schools: how administrators deal with student misconduct. Unpublished. Indiana: Teachers College, Ball State University.

Greenreese S, Johnson DJ, Campbell WA (1991) Teacher job-satisfaction and teacher job stress: school size, age and teaching experience. *Education* **112**: 247-252. (Linked to Reese *et al.*, 1988)

Haller EJ (1992) High-school size and student indiscipline: another aspect of the school consolidation issue. *Educational Evaluation and Policy Analysis* **14**: 145-156.

Haller EJ, Spotted Bear A, Griffith J, Moss P (1990) School size and program comprehensiveness: evidence from 'high school and beyond'. *Educational Evaluation and Policy Analysis* **12**: 109-120.

Haller EJ, Monk DH, Tien LT (1993) Small schools and higher-order thinking skills. *Journal of Research in Rural Education* **9**: 66 -73.

*Heck RH (1993) School characteristics, school academic indicators and student outcomes: implications for policies to improve schools. *Journal of Education Policy* **8**: 143-154.

Henkin AB, Wanat CL, Davis JH (1996) Social skills of principals: a profile in context. *Journal of School Leadership* **6**: 399-423.

Holland A, Andre T (1994) The relationship of self-esteem to selected personal and environmental resources of adolescents. *Adolescence* **29**: 345-360.

Hough D, Sills-Briegel T (1997) Student achievement and middle level programs, policies and practices in rural America: the case of community-based versus consolidated organizations. *Journal of Research in Rural Education* **13**: 64-70.

*Howley C (1995) The Matthew principle: a West Virginia replication? *Education Policy Analysis Archives* **3**. Available from: http://epaa.asu.edu/epaa/v3n18.html (Linked to Howley, 1996a; Howley, 1996b)

*Howley CB (1996a) Compounding disadvantage: the effects of school and district size on student achievement in West Virginia. *Journal of Research in Rural Education* **12**: 25-32. (Linked to Howley, 1995; Howley 1996b)

*Howley CB (1996b) Sizing up schooling: a West Virginia analysis and critique. Unpublished Dissertation Synopsis. West Virginia: West Virginia University. (Linked to Howley, 1995; Howley, 1996a)

*Howley CB (1999a) The Matthew project: state report for Montana. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley and Bickel, 1999)

*Howley CB (1999b) The Matthew project: state report for Ohio. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley and Bickel, 1999)

*Howley CB, Bickel R (1999) The Matthew project: national report. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley 1999a, 1999b, Bickel, 1999a, 1999b; Bickel and Howley 2000; Bickel *et al.,* 2000, 2001)

Huang G, Howley C (1993) Mitigating disadvantage: effects of small-scale schooling on student achievement in Alaska. *Journal of Research in Rural Education* **9**: 137-149.

Huber JD (1983) Comparison of disciplinary concerns in small and large schools. *Small School Forum* **4**: 7-9.

Hurd S (1995) Calumet region schools. Unpublished. Indiana: Heartland Centre.

Ingersoll RM, Quinn P, Bobbitt S (1997) *The Status of Teaching as a Profession: 1990-91.* NCES 97-104. Washington DC: Office of Educational Research and Improvement, US Department of Education. Available from: http://nces.ed.gov/pubs/97104.pdf

Jacobs LC, Chase CI (1989) Student participation in and attitudes toward high school activities: findings from a national study. *The High School Journal* **72**: 175-181.

Jewell RW (1989) School and school district size relationships: costs, results, minorities, and private school enrollments. *Education and Urban Society* **21**: 140-153.

*Johnson JD, Howley CB, Howley AA (2002) Size, excellence, and equity: a report on Arkansas schools and districts. Unpublished. Ohio: Educational Studies Department, College of Education, Ohio University. Available from: http://oak.cats.ohiou.edu/~howleyc/ARfin.htm

Kaufman P, Chen X, Choy S, Peter K, Ruddy S, Miller A, Fleury J, Chandler K, Planty M, Rand M (2001) *Indicators of School Crime and Safety.* NCES 2002-113. Washington DC: Office of Educational Research & Improvement, US Department of Education. Available from http://nces.ed.gov/pubs2002/2002113.pdf

Kearney JM (1994) The advantages of small rural schools: final report to the Idaho Rural School Association. Unpublished. Idaho: College of Education, University of Idaho.

*Kirjavainen T (1998) Efficiency of Finnish senior secondary schools: explaining high and low performance with resources, environmental and organisational factors. Paper presented at the European Conference of Educational Research, Lubljana, Slovenia, September 17-20. (Linked to Kirjavainen and Loikkanen, 1998)

*Kirjavainen T, Loikkanen HA (1998) Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. *Economics of Education Review* **17:** 377-394. (Linked to Kirjavainen, 1998)

Kowalski GS, Adams MC, Gundlach JH (1983) Structural determinants of juvenile offenses in school. *Urban Education* **18**: 190-197.

Langbein L, Bess R (2002) Sports in school: source of amity or antipathy? *Social Science Quarterly* **83**: 436-54.

Lee VE, Bryk AS (1989) A multilevel model of the social distribution of high school achievement. *Sociology of Education* **62**: 172-192.

Lee VE, Dedrick RF, Smith JB (1991) The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education* **64**: 190-208.

*Lee VE, Smith JB (1993) Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education* **66**: 164-87. (Linked to Lee and Smith, 1995, 1997; Lee *et al.,* 1997)

*Lee VE, Smith JB (1995) Effects of high school restructuring and size on early gains in achievement and learning. *Sociology of Education* **68**: 241-270. (Linked to Lee and Smith, 1993, 1997; Lee *et al.,* 1997)

*Lee VE, Smith JB (1997) High school size: which works best and for whom? *Educational Evaluation and Policy Analysis* **19**: 205-227. (Linked to Lee and Smith, 1993, 1995, Lee *et al.,* 1997)

*Lee VE, Smith JB, Croninger RG (1997) How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education* **70**: 128-50. (Linked to Lee and Smith, 1993, 1995, 1997)

Lee VE, Smerdon BA, Alfeld-Liro C, Brown SL (2000) Inside large and small high schools: curriculum and social relations. *Educational Evaluation and Policy Analysis* **22**: 147-171.

*Lee VE, Burkam DT (2001) Dropping out of high school: the role of school organization and structure. Paper presented at the conference 'Dropouts in America: How severe is the problem? What do we know about intervention and prevention?' Harvard Graduate School of Education, January 13.

Leithwood K, Jantzi D (1997) Explaining variation in teachers' perceptions of principals' leadership: a replication. *Journal of Educational Administration* **35**: 312-331.

*Leung A, Ferris JS (2002) School size and youth violence. Carleton Working Paper CEP 02-10. Paper presented at the Canadian Economic Association Meetings, Calgary, May 30-June 2. Available from: http://www.carleton.ca/economics/cep/cep02-10.pdf

Lien JL, Humphreys JT (2001) Relationships among selected variables in the South Dakota all-state band auditions. *Journal of Research in Music Education* **49**: 146-155.

Luyten H (1994) School size effects on achievement in secondary-education: evidence from the Netherlands, Sweden and the USA. *School Effectiveness and School Improvement* **5**: 75-99.

*Ma X (2001) Bullying and being bullied: to what extent are bullies also victims? *American Educational Research Journal* **38**: 351-370. (Linked to Ma, 2002)

*Ma X (2002) Bullying in middle school: individual and school characteristics of victims and offenders. *School Effectiveness and School Improvement* **13**: 63-89. (Linked to Ma, 2001)

Marsh HW (1993) The effects of participation in sport during the last two years of high-school. *Sociology of Sport Journal* **10**: 18-43.

*McLaughlin D, Drori G, Ross M (2000) *School-Level Correlates of Academic Achievement: Student Assessment Scores in SASS Public Schools.* NCES 2000-303. Washington DC: Office of Education Research and Improvement, US Department of Education. Available from: http://nces.ed.gov/pubs2000/2000303.pdf

*McMillen B, Zhang G, Cobb C, Williamson G, Kaase C, Williams J, Feifs H (2000) School Size and Its Relationship to Achievement and Behavior. Unpublished. North Carolina: State Board of Education. Available from: http://www.ncpublicschools.org/accountability/evaluation/legislative/size.pdf

*McNeely CA, Nonnemaker JM, Blum RW (2002) Promoting school connectedness: evidence from the national longitudinal study of adolescent health. *Journal of School Health* **72:** 138-146.

Melnick SA, Shibles MR, Gable RK (1987) School district size, student achievement and high school course offerings in Connecticut. *Research in Rural Education* **4**: 119-123.

Mertens SB, Flowers N, Mulhall PF (2001) School size matters in interesting ways. *Middle School Journal* **May**: 51-55. Available from: http://www.cprd.uiuc.edu/schools/MSJ%20article%20May01.pdf

Mok M, Flynn M (1996) School size and academic achievement in the HSC examination: is there a relationship? *Issues in Educational Research* **6**: 57-78. Available from: http://education.curtin.edu.au/iier/iier6/mok.html (Linked to Mok and Flynn, 1997)

Mok M, Flynn M (1997) Does school size affect quality of school life? *Issues in Educational Research* **7**: 69-86. Available from: http://education.curtin.edu.au/iier/iier7/mok.html (Linked to Mok and Flynn, 1996)

Monk D, Rice JK (1997) The distribution of mathematics and science teachers across and within secondary schools. *Educational Policy* **11**: 479-498.

Monk DH (1987) Secondary school size and curriculum comprehensiveness. *Economics of Education Review* **6**: 137-150.

Newmann FM, Rutter RA, Smith MS (1989) Organizational factors that affect school sense of efficacy, community and expectations. *Sociology of Education* **62**: 221-238.

Office of Academic affairs (1999) Oregon early options study. Unpublished. Eugene: Office of Academic affairs, Oregon University System. Available from http://www.ous.edu/aca/earlyoptions.htm

Page R (1990) High school size as a factor in adolescent loneliness. *High School Journal* **73**: 150-153.

Peterson KD, Marks HM, Warren VD (1996) SBDM in restructured schools: organizational conditions, pedagogy and student learning. Final deliverable for the Office of Educational Research and Improvement (OERI). Unpublished. Wisconsin: Centre on Organisation and Restructuring of Schools.

Pittman RB, Haughwout P (1987) Influence of high school size on dropout rate. *Educational Evaluation and Policy Analysis* **9**: 337-343.

Ramirez A (1990) High school size and equality of educational opportunity. *Journal of Rural and Small Schools* **4**: 12-19.

Reese SG, Johnson DJ (1988) School size and teacher job-satisfaction of urban secondary-school physical-education teachers. *Education* **108**: 382-4. (Linked to Greenreese *et al.,* 1991).

Ristau R, Crank F, Rogers H (2000) Wisconsin public high school business teachers' use of the internet as a teaching and learning tool. *Delta Pi Epsilon Journal* **42**: 171-188.

Rollins MM, Denton JJ, Janke DL (1983) Attainment of selected earth science concepts by Texas high school seniors. *Journal of Educational Research* **77**: 81-88.

Schoggen P, Schoggen M (1988) Student voluntary participation and high-school size. *Journal of Educational Research* **81**: 288-293.

Schreiber JB (2002) Institutional and student factors and their influence on advanced mathematics achievement. *Journal of Educational Research* **95**: 274-286.

Simpson WD, Marek EA (1988) Understandings and misconceptions of biology concepts held by students attending small high schools and students attending large high schools. *Journal of Research in Science Teaching* **25**: 361-374.

*Silins H, Mulford B (2000) Towards an optimistic future: schools as learning organisations: effects on teacher leadership and student outcomes. Paper presented at the annual AARE-NZARE Conference, Sydney, December. Available from: http://www.aare.edu.au/00pap/sil00273.htm

Smet M (2001) Determining the optimal size of study fields in Flemish secondary education. *Economics of Education Review* **20**: 443-457.

*Spielhofer T, O'Donnell L, Benton T, Schagen S, Schagen I (2002) *The Impact of School Size and Single-Sex Education on Performance.* Slough: National Foundation for Educational Research.

*Stiefel L, Berne R, Iatarola P, Fruchter N (2000) High school size: effects on budgets and performance in New York City. *Educational Evaluation and Policy Analysis* **22**: 27-39.

Stull WJ, Sanders NM, Stull JC (2000) The effects of school size and leadership on participation in the school-to-work movement. *Journal of Vocational Education Research* **25**: 472-502.

*Taylor J, Bradley S (2000) Resource utilization and economies of size in secondary school. *Bulletin of Economic Research* **52**: 123-150.

Thomas H, Bullock A (1992) School size and local management funding formulae. *Educational Management and Administration* **20**: 30-38.

Tomlinson JG, Mortimore P (1990) *Small, Rural and Effective: A Study of Secondary Schools.* Warwick Papers on Education Policy, No. 1. Coventry: Institute of Education, University of Warwick.

Uerling DF (1986) School district consolidations: selected educational implications. *Rural Educator* **7**: 20-22.

Van Batenburg TA, Lokman IH (1991) Effectiveness in senior secondary vocational education. Paper presented at the ICSEI, Cardiff, January.

Walberg HJ (1992) On local control: is bigger better? Unpublished. Washington DC: National Assessment of Educational Progress (NAEP), US Department of Education.

*Welsh WN, Greene JR, Jenkins PH (1999) School disorder: the influence of individual, institutional and community factors. *Criminology* **37**: 73-115.

*Welsh WN, Stokes R, Greene JR (2000) A macro-level model of school disorder. *Journal of Research in Crime and Delinquency* **37**: 243-283.

Whitehead R, Chillag S, Elliott D (1992) Anabolic-steroid use among adolescents in a rural state. *Journal of Family Practice* **35**: 401-405.

Whitener S, Gruber K, Lynch H, Tingos K, Perona M, Fondelier S (1997) *Characteristics of Stayers, Movers, and Leavers: Results from Teacher Follow up Survey 1994-1995.* NCES 97-450. Washington DC: National Center for Education Statistics, US Department of Education. Available from: http://nces.ed.gov/pubs97/97450.pdf

School within schools

Eight studies reported in nine papers; linked studies shown at the end of each reference; also included in the systematic map.

Ancess J, Ort SW (1999) How the coalition campus schools have re-imagined high school: seven years later. Unpublished. Columbia: National Center for Restructuring Education, Schools and Teaching. Available from: www.tc.columbia.edu/~ncrest/aera_ancess.doc (Linked to Darling-Hammond *et al.*, 2002)

Bruckerhoff C, Bruckerhoff T, Sheehan R (2000) Hartford Public High School: analysis of 1998 & 1999 first quarter ninth grade student records. Unpublished. Connecticut: Curriculum Research and Evaluation.

Darling-Hammond L, Ancess J, Ort SW (2002) Reinventing high school: outcomes of the coalition campus schools project. *American Educational Research Journal* **39**: 639-673. (Linked to Ancess and Ort, 1999)

Eichenstein, R (1994) Project achieve, part I: qualitative findings 1993-94. Unpublished. New York: Office of Educational Research, New York City Board of Education.

Fouts JT (1994) A school within a school: evaluation results of the first year of a restructuring effort. Unpublished. Seattle: School of Education, Seattle Pacific University.

Gordon R (1992) School within a school, grades 7, 8, 9, 10: focus on program evaluation. Unpublished. Iowa: Des Moines Public Schools.

Kemple JJ (2001) Career academies: impacts on students' initial transitions to post-secondary education and employment. Unpublished. New York: Manpower Demonstration Research Corporation.

Trybus MA, Li R (1998) Effects of a partnership academy on school and career success of at-risk high school students. Paper presented at the AERA annual meeting in San Diego, California, April 13-17.

Wasley PA, Fine M, Gladden M, Holland NE, King SP, Mosak E, Powell LC (2000) Small schools: great strides: a study of new small schools in Chicago. Unpublished. New York: Bank Street College of Education.

Studies in in-depth review

31 studies reported in 42 papers: linked studies are shown at the end of each reference.

Abbott ML, Joireman J, Stroh HR (2002) The influence of district size, school size and socioeconomic status on student achievement in Washington: a replication study using hierarchical linear modelling. Unpublished Technical Report 3. Washington: Washington School Research Centre. Available from: http://www.spu.edu/orgs/research/WSRC%20HLM%20District%20Size%20Final%2010-2-02.pdf

Atkinson A, Wilson D (2003) Measuring pupil attainment in English secondary schools: a preliminary analysis. Unpublished CMPO Working Paper Series No 03/063. Bristol: CMPO, University of Bristol.

Bedard K, Brown Jr WO, Helland E (1999) School size and the distribution of test scores. Unpublished. Claremont College: Working Papers in Economics. Claremont CA: Claremont Colleges. Available from: http://econ.claremontmckenna.edu/papers/1999-11.pdf

Bickel R (1999a) School size, socioeconomic status, and achievement: a Georgia replication of inequity in education. Unpublished. West Virginia: College of Education and Human Services, Marshall University. (Linked to Bickel and Howley, 2000; Howley and Bickel, 1999)

Bickel R (1999b) School size, socioeconomic status and achievement: a Texas replication of inequity in education with a single-unit school addendum.

Unpublished. West Virginia: College of Education and Human Services, Marshall University. (Linked to Bickel *et al.*, 2000, 2001; Howley and Bickel, 1999).

Bickel R, Howley C (2000) The influence of scale on school performance: a multi-level extension of the Matthew principle. *Education Policy Analysis Archives* **8**. Available from: http://epaa.asu.edu/epaa/v8n22/ (Linked to Bickel, 1999a; Howley and Bickel, 1999).

Bickel R, Howley C, Williams T, Glascock C (2000) Will the real 'Texas miracle in education' please stand up? Grade span configuration, achievement and expenditure per pupil. Unpublished. Vermont: Rural School and Community Trust. (Linked to Bickel, 1999b; Bickel *et al.*, 2001; Howley and Bickel, 1999)

Bickel R, Howley C, Williams T, Glascock C (2001) High school size, achievement equity and cost: robust interaction effects and tentative results. *Education Policy Analysis Archives* **9**. Available from: http://epaa.asu.edu/epaa/v9n40.html (Linked to Bickel, 1999b; Bickel *et al.,* 2000; Howley and Bickel, 1999)

Bowen GL, Bowen NK, Richman JM (2000) School size and middle school students' perceptions of the school environment. *Social Work in Education* **22**: 69-82.

Bowles TJ, Bosworth R (2002) Scale economies in public education: evidence from school level data. *Journal of Education Finance* **28**: 285-300.

Bradley S, Taylor J (2003) The economics of secondary schooling. In: Johnes G, Johnes J (eds) *International Handbook of Education Economics.* Cheltenham: Edward Elgar.

Bradley S, Taylor J (1998) The effect of school size on exam performance in secondary schools. *Oxford Bulletin of Economics and Statistics* **60**: 291-324.

Driscoll D, Halcoussis D, Svorny S (2003) School district size and student performance. *Economics of Education Review* **22**: 193-201.

Fetler M (1997) Staffing up and dropping out: unintended consequences of high demand for teachers. *Education Policy Analysis Archives* **5**. Available from: http://epaa.asu.edu/epaa/v5n16.html

Gill B, Dunn M, Goddard E (2002) *Student Achievement in England: Results in Reading Mathematical and Scientific Literacy among 15 year-olds for OECD PISA 2000 study.* London: The Stationery Office.

Heck RH (1993) School characteristics, school academic indicators and student outcomes: implications for policies to improve schools. *Journal of Education Policy* **8**: 143-154.

Howley C (1995) The Matthew principle: a West Virginia replication? *Education Policy Analysis Archives* **3** Available from: http://epaa.asu.edu/epaa/v3n18.html (Linked to Howley, 1996a, 1996b)

Howley CB (1996a) Compounding disadvantage: the effects of school and district size on student achievement in West Virginia. *Journal of Research in Rural Education* **12**: 25-32. (Linked to Howley, 1995, 1996b)

Howley CB (1996b) Sizing up schooling: a West Virginia analysis and critique. Unpublished Dissertation Synopsis. West Virginia: West Virginia University. (Linked to Howley, 1995, 1996a)

Howley CB (1999a) The Matthew project: state report for Montana. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley and Bickel, 1999)

Howley CB (1999b) The Matthew project: state report for Ohio. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley and Bickel, 1999)

Howley CB, Bickel R (1999) The Matthew project: national report. Unpublished. Ohio: Ohio University and Appalachia Educational Laboratory. (Linked to Howley 1999a, 1999b, Bickel, 1999a, 1999b; Bickel and Howley, 2000; Bickel *et al.,* 2000, 2001)

Johnson JD, Howley CB, Howley AA (2002) Size, excellence and equity: a report on Arkansas schools and districts. Unpublished. Ohio: Educational Studies Department, College of Education, Ohio University. Available from: http://oak.cats.ohiou.edu/~howleyc/ARfin.htm

Kirjavainen T (1998) Efficiency of Finnish senior secondary schools: explaining high and low performance with resources, environmental and organisational factors. Paper presented at the European Conference of Educational Research, Lubljana, Slovenia, September 17-20. (Linked to Kirjavainen and Loikkanen, 1998)

Kirjavainen T, Loikkanen HA (1998) Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. *Economics of Education Review* **17**: 377-394. (Linked to Kirjavainen, 1998)

Lee VE, Smith JB (1993) Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education* **66**: 164-187. (Linked to Lee and Smith, 1995, 1997; Lee *et al.,* 1997)

Lee VE, Smith JB (1995) Effects of high school restructuring and size on early gains in achievement and learning. *Sociology of Education* **68**: 241-270. (Linked to Lee and Smith, 1993, 1997; Lee *et al.,* 1997)

Lee VE, Smith JB (1997) High school size: which works best and for whom? *Educational Evaluation and Policy Analysis* **19**: 205-27. (Linked to Lee and Smith, 1993, 1995; Lee *et al.,* 1997)

Lee VE, Smith JB, Croninger RG (1997) How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education* **70**: 128-50. (Linked to Lee and Smith, 1993, 1995, 1997)

Lee VE, Burkam DT (2001) Dropping out of high school: the role of school organization and structure. Paper presented at the conference 'Dropouts in America: how severe is the problem? What do we know about intervention and prevention?' Harvard Graduate School of Education, January 13.

Leung A, Ferris JS (2002) School size and youth violence. Carleton Working Paper CEP 02-10. Paper presented at the Canadian Economic Association Meetings, Calgary, May 30 - June 2. Available from: http://www.carleton.ca/economics/cep/cep02-10.pdf

Ma X (2001) Bullying and being bullied: to what extent are bullies also victims? *American Educational Research Journal* **38**: 351-370. (Linked to Ma, 2002)

Ma X (2002) Bullying in middle school: individual and school characteristics of victims and offenders. *School Effectiveness and School Improvement* **13**: 63-89. (Linked to Ma, 2001)

McLaughlin D, Drori G, Ross M (2000) *School-Level Correlates of Academic Achievement: Student Assessment Scores in SASS Public Schools.* NCES 2000-303. Washington DC: Office of Education Research and Improvement, US Department of Education. Available from: http://nces.ed.gov/pubs2000/2000303.pdf

McMillen B, Zhang G, Cobb C, Williamson G, Kaase C, Williams J, Feifs H (2000) School size and its relationship to achievement and behavior. Unpublished. North Carolina: State Board of Education. Available from: http://www.ncpublicschools.org/accountability/evaluation/legislative/size.pdf

McNeely CA, Nonnemaker JM, Blum RW (2002) Promoting school connectedness: evidence from the national longitudinal study of adolescent health. *Journal of School Health* **72**: 138-146.

Silins H, Mulford B (2000) Towards an optimistic future: schools as learning organisations: effects on teacher leadership and student outcomes. Paper presented at the annual AARE-NZARE Conference, Sydney, December 2000. Available from: http://www.aare.edu.au/00pap/sil00273.htm

Spielhofer T, O'Donnell L, Benton T, Schagen S, Schagen I (2002) *The Impact of School Size and Single-Sex Education on Performance.* Slough: National Foundation for Educational Research.

Stiefel L, Berne R, Latarola P, Fruchter N (2000) High school size: effects on budgets and performance in New York City. *Educational Evaluation and Policy Analysis* **22**: 27-39.

Taylor J, Bradley S (2000) Resource utilization and economies of size in secondary school. *Bulletin of Economic Research* **52**: 123-150.

Welsh WN, Greene JR, Jenkins PH (1999) School disorder: the influence of individual, institutional, and community factors. *Criminology* **37**: 73-115.

Welsh WN, Stokes R, Greene JR (2000) A macro-level model of school disorder. *Journal of Research in Crime and Delinquency* **37**: 243-83.

6.2 Other references used in the text of the report

Add Health (2003). Add Health: The national longitudinal study of adolescent health. North Carolina: Carolina population centre, University of North Carolina. Retrieved from: http://www.cpc.unc.edu/projects/addhealth

Aitken M, Longford N (1986) Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society,* Series A, **149**: 1-43.

Barker R, Gump P (1964) *Big School, Small School: High School Size and Student Behavior*. Stanford, California: Stanford University Press.

Bedard K, Brown Jr WO, Helland E (2001) School size and the distribution of test scores. Unpublished. Claremont College: Working Papers in Economics.

Claremont CA: Claremont Colleges. Updated version of 1999 paper obtained from the author April 2004.

Boruch RF, Wortman PM (1979) Implications of education evaluation for education policy. In: Berliner DC (ed.) *Review of Research in Education.* Washington DC. American Educational Research Association, (pp 309-361).

Bradley S, Taylor J (2003) The economics of secondary schooling. In: Johnes G, Johnes J (eds) *International Handbook of Education Economics.* Cheltenham: Edward Elgar.

Bradley S, Taylor J (1998) The effect of school size on exam performance in secondary schools. *Oxford Bulletin of Economics and Statistics* **60**: 291-324.

Conant J (1959) *The American High School Today*. New York: McGraw-Hill.

Cook TD, Campbell DT (1979) *Quasi-experimentation: Design and Analysis Issues in Field Settings.* Chicago: Rand McNally.

Cotton K (1996a) *School Size, School Climate and Student Performance*. North West Regional Education Laboratory. School Improvement Research Series. Available from http://www.nwrel.org/scpd/sirs/10/c020.html

Cotton K (1996b) *Affective and Social Benefits of Small-Scale Schooling.* ERIC Digest (ED401088). Charleston WV: ERIC Clearinghouse on Rural Education and Small schools. Available from: http://www.ericfacility.net/databases/ERIC_Digests/ed401088.html

Cutshall S (2003) Is smaller better? When it comes to schools size does matter. Techniques. *ACTEonline*. Available from: http://www.acteonline.org/members/techniques/mar03_story2.cfm

Darling-Hammond L, Ancess J, Ort SW (2002) Reinventing high school: outcomes of the coalition campus schools project. *American Educational Research Journal* **39**: 639-673.

Department for Education and Skills (2003*) Statistics of Education: Schools in England 2003 Edition*. London: Department for Education and Skills. Available from: http://www.dfes.gov.uk/rsgateway/DB/VOL/v000417/index.shtml

Dewees S (1999) *The School-within-a-School Model.* ERIC Digest (ED438147). Charleston WV: ERIC Clearinghouse on Rural Education and Small Schools. Available from: http://www.ericfacility.net/databases/ERIC_Digests/ed438147.html

Egger M, Schneider M, Davey Smith G. (1998) Meta-analysis: spurious precision? Meta-analysis of observational studies. *British Medical Journal* **316**: 140-144.

EPPI-Centre (2003a) *Core Keywording Strategy: Data Collection for a Register of Educational Research. Version 0.9.7.* London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2003b) *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research. Version 0.9.7*. London: EPPI-Centre, Social Science Research Unit.

Gregory T (2001) *Breaking up Large High Schools: Five Common (and Understandable) Errors of Execution.* ERIC Digest (ED459049). Charleston WV:

ERIC Clearinghouse on Rural Education and Small Schools. Available from: http://www.ericfacility.ney/ericdigests/ed459049.html

Hanushek E (2004) What if there are no best practices? *Scottish Journal of Political Economy* **51**: 156-172.

Hanushek E (1979) Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources* **14**: 351-388.

Irmsher K (1997) *School Size* ERIC Digest (ED414615). Eugene OR: ERIC Clearinghouse on Educational Management. Available from: http://www.ed.gov/databases/ERIC_Digests/ed414615.html

Johnson J (2002) Will parents and teachers get on the bandwagon to reduce school size? *Phi Delta Kappen* **January**: 353-356.

Lawrence BK, Bingler S, Diamond BM, Hill B, Hoffman JL, Howley CB, Mitchell S, Rudolph D, Washor E (2002) Dollars and sense: cost effectiveness of small schools. Unpublished. Ohio: Knowledge Works Foundation. Available from: http://www.ruraledu.org/docs/dollars.pdf

Lee VE, Burkam DT (2003) Dropping out of high school: the role of school organisation and structure. *American Educational Research Journal* **40**: 353-393.

McGuire K (1989) School size: the continuing controversy. *Education and Urban Society* **21**: 164-174.

McKenzie P (1995) Scale economies in Australian secondary schools. *Australian Journal of Education* **39**: 109-125.

Muir (2000) Smaller schools: how much more than a fad? *American Educator* **Winter**: 40-46.

National Centre for Educational Statistics (2002) *Digest of Education Statistics 2002.* Washington DC: Institute of Education Sciences, US Department of Education. Available from: http://nces.ed.gov/programs/digest/d02/tables/dt094.asp

National Centre for Educational Statistics (2003) *Digest of Education Statistics 2003.* Washington DC: Institute of Education Sciences, US Department of Education. Available from: http://nces.ed.gov/programs/digest/d02/ch_2.asp#3

National Research Council and the Institute of Medicine (2004) *Engaging Schools: Fostering High School Students' Motivation to Learn.* Washington DC: Board on Children, Youth and Families, Division of Behavioral and Social Sciences and Education. The National Academic Press. Available from: http://books.nap.edu/books/0309084350/html/R1.html#pagetop

Raywid MA (1996) *Downsizing Schools in Big Cities.* ERIC Digest (ED393958). New York NY: ERIC Clearinghouse on Urban Education. Available from: http://www.ericfacility.net/ericdigests/ed393958.html

Raywid MA (1997) Small schools: a reform that works. *Educational Leadership* **December**: 34-39.

Raywid MA (1999) *Current Literature on Small Schools.* ERIC Digest (ED425049). Charleston WV: ERIC Clearinghouse on Rural Education and Small Schools.

Sherman J, Honegger S, McGivern M, Lernke M (2003) *Comparative Indicators of Education in United States and other G8 Countries: 2002.* Washington DC. National Centre for Educational Statistics. Available from: http://nces.ed.gov/pubs2003/2003026.pdf

Stiefel L, Berne R, Iatarola P, Fruchter N (2000) High school size: effects on budgets and performance in New York City. *Educational Evaluation and Policy Analysis* **22**: 27-39.

Tasker M (2003) Smaller Structures in Secondary Education: A Research Digest. Bristol: Human Scale Education.

Tate R (1982) Experimental design. In: Mitzel H and Rabinowitz W (eds) *Encyclopaedia of Education Research.* London: Collier Macmillan, (pp 553-561).

Taylor J, Bradley S (2000) Resource utilization and economies of size in secondary school. *Bulletin of Economic Research* **52**: 123-150.

Torgerson C, Torgerson D (2001) The need for randomised controlled trials in educational research. *British Journal of Educational Studies* **49**: 316-318.

Westergard E, Galloway D (2004) Parental disillusionment with school: prevalence and relationship with demographic variables, and phase, size and location of school. *Scandinavian Journal of Educational Research* **48**: 189-204.

Appendix 1.1: Advisory Group Membership

Members of the School Size Review Group

Zoe Garrett, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Mark Newman, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Diana Elbourne, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London
Steve Bradley, Department of Economics, University of Lancaster
Jim Taylor, Department of Economics, University of Lancaster
Anne West, Centre for Educational Research, Department of Social Policy, London School of Economics (LSE)
Philip Noden, Centre for Educational Research, Department of Social Policy, LSE
Jennifer Sinclair, Research Associate, Institute of Education, University of London
Lorraine Fincham, Research Associate, Insititute of Education, University of London
Veena Meetoo, Research Associate, Institute of Education, University of London
Tom Korolewicz, Research Associate, Institute of Education, University of London

Members of the School Size Advisory Group

Caroline Macready, Department for Education and Skills
Audrey Brown, Department for Education and Skills
Judy Sebba, Department for Education and Skills (until 2004)
Anthony Zacharzewski, HM Treasury
Will Straw, HM Treasury

Appendix 2.1: Inclusion and exclusion criteria

The following indicate the boundaries of the review and the types of studies to be included.

Populations

The review is interested in those studies that investigate the impact in secondary education as defined by schooling aged 11-18 years. Studies including middle school education and sixth-form college education where the average age of students is between 11-18 years were included in the initial searches, although it is recognised that not all students attending these educational institutions will be between 11 and 18 years.

Interventions

The study may focus on schools or schools-within-schools as defined by the presence of an autonomous principal/headteacher. Units and resource bases with separate teaching staff, but sharing a principal with a main school are considered to be beyond the scope of the review, as are special education establishments.

Outcomes

During the initial stages no outcomes were excluded from the review. Therefore the review includes variables that relate to student outcomes (e.g. attainment, truancy), school level outcomes (e.g. ethos, parental involvement) and economic outcomes (e.g. economies of scale).

Study types

To be included in the review, studies needed to be empirical and to include outcomes. Studies that are polemical, literature reviews or purely descriptive without outcomes were excluded (other than for the schools-within-schools literature). Studies had to investigate school size as a variable, although in the initial searches a comparator was required.

Additional criteria

It was beyond the scope of the review to consider studies that are not written in English and that were carried out outside OECD countries, or that were carried out before 1980.

A summary table of the inclusion and exclusion criteria is presented overleaf.

Appendix Table: 2.1.1 Summary of inclusion and exclusion criteria

Inclusion	**Exclusion**
Schools, schools-within-schools *Schools within schools need to be autonomous with their own principal.*	Units, resource bases, special education establishments *Any sub-component of a school that does not have its own principal/headteacher*
Secondary education (11-18 years) *Studies will be included if the students have an average age between 11 and 18 years. Middle schools in England will be included in the initial screening.*	Education outside 11-18 years *Primary and elementary education will be excluded.*
Inclusion of a variable considering school size	No variable considering school size *Resource allocation, class size, district size are to be excluded unless they also focus on school size.*
English language	Not English language
Date of publication post-1980	Date of publication pre-1980
Studies including empirical data and outcomes *A range of study types including exploration of relationships and evaluations* *descriptive studies, case studies for schools within schools only*	Studies without empirical data or outcomes *Including polemic or opinion pieces, literature reviews (except to identify primary studies), simulation exercises*
OECD countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, The Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States	Non-OECD countries

Appendix 2.2: Search strategy for electronic databases

British Education Index

Database: British Education Index 1976 - June 2003
Searched: 18 September 2003
#1: BEI Subject Headings=("TRANSITIONAL SCHOOLS")
#2: BEI Subject Headings=("DEVELOPING INSTITUTIONS")
#3: BEI Subject Headings=("ONE TEACHER SCHOOLS")
#4: BEI Subject Headings=("INSTITUTIONAL SURVIVAL")
#5: BEI Subject Headings=("QUOTAS")
#6: BEI Subject Headings=("SCHOOL SPACE")
#7: BEI Subject Headings=("SCHOOL EXPANSION")
#8: BEI Subject Headings=("SMALL SCHOOLS")
#9: BEI Subject Headings=("SCHOOL SIZE")
#10: #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9
#11: SCATTERPLEX SCHOOL
#12: MULTIPLEX SCHOOL
#13: CHARTER SCHOOL
#14: SPECIAL CURRICULUM SCHOOL
#15: MAGNET SCHOOL
#16: SCHOOL WITH A SPECIAL
#17: SCHOOL COMMUNITY
#18: SMALL SCALE EDUCATIONAL INSTITUTION
#19: MINISCHOOL
#20: SUBSCHOOL
#21: SMALL SCALE SCHOOL
#22: SCHOOL DOWNSIZE
#23: AUTONOMOUS UNIT
#24: SCHOOL WITHIN A SCHOOL
#25: SCHOOL WITHIN SCHOOL
#26: SCHOOL INTAKE
#27: SPECIALIST SCHOOL
#28: FACILITY EXPANSION
#29: TRANSITIONAL SCHOOLS
#30: DEVELOPING INSTITUTIONS
#31: ONE TEACHER SCHOOLS
#32: NON GRADED INSTRUCTIONAL DEVELOPMENT
#33: HOUSE PLAN
#34: COMPREHENSIVE SCHOOL REFORM
#35: MULTIUNIT SCHOOL
#36: CONSOLIDATED SCHOOL
#37: HIGH SCHOOL SIZE
#38: INSTITUTIONAL SURVIVAL
#39: QUOTAS
#40: SCHOOL SPACE
#41: SCHOOL EXPANSION
#42: SCHOOL RESTRUCTURING
#43: SMALL SCHOOLS
#44: SCHOOL SIZE
#45: #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR 22 OR #23 OR #24 OR 25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44
#46: #10 OR #45

#47: BEI Subject Headings=("CHURCH OF ENGLAND SCHOOLS")
#48: BEI Subject Headings=("MIDDLE SCHOOLS")
#49: BEI Subject Headings=("SECONDARY EDUCATION")
#50: BEI Subject Headings=("SECONDARY SCHOOLS")
#51: #47 OR #48 OR #49 OR #50
#52: CATHOLIC SCHOOL
#53 CHURCH OF ENGLAND SCHOOL
#54 VOLUNTARY CONTROLLED SCHOOL
#55 VOLUNTARY AIDED SCHOOL
#56 PUBLIC SCHOOL
#57 PRIVATE SCHOOL
#58 YEAR 13 OR YEAR 12 OR YEAR 11 OR YEAR 10 OR YEAR 9 OR YEAR 8 OR YEAR 7
#59 SECONDARY MODERN
#60 SIXTH FORM
#61 SIXTH FORM COLLEGE
#62 SIXTH FORM SCHOOL
#63 HIGH SCHOOL EDUCATION
#64 COMPREHENSIVE SCHOOL
#65 COMPREHENSIVE EDUCATION
#66 GRAMMAR SCHOOL
#67 GRADE 12 OR GRADE 11 OR GRADE 10 OR GRADE 9
#68 VOCATIONAL HIGH SCHOOLS
#69 VOCATIONAL HIGH SCHOOL
#70 MIDDLE SCHOOL STUDENTS
#71 MIDDLE SCHOOLS
#72 HIGH SCHOOLS
#73HIGH SCHOOL STUDENTS
#74 SECONDARY EDUCATION
#75 JUNIOR HIGH SCHOOL STUDENTS
#76 JUNIOR HIGH SCHOOLS
#77 SECONDARY SCHOOLS
#78 HIGH SCHOOL FRESHMAN
#79 #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61
OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71 OR #72 OE #73 OR #74 OR #75 OR #76 OR #77 OR #78
#80 #51 OR #79
#81 #46 AND #80

Australian Education Index

Australian Education Index 1976 - June 2003
Searched: 18 September 2003
#1: AEI Subject Headings=("FACILITY EXPANSION")
#2: AEI Subject Headings=("TRANSITIONAL SCHOOLS")
#3: AEI Subject Headings=("DEVELOPING INSTITUTIONS")
#4: AEI Subject Headings=("HOUSE PLANS")
#5: AEI Subject Headings=("INSTITUTIONAL CHARACTERISTICS")
#6: AEI Subject Headings=("SCHOOL ORGANIZATION")
#7: AEI Subject Headings=("MULTIUNIT SCHOOLS")
#8: AEI Subject Headings=("CONSOLIDATED SCHOOLS")
#9: AEI Subject Headings=("INSTITUTIONAL SURVIVAL")
#10: AEI Subject Headings=("QUOTAS")
#11: AEI Subject Headings=("SCHOOL SPACE")
#12: AEI Subject Headings=("SCHOOL EXPANSION")
#13: AEI Subject Headings=("SCHOOL RESTRUCTURING")

#14: AEI Subject Headings=("SMALL SCHOOLS")
#15: AEI Subject Headings=("SCHOOL SIZE")
#16: #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15
#17: SCATTERPLEX SCHO
#18: MULTIPLEX SCHOOL
#19: MAGNET SCHOOL
#20: SCHOOL WITH A SPECIAL
#21: SMALL SCALE EDUCATION INSTITUTION
#22: MINISCHOOL
#23: SUBSCHOOL
#24: SMALL SCALE SCHOOL
#25: SCHOOL DOWNSIZ
#26: AUTONOMOUS UNIT
#27: SCHOOL WITHIN A SCHOOL
#28: SCHOOL WITHIN SCHOOL
#29: SCHOOL INTAKE
#30: SPECIALIST SCHOOL
#31: FACILITY EXPANSION
#32: TRANSITIONAL SCHOOLS
#33: DEVELOPING INSTITUTIONS
#34: ONE TEACHER SCHOOLS
#35: NON GRADED INSTRUCTIONAL DEVELOPMENT
#36: HOUSE PLAN
#37: COMPREHENESIVE SCHOOL REFORM
#38: INSTITUTIONAL CHARACTERISTICS
#39: SCHOOL ORGANIZATION
#40: MULTIUNIT SCHOOLS
#41: CONSOLIDATED SCHOOLS
#42: HIGH SCHOOL SIZE
#43: INSTITUTIONAL SURVIVAL
#44: QUOTAS
#45: SCHOOL SPACE
#46: SCHOOL EXPANSION
#47: SCHOOL RESTRUCTURING
#48: SMALL SCHOOLS
#49: SCHOOL SIZE
#50: #17 OR #18 OR #19 OR #20 OR #21 OR 22 OR #23 OR #24 OR 25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49
#51: #16 OR #50
#52: AEI Subject Headings=("YEAR 7" OR "YEAR 8")
#53: AEI Subject Headings=("YEAR 10" OR "YEAR 11" OR "YEAR 12" OR "YEAR 13") AEI Subject Headings=("YEAR 9")
#54: AEI Subject Headings=("GRADE 10" OR "GRADE 11" OR "GRADE 12")
#55: AEI Subject Headings=("GRADE 9")
#56: AEI Subject Headings=("VOCATIONAL HIGH SCHOOLS")
#57: AEI Subject Headings=("MIDDLE SCHOOLS")
#58: AEI Subject Headings=("HIGH SCHOOL STUDENTS")
#59: AEI Subject Headings=("SECONDARY EDUCATION")
#60: AEI Subject Headings=("JUNIOR HIGH SCHOOLS")
#61: AEI Subject Headings=("SECONDARY SCHOOLS")
#62: AEI Subject Headings=("JUNIOR HIGH SCHOOL" OR "JUNIOR HIGH SCHOOL STUDENT”)
#63: AEI Subject Headings=("HIGH SCHOOLS")

#64: #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62
#65: CATHOLIC SCHOOL
#66: CHURCH OF ENGLAND SCHOOL
#67: VOLUNTARY AIDED SCHOOL
#68: PUBLIC SCHOOL
#69: PRIVATE SCHOOL
#70: YEAR 13 OR YEAR 12 OR YEAR 11 OR YEAR 10 OR YEAR 9 OR YEAR 8 OR YEAR 7
#71: SECONDARY MODERN
#72: SIXTH FORM
#73: HIGH SCHOOL EDUCATION
#74: COMPREHENSIVE SCHOOL
#75: COMPREHENSIVE EDUCATION
#76: COMPREHENSIVE SCHOOL
#77: GRAMMAR SCHOOL
#78: GRADE 12 OR GRADE 11 OR GRADE 10 OR GRADE 9
#79: VOCATIONAL HIGH SCHOOL
#80: MIDDLE SCHOOL STUDENTS
#81: MIDDLE SCHOOLS
#82: HIGH SCHOOLS
#83: HIGH SCHOOL STUDENTS
#84: SECONDARY EDUCATION
#85: JUNIOR HIGH SCHOOL STUDENTS
#86: JUNIOR HIGH SCHOOLS
#87: SECONDARY SCHOOLS
#88: HIGH SCHOOL FRESHMAN
#89: #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71 OR #72 OE #73 OR #74 OR #75 OR #76 OR #77 OR #78 OR #79 OR #80 OR #81 OR #82 OR #83 OR #84 OR #85 OR #86 OR #87 OR 88
#90: 64 OR #89
#91: #51 OR #90

ERIC

ERIC 1980-2003
Searched: 15 September 2003
Cambridge Scientific Abstracts
#1: KW=((year* 10) OR (year* 11) OR (year* 12))
#2: KW=((year* 7) OR (year* 8) OR (year* 9))
#3: KW=((grade* 7) OR (grade* 8) OR (grade* 9))
#4: KW=((grade* 10) OR (grade* 11) OR (grade* 12))
#5: KW=((secondary modern) OR (grammar school*))
#6: KW=((middle school*) OR (six* form*) OR (comprehensive school*))
#7: KW=((secondary educat*) OR (secondary school*) OR (high school*))
#8: #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
#9: ((de=((middle school students)
#10:((de=((secondary education)
#11:((de=((high school freshmen)
#12:((de=((high school seniors)
#13:((de=((high schools)
#14:((de=((high school students)
#15:((de=((junior high school students)
#16:((de=((secondary school teachers)
#17:((de=((middle schools)
#18:((de=((middle school teachers)
#19:((de=((secondary school teachers)

#20:((de=((secondary education)
#21:((de=((secondary schools)
#22:((de=((secondary school students)
#23:((de=((elementary secondary education)
#24:((de=((junior high schools)
#25:((de=((grade 7) or (grade 8) or (grade 9) or (grade 10) or (grade 11) or (grade 12))
#26: #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25
#27: #8 OR #26
#28: KW=((facility expansion))
#29: KW=((develop* institution*))
#30: KW=((transition* school*))
#31: KW=((one teacher school*))
#32: KW=((school expansion))
#33: KW=((house plan))
#34: KW=((consolidated school*))
#35: KW=((multiunit school*))
#36: KW=((school with a special*))
#37: KW=((minischool*))
#38: KW=((subschool*))
#39: KW=((small scale school*))
#40: KW=((school downsiz*))
#41: KW=((autonomous unit))
#42: KW=((school size))
#43: KW=((small school))
#44: #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43
#45: ((de=((multiunit schools)))
#46: ((de=((transitional schools)))
#47: ((de=((facility expansion)))
#48: ((de=((comprehensive school reform)))
#49: ((de=((small schools)))
#50: ((de=((school size)))
#51: ((de=((house plan)))
#52: ((de=((school expansion)))
#53: ((de=((consolidated schools)))
#54: ((de=((one teacher schools)))
#55: ((de=((developing institutions)))
#56: #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55
#57: #44 OR #56
#58: #27 AND #57

ASSIA

Cambridge Scientific Abstracts
Searched: 10 September 2003
Database: ASSIA: Applied Social Sciences Index and Abstracts
#1: KW=((year* 10) OR (year* 11) OR (year* 12))
#2: KW=((year* 7) OR (year* 8) OR (year* 9))
#3: KW=((grade* 7) OR (grade* 8) OR (grade* 9))
#4: KW=((grade* 10) OR (grade* 11) OR (grade* 12))
#5: KW=((secondary modern) OR (grammar school*))
#6: KW=((middle school*) OR (six* form*) OR (comprehensive school*))
#7: KW=((secondary educat*) OR (secondary school*) OR (high school*))
#8: #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7

#9: ((de=((middle school students)
#10:((de=((secondary education)
#11:((de=((high school freshmen)
#12:((de=((high school seniors)
#13:((de=((high schools)
#14:((de=((high school students)
#15:((de=((junior high school students)
#16:((de=((secondary school teachers)
#17:((de=((middle schools)
#18:((de=((middle school teachers)
#19:((de=((secondary school teachers)
#20:((de=((secondary education)
#21:((de=((secondary schools)
#22:((de=((secondary school students)
#23:((de=((elementary secondary education)
#24:((de=((junior high schools)
#25:((de=((grade 7) or (grade 8) or (grade 9) or (grade 10) or (grade 11) or (grade 12))
#26: #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25
#27: #8 OR #26
#28: KW=((facility expansion))
#29: KW=((develop* institution*))
#30: KW=((transition* school*))
#31: KW=((one teacher school*))
#32: KW=((school organisation))
#33: KW=((school organization))
#34: KW=((school expansion))
#35: KW=((school restructuring))
#36: KW=((house plan))
#37: KW=((consolidated school*))
#38: KW=((multiunit school*))
#39: KW=((charter school*))
#40: KW=((magnet school*))
#41: KW=((school with a special*))
#42: KW=((minischool*))
#43: KW=((subschool*))
#44: KW=((small scale school*))
#45: KW=((school downsiz*))
#46: KW=((autonomous unit))
#47: KW=((school size))
#48: KW=((small school))
#49: #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48
#50: de=((multiunit schools))
#51: de=((transitional schools)
#52: de=((facility expansion))
#53: de=((comprehensive school reform))
#54: de=((small schools))
#55: de=((school size))
#56: de=((school restructuring))
#57: de=((house plan))
#58: de=((school expansion))
#59: de=((quotas))
#60: de=((consolidated schools))

#61: de=((one teacher schools))
#62: de=((developing institutions))
#63: de=((school organization))
#64: #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63
#65: #49 OR #64
#66: #27 AND #65

Social Science Citation Index

Searched: 10 September 2003
Databases SCI-Expanded, SSCI, A&HCI
Time span=1981-2003
#1: TS=(school size OR small school OR autonomous unit OR school downsiz*)
#2: TS=(small scale school* OR subschool*)
#3: TS=(minischool* OR school with a special* OR magnet school* OR charter school* OR multiunit school*)
#4: TS=(consolidated school* OR house plan OR school restructur* OR school expansion* OR school organ*)
#5: TS=(one teacher school* OR transition* school* OR develop* institut*)
#6: #1 OR #2 OR #3 OR #4 OR #5
#7: TS=(high school* OR secondary school* OR comprehensive school* OR six* form* OR middle school* OR grammar school*)
#8: TS=(secondary educat*)
#9: TS=(year* 7 OR year* 8 OR year* OR year* 10 OR year* 11 OR year* 12)
#10: TS=(grade* 7 OR grade* 8 OR grade* 9 OR grade* 10 OR grade* 11 OR grade* 12)
#11: #7 OR #8 OR #9 OR #10
#12: #6 AND #11

PsycInfo

PsycINfo 1872-2003/09 wk3
Limits: 1980-2004; English
Searched: 19 September 2003
#1 school size
#2 small schools
#3 school restructuring
#4 school expansion
#5 school space
#6 quotas
#7 institutional survival
#8 high school size
#9 consolidated schools
#10 multiunit schools
#11 comprehensive school reform
#12 house plan
#13 non graded instructional grouping
#14 one teacher schools
#15 developing institutions
#16 transitional schools
#17 facility expansion
#18 school intake
#19 school* within school*
#20 school* within a School*
#21 autonomous unit*
#22 school downsiz*

#23 small scale school*
#24 subschool*
#25 minischool*
#26 small scale education* institution
#27 magnet school*
#28 charter school*
#29 multiplex school*
#30 scatterplex school*
#31 school *enrollmen*t
#32 ("School-*Enrollmen*t" in DE)
#33 ("Nongraded-Schools" in DE)
#34 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33
#35 high school freshman
#36 secondary school*
#37 junior high school*
#38 junior high school* students
#39 secondary education
#40 high school students
#41 high school*
#42 middle school*
#43 middle school students
#44 vocational high school*
#45 grade 9
#46 grade 10
#47 grade 11
#48 grammar school*
#49 comprehensive educat*
#50 comprehensive school*
#51 high school educat*
#52 sixth form school*
#53 sixth form
#54 sixth form colleg*
#55 secondary modern
#56 year 7
#57 year 8
#58 year 9
#59 year 10
#60 year 11
#61 year 12
#62 year 13
#63 private school*
#64 public school*
#65 voluntary aided school*
#66 voluntary controlled school*
#67 church of England School*
#68 catholic school*
#69 (("Junior-High-School-Students" in DE) or ("Junior-High-Schools" in DE))
#70 ("Secondary-Education" in DE)
#71 (("High-School-Students" in DE) or ("High-Schools" in DE))
#72 (("Middle-School-Education" in DE) or ("Middle-School-Students" in DE))
#73 #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53

OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71 OR #72
#74 “Elementary-Education" in DE) or ("Elementary-School-Students" in DE) or ("Elementary-School-Teachers" in DE) or ("Elementary-Schools" in DE)
#75 #73 not #74
#76 #75 AND #34

Appendix 2.3: EPPI-Centre Keyword sheet, including review-specific keywords

A.1 Identification of report (or reports)

A.1.1 Citation
A.1.2 Contact
A.1.3 Handsearch
A.1.4 Unknown
A.1.5 Electronic database

A.2 Status

A.2.1 Published
A.2.2 In press
A.2.3 Unpublished

A.3 Linked reports

A.3.1 Not linked
A.3.2 Linked

A.4 Language (please specify)

A.4.1 Details

A.5 In which country/countries was the study carried out?

A.5.1 Details

A.6 What is/are the topic focus/foci of the study?

A.6.1 Assessment
A.6.2 Classroom management
A.6.3 Curriculum
A.6.4 Equal opportunities
A.6.5 Methodology
A.6.6 Organisation and management
A.6.7 Policy
A.6.8 Teacher careers
A.6.9 Teaching and learning
A.6.10 Other topic focus

A.7 Curriculum

A.7.1 Art
A.7.2 Business Studies
A.7.3 Citizenship
A.7.4 Cross-curricular
A.7.5 Design & Technology
A.7.6 Environment
A.7.7 General
A.7.8 Geography
A.7.9 Hidden
A.7.10 History
A.7.11 ICT
A.7.12 Literacy - first language
A.7.13 Literacy further languages
A.7.14 Literature
A.7.15 Mathematics
A.7.16 Music
A.7.17 PSE
A.7.18 Phys. Ed.
A.7.19 Religious Ed.
A.7.20 Science
A.7.21 Vocational
A.7.22 Other curriculum

A.7.23 The material does not focus on curriculum issues.

A.8 Programme name (Please specify.)

A.8.1 Details

A.9 What is/are the population focus/foci of the study?

A.9.1 Learners
A.9.2 Senior management
A.9.3 Teaching staff
A.9.4 Non-teaching staff
A.9.5 Other education practitioners
A.9.6 Government
A.9.7 Local education authority officers
A.9.8 Parents
A.9.9 Governors
A.9.10 Other population focus

A.10 Age of learners (years)

A.10.1 0-4
A.10.2 5-10
A.10.3 11-16
A.10.4 17-20
A.10.5 21 and over

A.11 Sex of learners

A.11.1 Female only
A.11.2 Male only
A.11.3 Mixed sex

A.12 What is/are the educational setting(s) of the study?

A.12.1 Community centre
A.12.2 Correctional institution
A.12.3 Government department
A.12.4 Higher education institution
A.12.5 Home
A.12.6 Independent school
A.12.7 Local education authority
A.12.8 Nursery school
A.12.9 Post-compulsory education institution
A.12.10 Primary school
A.12.11 Pupil referral unit
A.12.12 Residential school
A.12.13 Secondary school
A.12.14 Special needs school
A.12.15 Workplace
A.12.16 Other educational setting

A.13 Which type(s) of study does this report describe?

A.13.1 Description
A.13.2 Exploration of relationships
A.13.3 Evaluation: naturally occurring
A.13.4 Evaluation: researcher-manipulated
A.13.5 Methodology
A.13.6 Review: systematic review
A.13.7 Review: other review

Review-specific keywords

A.1 Study school focus

A.1.1 School within a school
A.1.2 School size
A.1.3 School consolidation or decline

A.2 Student characteristics

A.2.1 SEN
A.2.2 Gifted or talented
A.2.3 Mainstream
A.2.4 Ethnic minority

A.3 Methods

A.3.1 Categorical
A.3.2 Continuous
A.3.3 Other

A.4 Nature of the variable

A.4.1 Whole school size studied
A.4.2 Proxy measure

A.5 Student outcomes

A.5.1 Performance measured with prior attainment
A.5.2 Performance measured without prior attainment
A.5.3 Post school destination
A.5.4 Student attitudes
A.5.5 Student behaviour
A.5.6 Attendance
A.5.7 Long term economic outcomes
A.5.8 Other
A.5.9 Not studied

A.6 Teacher outcomes

A.6.1 Morale and stress
A.6.2 Retention
A.6.3 Salary
A.6.4 Experience
A.6.5 Other
A.6.6 Not studied

A.7 School outcomes

A.7.1 Class size
A.7.2 Grouping arrangements
A.7.3 School accountability and governance
A.7.4 Student teacher relationships
A.7.5 Communication within the school
A.7.6 Parental involvement
A.7.7 Relationships between school and wider community
A.7.8 Range of curricular provision
A.7.9 Range of extra-curricular provision
A.7.10 Education economics
A.7.11 Other
A.7.12 Not studied

A.8 Other outcomes

A.8.1 Detail

A.9 Country

A.9.1 North America
A.9.2 United Kingdom
A.9.3 Europe (not UK)
A.9.4 Australasia
A.9.5 Other

Appendix 2.4: Framework for assessing study 'validity'

The policy level question addressed by the review is whether systematic intervention to change school size (and/or to stop changes) will lead to additional 'benefit' (and/or prevent 'harm'). The review question is therefore about impact: that is, does school size 'determine' an outcome? In this context, the argument is made that differences in the independent variable (school size) 'cause' systematic differences in the dependent outcome variable(s). An important issue for interpretation and synthesis is to what extent the results obtained in an individual study that attempts to answer this question can be said to be an estimate of the 'true' state rather than an artefact of the study design and/or method of analysis (i.e. bias).

A primary research study that aimed to investigate this question would attempt to minimise the various threats to validity that can produce bias in the study results. Empirical evidence has demonstrated that the optimal research design for questions of impact are prospective randomised experiments. This design is regarded as optimal both because it establishes the direction of causality and minimises the effects of bias (Boruch and Wortman, 1979; Tate, 1982; Torgerson and Torgerson, 2001). Observational studies, such as those included in the review, yield estimates that may deviate from the true underlying relationship beyond the play of chance due to the 'effects' of confounding factors, biases or both. The main problem is not the lack of precision but that studies produce findings that are seriously biased or confounded (Egger *et al.,* 2001). When considering the interpretation and synthesis of the results, it is therefore appropriate to estimate the extent to which each of the individual studies 'controlled' for the various threats to validity.

Table 2.4.1: Threats to validity (after Cook and Campbell, 1979)

Threats to statistical conclusion validity (a) Low statistical power (b) Violated statistical assumptions (c) Error rate (d) Reliability of measures (e) Reliability of treatment (f) Random irrelevancies in setting (g) Random heterogeneity of respondent **Threats to construct validity** (a) Inadequate explication of the constructs (b) Mono-operation bias (c) Mono-method bias (d) Hypothesis guessing (e) Evaluation apprehension (f) Experimenter expectancies (g) Confounding levels of construct (h) Interaction of different treatments (i) Interaction of testing and treatment (j) Restricted generalisability	**Threats to internal validity** (a) History (b) Maturation (c) Testing (d) Instrumentation (e) Statistical regression (f) Selection (g) Mortality (h) Interaction of selection with maturation, history and testing (i) Ambiguity about direction of causality (j) Diffusion/imitation of treatment (j) Compensatory equalisation of treatment (k) Demoralisation of respondents **Threats to external validity** (a) Interaction of selection and treatment (b) Interaction of setting and treatment (c) Interaction of history and treatment

Some of the threats to validity listed in Table 2.4.1 are inherent in real world settings and, as such, the researcher attempts to design them out. The process of study design and conduct creates some of the threats to validity and thus requires other study design elements and/or the analysis to account for the 'additional' threats generated, the review criteria specified a number of requirements that studies should meet which 'control' for some threats to validity and/or component parts of some threats. After selecting studies which met the review

inclusion criteria, it is also apparent that, due to the design of the included studies, other threats to validity are either not relevant and/or are simply not dealt with in any of the studies considered. On this basis, a more limited list of threats was selected against which the studies in the review could be considered and compared. The list of threats not considered further (and the reasons for excluding them) are given in Table 2.4.2. The threats to external validity identified by Cook and Campbell are 'products' of the internal validity. So lack of internal validity creates by definition threats to external validity and so external 'threats' were not considered separately.

Table 2.4.2: Threats not considered further

'Threat'	Reason for not considering further
Low statistical power	Power calculation not reported in any study
Error rate	Given the methods used, this is a potential problem in all studies.
Reliability of treatment Inadequate explication of the constructs Diffusion/imitation of treatment Compensatory equalisation of treatment Restricted generalisability	From the perspective of the analysis used, school size is the treatment and therefore these threats are not an issue. What might be an issue is whether there is sufficient number of schools of different sizes in the sample which is considered.
Demoralisation of respondents Hypothesis guessing Maturation	NA retrospective study design
Testing (and interaction with treatment)	Not pre-/post-test design. Where 'gain' scores used is same for all cases in study.
Instrumentation	NA use national datasets and/or survey instruments in cross sectional designs

Hence we drew up a table adapted from Cook and Campbell but revised for this review (Table 2.4.3). This was completed for each type of outcome in every study. The more 'threats' identified, the lower the WoE the study was assessed to provide in terms of its design and analysis. The number of these threats was used as indicative, rather than deterministic, both because the different threats may not be equally problematic for the validity of studies, but also because other factors (for example, context and ethics) are relevant in judging the WoE of a study for the review.

Table 2.4.3: Evidence that threat to validity 'controlled'

Threats	*Features*	*Indicators that threat to validity met*
Violated statistical assumptions	All assumptions about random selection, distribution of the data, and unit of analysis should be reported.	Normal distribution/Outliers (or corrected for) Independent observations (or corrected for) Random sample (at the school level) (or weighting applied) Confirmation that data are linear (or corrected for) Data are homoscedastic (or corrected for) Specification of the model: Adjusted R^2 F Test for model fit Analysis of residuals
Reliability of measures	Low reliability indicates high standard errors.	Only applicable where survey instruments used Scale reliability coefficients
Random irrelevancies in setting	Environmental effects which may cause or interact with treatment effects.	Random selection of sample Adequate sample size Check for multicollinearity (use instrumental variable estimation technique)
Random heterogeneity of respondents	Certain characteristics in subjects may be correlated with dependent variables.	Random selection of sample Longitudinal data Adequate sample size Check for multicollinearity (as above)
Mono-operation bias	Measure single dependent variable	Multiple dependent variables for the same construct (e.g. attainment)
Mono-method bias	Measure dependent variable in one way	Measure dependent variable in more than one way (e.g. satisfaction measured in different ways)
History	Event external to treatment which may affect dependent variable	Random sampling Longitudinal data
Maturation	Biological and psychological changes in subjects which will affect their responses	Random sampling Longitudinal data
Statistical regression	Extreme scores tend to move to middle on post-testing regardless of treatment.	Use of gain or value-added scores
Selection (interaction effects)	Differences in subjects prior to treatment	Random allocation not feature of studies. All studies control for SES. Other 'key' independent variables 'controlled for' (Y/N): Class size Student/teacher ratio Funding: Public/Private
Mortality/missing data	Differential loss of subjects during study	Sensitivity/Intention to treat analysis/weighting
Ambiguity about direction of causality	In studies conducted at one point in time, problem inferring direction of causality	Random allocation not a feature of studies Longitudinal data Use of path analysis Assessment of simultaneity (Structural Equations Modelling)

Appendix 3.1: Details of studies in systematic map

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Abbott *et al. (*2002)	USA	Exploration of relationships	School size	Proxy measure: span size	Continuous	Student outcomes: (i) Performance measured without prior attainment
Alexander (2002)	USA	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Continuous	Student outcomes: (i) Course taking patterns School organisation and management: (i) Range of curricular provision
Aptekar (1983)	USA	Evaluation: naturally occurring	School size	Proxy measure: number of seniors	Categorical One school with 67 seniors compared with one with 38	Student outcomes: (i) Student attitudes School organisation and management: (i) School accountability and governance, (ii) communication within the school, (iii) range of curricular provision
Atkinson and Wilson (2003)	England	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) performance measured without prior attainment
Barker (1985)	USA	Exploration of relationships	School size	Whole school size studied	Categorical Small<500 Large>1,000	School organisation and management: (i) Range of curricular provision
Barker (1986)	USA	Exploration of relationships	School size	Whole school size studied	Categorical Small = <500 large = >1,000	Student outcomes: (i) Opportunity to use microcomputers Teacher outcomes: (i) Experience of computer use
Barnett *et al. (*2002)	Northern Ireland, UK	Exploration of relationships	School size	Whole school size studied	Categorical 0-299, 300-399,	Student outcomes: (i) Performance measured without prior attainment

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
		Evaluation: naturally occurring			400-499, 500-599, 600-799, 800-999, >1,000	
Barrow *et al. (*2001)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <200 201-500 >500	Student outcomes: (i) Student attitudes, (ii) student behaviour
Bedard *et al. (*1999)	USA	Exploration of relationships	School size	Whole school size studied Proxy measure: average enrolment per grade (span size)	Continuous	Student outcomes: (i) Performance measured without prior attainment
Belden Russonello and Stewart, Research and Communications (2001)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <2,000, 2,000-3,000 >3,000	Student outcomes: (i) Post-school destination, (ii) student attitudes, (iii) student behaviour Teacher outcomes: (i) Teaching quality School organisation and management: (i) Class size, (ii) school accountability and governance, (iii) range of curricular provision
Berk and Goebel (1987a)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 1-500, 501-1,200, 1,201-1,750, 1,751-3,700	Student outcomes: (i) Participation in extra curricular activities School organisation and management: (i) Range of extra curricular provision
Berk and Goebel (1987b)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 1-799, 800-1,599,	Student outcomes: (i) Participation in extracurricular activities School organisation and management:

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
					1600-2,299 2300-5,000	(i) Range of extra-curricular provision
Bickel and Howley (2000)	USA	Exploration of relationships	School size	Proxy measure Students per grade level	Continuous	Student outcomes: (i) Performance measured without prior attainment
Bickel *et al. (*2001)	USA	Exploration of relationships	School size	Whole school size studied Proxy measure: span size	Continuous	Student outcomes: (i) Performance measured without prior attainment School organisation and management: (i) Grouping arrangements, (ii) education economics: cost per student
Bonesronning (1996a)	Norway	Exploration of relationships	School size	Proxy measure: total teacher input defined as teacher man years available per graduate	Continuous	Student outcomes: (i) Performance measured with prior attainment
Bonesronning (1996b)	Norway	Exploration of relationships	School size	Proxy measure: total teacher man-years available for the graduates	Continuous	Student outcomes: (i) Performance measured with prior attainment School organisation and management: (i) Grouping arrangements: allocation of students to departments and resources
Bos *et al. (*1990)	Netherlands	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; dropout and class repeating (ii) attendance; truancy
Boswell and Carr (1988)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <500 500-1,000	Teacher outcomes: (i) Perceptions of their role and duties

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
					>1,000	
Bowen *et al. (*2000)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 0-399, 400-599, 600-799, 800-999, 1,000-1,344	Student outcomes: (i) Student attitudes; safety, support, satisfaction School organisation and management: (i) School accountability and governance, (ii) student teacher relationships
Bowles and Bosworth (2002)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment School organisation and management: (i) Education economics: economies of scale
Bradley and Taylor (1998)	England	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment
Bradley and Taylor (2003)	England	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied Proxy measure: Change in school size	Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) performance measured without prior attainment, (iii) student behaviour; attendance and truancy
Bruckerhoff *et al. (*2000)	USA	Evaluation: naturally occurring	School within a school	Proxy measure: Number of 9th grade students	Categorical Before and after school within school changes	Student outcomes: (i) Performance measured without prior attainment, (ii) attendance
Bryk and Thum (1989)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour, (ii) attendance and (iii) dropout

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
		Evaluation: naturally occurring				Teacher outcomes: (i) Morale and stress School organisation and management: (i) Grouping arrangements, (ii) resource availability
Bulach and Williams (2002)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Perceptions of school climate and culture School organisation and management: (i) Communication in school, (ii) accountability and governance
Caldas (1993)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment
Cicmanec *et al. (*2001)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Grade assigning behaviours
Coladarci and Cobb (1996)	USA	Exploration of relationships	School size	Whole school size studied	Categorical Small <800 large >1,600 Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) student attitudes, (iii) student behaviour; extra-curricular participation
Darling-Hammond *et al. (*2002)	USA	Evaluation: naturally occurring	School within a school	Whole school size studied	Categorical Comparison with	Student outcomes: (i) Performance measured with prior attainment, (ii) performance measured without prior attainment, (iii) post- school destination, (iv)

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
					before an after creation of school within school	student behaviour, (v) attendance Teacher outcomes: (i) Morale and stress School organisation and management: (i) Class size, (ii) grouping arrangements, (iii) student-teacher relationships, (iv) communication within the school, (v) range of curricular provision
Driscoll *et al.* (2003)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment
Echternacht (1981)	USA	Exploration of relationships	School size	Whole school size studied	Categorical Divided into two equal sized groups	Teacher outcomes: (i) Instructional difficulties
Edge and Friedberg (1984)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment
Edington and Gardener (1984)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Post-school destination, (ii) student attitudes; to society, regarding self, towards school, character, co-operation, change (iii) physical fitness and wellbeing
Egelund and Hansen (2002)	Denmark	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour School organisation and management: (ii) Perceptions of class disturbance
Eichenstein (1994)	USA	Evaluation: naturally occurring	School within a school	Proxy measure: sample of students from each house	Categorical No comparison	Student outcomes: (i) Student attitudes, (ii) attendance

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
					group, implicitly comparing to before school within school was created	Teacher outcomes: (i) Morale and stress School organisation and management: (i) School accountability and governance, (ii) range of curricular provision, (iii) relationships between school and wider community (iv) grouping arrangements
Elsworth (1998)	Australia	Exploration of relationships	School size	Proxy measure: VCE (Victorian Certificate of Education) enrolments	Continuous	School organisation and management: (i) Range of curricular provision
Epstein (1990)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	School organisation and management: (i) Grouping arrangements; grade span
Fetler (1997)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Dropout rate Teacher outcomes: (i) Experience; qualifications, percentage of newly qualified teachers, faculty growth
Forbes *et al.* (1993)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <450 >450	Student outcomes: (i) Performance measured without prior attainment School organisation and management: (i) Range of curricular provision, (ii) levels of funding
Foster and Martinez (1985)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student attitudes towards school Teacher outcomes: (ii) Morale and stress

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Fouts (1994)	USA	Evaluation: Naturally occurring	School within a school	Proxy measure Measures from a sample of students	Categorical Comparison with a control group not in the school within school	Student outcomes: (i) Performance measured without prior attainment, (ii) student behaviour; discipline (iii) attendance; (iv) student attitudes Teacher outcomes: (i) Attitudes towards role and school; (ii) morale and stress School organisation and management: (i) Parental perceptions of school
Fowler and Walberg (1991)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment, (ii) post- school destination, (iii) attendance
Franklin and Crone (1992)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <362 363-505 506-696 >697 (inclusive)	Student outcomes: (i) Performance measured without prior attainment, (ii) student behaviour, (iii) attendance Teacher outcomes: (i) Qualifications School organisation and management: (i) Class size
Franklin and Glascock (1996)	USA	Exploration of relationships	School size	Proxy measure: End-of-year membership	Categorical <372 372-552 >552	Student outcomes: (i) Performance measured without prior attainment, (ii) student behaviour, (iii) attendance School organization and management: (i) School configuration
Gill *et al.* (2002)	United Kingdom	Exploration of relationships	School size	Proxy measure: number of students aged 11-15	Continuous	Student outcomes: (i) Performance measured without prior attainment

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Gordon (1992)	USA	Evaluation: naturally occurring	School within a school	Whole school size studied Proxy measure: sample of students from school within school	Categorical Implicit comparison with non schools within schools	Student outcomes: (i) Performance measured without prior attainment, (ii) student attitudes, (iii) student behaviour; dropout, (iv) attendance Teacher outcomes: (i) role and duties School organisation and management: (i) School accountability and governance, (ii) relationship with parents (iii) relations with wider community
Grabe (1981)	USA	Exploration of relationships	School size	Proxy measure: total enrolment in the upper three grades	Categorical >580 = large <580 = small	Student outcomes: (i) Student attitudes; perceptions of alienation, (ii) student behaviour participation in school activities
Green and Barnes (1993)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 1-400 400-799 800-1,199 >1,200	Student outcomes: (i) Actions of administrators in regard to student misconduct
Haller (1992)	USA	Exploration of relationships Evaluation: naturally occurring	School size School consolidation or decline	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour, (ii) attendance
Haller *et al.* (1990)	USA	Exploration of relationships Evaluation: naturally occurring	School size	Proxy measure: number of students in graduating class	Categorical <25, 25-49, 50-99, 100-199, 200-299, 300-399, >400	School organization and management: (i) Comprehensiveness of curricular provision

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Haller *et al.* (1993)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment
Heck (1993)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment, (ii) student behaviour; suspensions, (iv) attendance
Henkin *et al.* (1996)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Social skills of principals
Holland and Andre (1994)	USA	Exploration of relationships	School size	Proxy measure: number of students per grade	Categorical <100 >250	Student outcomes: (i) Student attitudes; self-esteem, sex role orientation and attitudes to women's roles, (ii) student behaviour; participation in extra curricular activities School organisation and management: (i) Range of extra-curricular provision
Hough and Sills-Briegel (1997)	USA	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Categorical SS <100 SS 101-200 SS 201-300 SS >300	Student outcomes: (i) Performance measured without prior attainment School organisation and management: (i) Comparison between rural and consolidated schools
Howley (1996a)	USA	Exploration of relationships	School size	Proxy measure: average enrolment in a grade	Continuous	Student outcomes; (i) Performance measured without prior attainment
Howley (1999a)	USA	Exploration of relationships	School size	Proxy measure: whole school size divided by number of grades in school	Continuous	Student outcomes: (i) Performance measured without prior attainment

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Howley (1999b)	USA	Exploration of relationships	School size	Proxy measure: average enrolment per grade	Continuous	Student outcomes: (i) Performance measured without prior attainment
Huang and Howley (1993)	USA	Exploration of relationships	School size	Proxy measure: average enrolment per grade	Categorical <20 21-59 >60	Student outcomes: (i) Performance measured without prior attainment
Huber (1983)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <200 500 and over	Student outcomes: (i) Student attitudes, (ii) student behaviour, (iii) attendance
Hurd (1995)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Composite measure based on academic grades, aptitude, attendance, graduation
Ingersoll *et al.* (1997)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <300 300-599 >600	Teacher outcomes: (i) Salary, (ii) experience, (iii) induction and professional development School organisation and management: (i) school accountability and governance
Jacobs and Chase (1989)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <800 >800	Student outcomes: (i) Performance measured without prior attainment, (ii) student attitudes; satisfaction with school, (iii) student behaviour; participation in school activities

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Jewell (1989)	USA	Exploration of relationships	School size	Proxy measure: average school size for the state	Continuous	Student outcomes: (i) Performance measured without prior attainment Teacher outcomes: (i) Salary School organisation and management: (i) Nature of enrolment
Johnson *et al.* (2002)	USA	Exploration of relationships	School size	Proxy measure: span size	Continuous	Student outcomes: (i) Performance measured without prior attainment
Kaufman (2001)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <300 300-1,000 >1,000	Student outcomes: (i) Student attitudes, (ii) student behaviour School organisation and management: (i) School accountability and governance; policies (e.g zero tolerance policies)
Kearney (1994)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <200 200-400 400-600 600-800 >800	Student outcomes: (i) Performance measured without prior attainment Teacher outcomes: (i) Salary, (ii) number of teachers School organisation and management: (i) Education economics; growth and expenditure
Kemple (2001)	USA	Evaluation: researcher-manipulated	School within a school	Whole school size studied	Categorical Comparison of group in and outside the school within a school	Student outcomes: (i) Performance measured without prior attainment; graduation rates, (ii) post-school destination

Study	Country	Type of study		School size variable	Nature of school size variable	Outcomes
Kirjavainen and Loikkanen (1998)	Finland	Exploration of relationships	School size	Whole school size studied	Continuous	School organisation and management: (i) Education economics; efficiency/ inefficiency
Kowalski *et al.* (1983)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; school crime Teacher outcomes: (i) Principals perceptions of crime
Langbein and Bess (2002)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student attitudes, (ii) student behaviour; incidents and suspensions, participation in sports
Lee and Bryk (1989)	USA	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) performance measured without prior attainment
Lee and Burkam (2001)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <600, 601-1,500 1,501-2,500 >2,500	Student outcomes: (i) Attendance; dropping out
Lee *et al.* (1991)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Self-efficacy and satisfaction, (ii) perceptions of school organisation and management
Lee *et al.* (2000)	USA	Evaluation: naturally occurring	School size	Whole school size studied	Categorical <500 >1,500	School organisation and management: (i) School accountability and governance, (ii) student-teacher relationships, (iii) communication within the school, (iv) range of curricular provision

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Lee and Smith (1997)	USA	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Categorical <300, 301-600 601-900 901-1,200 1201-1,500 1501-1,800 1801-2,100 >2,100 Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) student attitudes; engagement
Leithwood and Jantzi (1997)	Canada	Exploration of relationships	School size	Whole school size studied	Continuous	School organisation and management: (i) School accountability and governance
Leung and Ferris (2002)	Canada	Exploration of relationships	School size	Whole school size studied	Categorical >1,000 1,000-1,499 1,500-1,999 >2,000 Continuous	Student outcomes: (i) Student behaviour; self-reported violence at school
Lien and Humphreys (2001)	USA	Exploration of relationships	School size	Proxy measure: enrolment per grade	Continuous	Student outcomes: (i) Performance measured without prior attainment
Luyten (1994)	The Netherlands, Sweden, USA	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Categorical <240, 240-359 360-499 500-999 >1,000 Continuous	Student outcomes: (i) Performance measured with prior attainment

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Ma (2001)	Canada	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; bullying and being bullied
Marsh (1993)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; participation in sport
McLaughlin *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment Teacher outcomes: (i) Perceptions of climate, self-influence and normative cohesion School organisation and management: (i) Class size
McMillen *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Continuous Categorical <200 201-500 501-750 751-1,000 1,001-1,250 1,251-1,500 1,501-1,800 >1,800	Student outcomes: (i) Performance measured without prior attainment, (ii) performance measured with prior attainment, (iii) student behaviour; violence report, (iv) attendance; dropout
McNeely *et al.* (2002)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student attitudes; connectedness with school
Melnick *et al.* (1987)	USA	Exploration of relationships	School size	Whole school size studied	Categorical Small<600 Medium 601-900	Student outcomes: (i) Performance measured without prior attainment School organisation and management:

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
					Large >900	(i) Range of curricular provision, (ii) range of extra-curricular provision
Mertens *et al.* (2001)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <500 500-749 750> distinction also made between urban and rural schools	Student outcomes: (i) Performance measured without prior attainment, (ii) student attitudes; self-esteem, collaboration, (iii) student behaviour; behavioural problems Teacher outcomes: (i) Collaboration School organisation and management: (i) Parental involvement, (ii) school climate, (iii) classroom practices, (iv) grouping arrangements
Mok and Flynn (1997)	Australia	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Performance measured without prior attainment, (ii) student attitudes; quality of life, attitudes, religious practices School organisation and management: (i) Student-teacher relationships
Monk and Rice (1997)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Coursework preparation
Monk (1987)	USA	Exploration of relationships	School size	Proxy measure: enrolments in grade 9-12 used	Categorical <100 100-200 200-300 300-400 400-500 500-1,000 1,000-1,500 1,500-2,000 2,000-2,500 2,500-3,000	Student outcomes: (i) Student behaviour; student take up of courses (enrolment) School organisation and management: (i) Range of curricular provision; length and depth

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Newmann *et al.* (1989)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Morale and stress, (ii) perceptions of community School organization and management: (i) School accountability and governance
Office of Academic Affairs (1999)	USA	Exploration of relationships	School size	Whole school size studied	Categorical small less than 299 medium 300-999 large 1,000+	Student outcomes: (i) Student attitudes; students' demands for early college options (programme to help transition to college) Teacher outcomes: (i) Teacher demands for early college options School organisation and management: (i) Range of curricular provision, (ii) range of extra-curricular provision
Page (1990)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <500 500-1000 >1000	Student outcomes: (i) Student attitudes; score on Loneliness Scale
Peterson *et al.* (1996)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Teacher outcomes: (i) Teacher morale and stress; sense of empowerment School organization and management: (i) School accountability and governance; decision-making structures and power relations
Pittman and Haughwout (1987)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Attendance School organisation and management: (i) Range of curricular provision, (ii) school climate

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Ramirez (1990)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 22-99 109-440 530-1,468 1,609-2,493	Student outcomes: (i) Performance measured without prior attainment Teacher outcomes: (i) Staff retention, (ii) teacher certification School organization and management: (i) Class size, (ii) range of curricular provision
Reese and Johnson (1988)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 500-1,000 1,001-1,500 1,501-2,000 >2,000	Teacher outcomes: (i) Morale and stress
Ristau *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <499 500-999 >1,000	Teacher outcomes: (i) Use of computers in business courses
Rollins *et al.* (1983)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <129 130-249 250-579 580-1,259 >1,260	Student outcomes: (i) Performance measured without prior attainment
Schoggen and Schoggen (1988)	USA	Exploration of relationships	School size	Proxy measure: number of students in the senior class	Continuous	Student outcomes: (i) Student behaviour; participation in school extra-curricula activities
Schreiber (2002)	USA	Exploration of relationships	School size	Proxy measure: number of full-time faculty	Continuous	Student outcomes: (i) Performance measured with prior attainment

Study	Country	Type of study	Study focus	School size variable	Nature of school size variable	Outcomes
Silins and Mulford (2000)	Tasmania & Australia	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student attitudes; engagement, (ii) student behaviour; participation Teacher outcomes: (i) perceptions of climate and culture School organisation and management: (i) School accountability and governance, (ii) communication within the school, (iii) relationships between school and wider community
Simpson and Marek (1988)	USA	Exploration of relationships	School size	Proxy measure: 10th grade enrolment	Categorical small = <150, large = >900, in grades 10-12	Student outcomes: (i) Performance measured without prior attainment
Smet (2001)	Belgium	Exploration of relationships	School size	Whole school size studied	Continuous	School organization and management: (i) Education economics; optimal size of education organisations
Spielhofer *et al.* (2002)	England	Exploration of relationships Evaluation: naturally occurring	School size	Proxy measure: size of year 11	Continuous	Student outcomes: (i) Performance measured with prior attainment, (ii) opportunity School organisation and management: (i) Range of curricular provision
Stiefel *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Categorical 0-600 600-2000 >2000 Continuous	School organisation and management: (i) Education economics; budget per graduate and per student
Stull *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Post-school destination School organisation and management: (i) Range of curricular provision

Study	Country	Type of study		School size variable	Nature of school size variable	Outcomes
Taylor and Bradley (2000)	England	Exploration of relationships	School size	Proxy measure: capacity and capacity utilisation rate	Continuous	School organisation and management: (i) Education economics; costs per student and staff hours
Thomas and Bullock (1992)	England and Wales	Exploration of relationships Evaluation: naturally occurring	School size	Whole school size studied	Categorical <400, 400-499 500-599 600-699 700-799 800-899 900-999 1,000+	Teacher outcomes: (i) Salary School organisation and management: (i) Education economics; impact of LEA funding formulas
Tomlinson and Mortimore (1990)	England	Evaluation: naturally occurring	School size	Whole school size studied	Categorical Comparison between six small secondary schools	Student outcomes: (i) Performance measured without prior attainment, (ii) post- school destination, (iii) student behaviour Teacher outcomes: (i) Experience School organisation and management: (i) Student teacher relationships, (ii) range of curricular provision, (iii) parental perceptions
Trybus and Li (1998)	USA	Evaluation: naturally occurring	School within a school	Whole school size studied	Categorical Comparison with students not in the school within school programme	Student outcomes: (i) Performance measured without prior attainment, (ii) post- school destination, (iii) student behaviour, (iv) attendance
Uerling (1986)	USA	Exploration of relationships	School size	Proxy measure grades 9-12 enrolment	Continuous	School organisation and management: (i) Range of curricular provision; school units meeting standard of state endorsement

Study	Country	Type of study		School size variable	Nature of school size variable	Outcomes
Van Batenburg and Lokman (1991)	The Netherlands	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Attendance; continuation into next year of vocational education
Walberg (1992)	USA	Exploration of relationships	School size	Proxy measure: average size of school in state	Continuous	Student outcomes: (i) Performance measured without prior attainment
Wasley *et al.* (2000)	USA	Evaluation: naturally occurring	School within a school School size	Whole school size studied	Categorical Comparison between small schools and implicitly with larger schools	Student outcomes: (i) Performance measured without prior attainment, (ii) student behaviour, (iii) attendance Teacher outcomes: (i) Teacher perceptions of work School organisation and management: (i) School accountability and governance, (ii) student-teacher relationships
Welsh *et al.* (1999)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; school misconduct
Welsh *et al.* (2000)	USA	Exploration of relationships	School size	Whole school size studied	Continuous	Student outcomes: (i) Student behaviour; school incidents
Whitehead *et al.* (1992)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <298 299-604 >605	Student outcomes: (i) Student behaviour; drug use - steroids and sport activity, illicit drug use
Whitener (1997)	USA	Exploration of relationships	School size	Whole school size studied	Categorical <150 150-299 300-499 500-749 >750	Teacher outcomes: (i) Staff retention; turnover rate

Appendix 4.1: Comparison of studies in systematic map and in-depth review

The following tables present the number of studies coded for each category of the EPPI-Centre generic and review-specific keywords; in most cases the categories are not mutually exclusive (shown at the bottom of each table). Proportions are given in brackets; this figure shows the proportion of studies coded within each category, as a proportion of all the studies either in the map (N=119) or in the in-depth review (N=31), rather than as a proportion of all the codes. This means that proportions do not add up to 1.0 unless the categories are mutually exclusive.

EPPI-Centre Generic keywords

Table 4.2.1: Comparison of identification of studies (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Contact	2 (0.02)	2 (0.06)
Handsearch	6 (0.05)	4 (0.13)
Electronic database	111 (0.93)	25 (0.81)

*categories mutually exclusive

Table 4.2.2: Comparison of origin of the studies (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
North America	100 (0.84)	23 (0.74)
Europe	8 (0.07)	1 (0.03)
Australasia	3 (0.02)	1 (0.03)
United Kingdom	9 (0.08)	6 (0.19)

*categories mutually exclusive, except for one study in systematic map coded twice (America and Europe)

Table 4.2.3: Comparison of topic focus of the studies (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Classroom management	3 (0.02)	0 (0.00)
Curriculum	71 (0.60)	18 (0.58)
Equal opportunities	5 (0.04)	2 (0.06)
Organisation and management	119 (1.0)	31 (1.0)
Policy	8 (0.07)	2 (0.06)
Teacher careers	9 (0.08)	1 (0.03)
Teaching and learning	66 (0.55)	21 (0.68)
Other topic focus	41 (0.34)	15 (0.48)

*categories not mutually exclusive

Table 4.2.4: Comparison of curriculum focus of the studies (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Art	1 (0.01)	0 (0.00)
Business Studies	3 (0.02)	0 (0.00)
Citizenship	3 (0.02)	1 (0.03)

Cross-curricular	20 (0.17)	7 (0.23)
Design and Technology	1 (0.01)	1 (0.03)
Environment	1 (0.01)	0 (0.00)
General	1 (0.01)	0 (0.00)
History	4 (0.03)	3 (0.10)
ICT	2 (0.02)	0 (0.00)
Literacy – first language	28 (0.24)	13 (0.42)
Literacy further languages	7 (0.06)	2 (0.06)
Mathematics	36 (0.30)	13 (0.42)
PSE	1 (0.01)	1 (0.03)
Physical education	4 (0.03)	0 (0.00)
Science	22 (0.18)	8 (0.26)
Vocational	5 (0.04)	0 (0.00)
Other curriculum	18 (0.15)	4 (0.13)

*Categories not mutually exclusive

Table 4.2.5: Comparison of population focus of the studies (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Learners	103 (0.87)	31 (1.0)
Senior management	7 (0.06)	1 (0.03)
Teaching staff	23 (0.19)	4 (0.13)
Non-teaching staff	2 (0.02)	0 (0.00)
Government	4 (0.03)	2 (0.06)
LEA officers	2 (0.02)	0 (0.00)
Parents	3 (0.03)	0 (0.00)
Governors	1 (0.01)	0 (0.00)
Other population focus	5 (0.04)	1 (0.03)

*Categories not mutually exclusive

Table 4.2.6: Comparison of the age of the learners (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Ages 0-4	2 (0.02)	1 (0.03)
Ages 5-10	33 (0.28)	14 (0.45)
Ages 11-16	114 (0.96)	30 (0.97)
Ages 17-20	75 (0.63)	16 (.52)
Ages 21 and over	3 (0.03)	0 (0.00)

*Categories not mutually exclusive

Table 4.2.7: Comparison of the educational establishment (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Higher education institution	4 (0.04)	0 (0.00)
Independent school	12 (0.10)	5 (0.16)
Nursery school	1 (0.01)	0 (0.00)
Post-compulsory education institution	1 (0.01)	0 (0.00)
Primary school	32 (0.27)	14 (0.45)
Secondary school	118 (0.99)	31 (1.0)
Other educational setting	1 (0.01)	0 (0.00)

*Categories not mutually exclusive

Table 4.2.8: Comparison of study types (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Exploration of relationships	108 (0.91)	31 (1.0)
Evaluation: naturally occurring	24 (0.20)	5 (0.16)
Evaluation: researcher-manipulated	1 (0.01)	0 (0.00)

*Categories not mutually exclusive

Review-specific keywords

Table 4.2.9: Comparison of student mix (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Mainstream	119 (1.0)	31 (1.0)
Ethnic minority	8 (0.06)	4 (0.13)

*Categories not mutually exclusive

Table 4.2.10: Comparison of the presentation of the school size variable measured (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Categorical	55 (0.46)	6 (0.19)
Continuous	70 (0.59)	29 (0.94)

*Categories not mutually exclusive

Table 4.2.11: Comparison of the measurement of the school size variable (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Whole school size	90 (0.76)	22 (0.71)
Proxy measure	33 (0.28)	12 (0.39)

*Categories not mutually exclusive

Table 4.2.12: Comparison of student level outcomes (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Performance measured with prior attainment	12 (0.10)	5 (0.16)
Performance measured without prior attainment	47 (0.39)	17 (0.55)
Performance (total)	54 (0.45)	19 (0.61)
Post-school destination	8 (0.06)	0 (0.00)
Student attitudes	23 (0.19)	4 (0.13)
Student behaviour	36 (0.30)	7 (0.23)
Attendance	21 (0.18)	5 (0.16)
Other	6 (0.05)	1 (0.03)

*Categories not mutually exclusive

Table 4.2.13: Comparison of teacher level outcomes (number (proportion))*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Morale and stress	9 (0.08)	0 (0.00)
Retention	2 (0.02)	0 (0.00)
Salary	4 (0.03)	0 (0.00)
Experience	6 (0.05)	1 (0.03)
Other	21 (0.18)	2 (0.06)

*Categories not mutually exclusive

Table 4.2.14: Comparison of school level outcomes*

Attribute	Number in systematic map (N=119)	Number in in-depth review (N=31)
Class size	5 (0.04)	1 (0.03)
Grouping arrangements	8 (0.06)	1 (0.03)
School accountability and governance	15 (0.13)	2 (0.06)
Student teacher relationships	6 (0.05)	1 (0.03)
Communication within the school	5 (0.04)	1 (0.03)
Parental involvement	3 (0.03)	0 (0.00)
Relationships between school and wider community	3 (0.03)	1 (0.03)
Range of curricular provision	19 (0.16)	1 (0.03)
Range of extra-curricular provision	5 (0.04)	0 (0.00)
Economics of education	8 (0.06)	5 (0.16)
Other	8 (0.06)	0 (0.00)

*Categories not mutually exclusive

Appendix 4.2: Aims and overview of studies in in-depth review

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	analysis	Date of data collection
Abbott *et al.* (2002)	Using data from Washington middle schools, the study attempts to identify the ways in which district size, school size and family income level interact to effect student achievement. This paper is a replication study of Bickel and Howley (2000).	1. Fourth and seventh grade student Washington assessment of student learning (WASL) scale scores in reading and maths aggregated to the school level WASL scale score Grade 7: Maths 366.29 (SD 17.81) Reading 393.79 (SD 6.81)	Hierarchical linear modelling	Data are from the testing year 2001
Atkinson and Wilson (2003)	Using national matched exam datasets, the study examines patterns of attainment of boys and girls at different stages of their secondary education in the UK and investigates factors that may affect student outcomes.	1. KS3 mean and subject level, result (Maths, English, Science) (range 1-9 average 5) 2. GCSE total points and subject level (Maths, English, Science) (range 0 (ungraded) -8 (A*) for each GCSE) 3. Value added scores (between KS3 and GCSE) 4. 5 GCSEs A* to C Overall KS3: 4.84 (SD 1.23) English KS3: 4.66 (SD 1.56) Maths KS3: 5.02 (SD 1.35) Science KS3: 4.93 (SD 1.18) Total GCSE: 40.02 (SD 17.85) GCSE English: 4.58 (SD 1.59) GCSE Maths: 4.11 (SD 1.86) GCSE Science: 4.30 (SD 1.74) Value added: -0.46 (9.62)	Regression analyses Probit analysis of likelihood of achieving five GCSEs A*-C	Data are from 1997 and 1999
Bedard *et al.* (1999)	This paper uses newly available STAR test score data from California to explore the relationship between school size and the distribution of test scores across elementary, middle and high schools.	1. The distribution of test scores (maths, reading, language, social science, science and spelling) within the schools in terms of whether the schools distribution is poor, average or good across grades 3, 8 and 10. Percentage of schools by type; based on distribution of maths scores: Middle schools: Good 19.2, Average 5.7, Poor 75.1 High schools: Good: 19.6, Average: 16.0, Poor 64.4	Standard and modified ordered probits	It would seem to be 1998, but this is a little unclear

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Bickel *et al.* (2001)	Using state data from Texas, the 2001 paper has two objectives: first to determine if the size by SES interaction effect proves robust across alternative regression model specifications, as it did across differing states; and second, to make a tentative judgment as to whether the equity gains associated with smaller schools are incompatible with the need for fiscal efficiency. The 1999 paper focuses on replicating the size, SES and interaction effect seen in other states in USA using a Texas dataset. (See, for example, Howley, 1999a and b.)	2001 paper: 1. Achievement: measured at grade 10 from reading, writing, maths, and as a composite score of the Texas Assessment of Academic Skills 2. Expenditure per student (dollars) 1999 paper: 1. Achievement: measured at grades 3, 5, 8,10 from reading, maths, writing (not grades 3 and 5) of the Texas Assessment of Academic Skills Grade 10 (2001): Reading: 39.17 (SD 2.30) Maths: 45.51 (SD 4.08) Writing: 32.88 (SD 1.80) Expenditure: 4,745.67 (SD 1318.94) Grade 8 (1999): No overall statistics presented. Only presented for single unit (SU) and all others (O) Maths: SU: 46.7 (SD 4.3), O: 45.3 (SD 5.2) Writing: SU: 31.4 (SD 2.9), O: 30.5 (SD 2.7) Reading: SU: 39.2 (SD 3.0), O: 30.5 (SD 2.7)	Regression analyses Hierarchical linear modelling	Data are from the school year 1996-97
Bickel and Howley (2000)	The 1999 paper aims to see if relationship between school size and achievement and SES in California and Virginia can be replicated in Georgia. The 2000 paper aims to extend this to take into account district (and class) size as well as school level interactions.	1. Iowa test of basic skills at grade 8 (+ grades 3 and 5 in 1999 paper only), mean school-level percentile scores for seven subtests (reading comprehension, reading vocabulary, maths, language arts, science, social studies, research skills) and a composite global gauge of achievement. 2. Georgia high school graduation test (grade 11): school-level percentage of students passing the first administration of the GHSG, includes English, maths, science, social studies, composite score. **Grade 8** Reading comprehension: 47.02 (SD 12.88) Maths: 2.26 (SD 12.42) Reading vocabulary: 43.82 (SD 15.05) Language arts: 54.20 (SD 12.72) Social studies: 51.31 (SD 12.04) Science: 51.07 (SD 13.88) Research skills: 53.01 (SD 12.60)	Regression analyses	Data are from school year 1996-97

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
		Composite: 51.25 (SD 13.71) **Grade 11** English: 92.87 (SD 5.18) Maths: 85.33 (SD 9.77) Science: 70.66 (SD 15.22) Social studies: 75.14 (SD 12.97) Composite: 63.89 (SD 16.41)		
Bowen *et al.* (2000)	Using data from a national sample of middle school students, the study examines the 'effects' of school size on the social environment at school. It was hypothesised that larger school size would be associated with lower levels of school satisfaction, teacher support and school safety.	1. School satisfaction: Five-item dichotomous scale 2. Teacher support eight-item dichotomous scale 3. School safety: ten-item scale; responses: three-item Likert scale School satisfaction: 3.87 (SD 1.36) Teacher support: 6.52 (SD 2.06) School safety: 24.14 (SD 5.41)	Factorial analysis of variance	Data collected between October 1996 and February 1997
Bowles and Bosworth (2002)	Using a data set for 17 Wyoming school districts, the study aims to investigate economies of scale in schooling by considering the 'effect' of school size on the cost of education per student. Asks if small schools receive the same funding per student as large schools: will the students in the small schools receive an equal education? Does school size affect the average cost of producing education?	1. Per student cost average across four years, 2. Average student test scores No descriptive information given	Regression analyses; least squares and simultaneous equations	Data were collected across four school years 1994-1998
Bradley and Taylor (1998)	Uses English school data sets from 1992-96 to explore the relationship between academic achievement and school size, whilst controlling for a range of other explanatory variables including student level and school level characteristics.	1. Exam performance (GCSEs A*-C) of all secondary schools in England covering the period 1992-96 Proportion of students receiving five or more GCSEs A*-C 1992: 35.5 1993: 37.8 1994: 39.9 1995: 40.7 1996: 42.1	Ordered logit equations	Data were used from the school years 1992 - 96
Bradley and Taylor (2003)	The study investigates the determinants of school outcome for the secondary education sector in England 1992-2002. It also investigates residual impact of school size on exam performance and truancy after accounting for SES of school intake. It considers the extent that the introduction of market forces into	1. Value added KS3 to KS4 2. Average score in KS4 exams (GCSE/GNVQ -year 11) 3. Change in school exam performance (based on proportion of students receiving five or more GCSE results A*-C) 1993-2002 4. Truancy rate% and absence rate % Value added KS3 to KS4: 98.7 (SD 2.6)	Multivariate regression analysis	Based on data from 1992-2002 to look for change over time

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
	secondary education affected the performance (i.e. efficiency) of schools and whether there have been any detrimental equity effects.	Average score in KS4 exams: 34.5 (SD 7.5) Proportion of unauthorised absence: 1.2 (SD 1.2) Proportion of authorised absence: 7.8 (SD 1.9) Change in exam performance 1992-2002 = 13.3%		
Driscoll *et al.* (2003)	Using state data from California, the study examines the impact of district size on student academic performance, whilst controlling for characteristics of the student population and other environmental factors, including class and school size.	1. 1999 California academic performance index (API) a weighted average of Stanford test scores. Analysed across elementary, middle and high schools. Mean API: 629 (SD 131) This appears to be an average across elementary, middle and high-school settings.	Instrumental variable estimation technique	1999 school level data were used; 2000 data were drawn on to consider change scores but not considered reliable.
Fetler (1997)	Using published data from the California Department of Education (1993-96), this study analyses and discusses high-school dropout rates in relation to measures of school size, location, growth, student poverty, teacher education and experience. The main focus of the study is on teacher education and experience. Asks what is the impact of teacher education and experience on student dropout rates, controlling for school size, location and growth, and student poverty?	1. School average dropout rate Mean dropout rate: 3.8 (SD 3.7) Median dropout: 2.1	Stepwise Backwards Regression analysis	The study used published data collected between 1993 and 1996; inclusive; data were used to compute an average score.
Gill *et al.* (2002)	The survey is concerned with students literacy in three subject areas, reading, maths, science. This report is of the English Survey carried out as part of an international comparison of literacy in these areas of students aged 15 years old. 1. What is the level of literacy (as defined above) in the three areas in England? 2. How does the level of literacy in England compare internationally? 3. What, if any, relationships can be identified between levels of literacy, student, school and/or home characteristics?	1. Reading literacy: The only outcome that is reported by school size is reading literacy. Distribution of student proficiency on the combined reading scale Mean score: 523 (SD 100) 5^{th} %tile: 352 25^{th} %tile: 458 75^{th} %tile: 595 95^{th} %tile: 682	Stepwise backwards multiple regression	Data collected in 2000

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Heck (1993)	Using state data, the study attempts to further understanding of how school characteristics and school attitudes toward knowledge and achievement may be useful in explaining the types of outcomes that are produced. It is hypothesised that a set of variables representing the school academic indicators (access to knowledge, press for achievement, level of teacher professionalism) will be significantly related to achievement, attendance and behavioural outcomes the school produces.	1. Reading and maths attainment scores on Stanford achievement test 2. Average daily attendance 3. Number of student suspensions for significant offences No descriptive statistics provided	Stepwise forced entry linear regression	Student outcome scores were collected over a two-year period 1989-91.
Howley (1996)	The study tests hypotheses about the relationship between size of educational units (schools and districts) and aggregate student achievement in West Virginia. Null hypotheses: 1. What is the (zero-order) relationship between school size and student achievement in West Virginia schools? 2. Does SES regulate the relationship of school size and student achievement in West Virginia?	1. Achievement in grades 3, 6, 9 and 11 in the comprehensive test of basic skills. No overall descriptive statistics	Regression analyses	School size data were taken from 1990, whilst achievement scores were taken from either spring 1991 or autumn 1990. Free schools meal data were taken from autumn 1990.
Howley CB (1999a)	Through the use of national data sets for Montana elementary and high schools, the authors assess the relationship between school size, SES and academic performance across grades 4, 8 and 11. This line of inquiry tests the 'interaction hypothesis' of school and district size. The interaction hypothesis expresses the possibility that the degree (i.e. strength or weakness) and directionality (positively or negatively) of the relationship of size to achievement is contingent on community SES.	1. Student tests at grade 4, 8, 11 using comprehensive test of basic skills, Iowa test of basic skills and the Stanford achievement test. (Note: not everybody uses the same test.) No overall descriptive statistics Grade 8 Larger schools: Mean: 57.3 Smaller schools: Mean 58.2 Grade 11 Larger schools: Mean: 56.2 Smaller schools: Mean 55.7	Regression analysis	Data were for the academic year 1995-96.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Howley (1999b)	Through the use of national data sets for Ohio elementary and high schools, the authors assess the relationship between school size, SES and academic performance across grades 4, 6, 9 and 12. This line of inquiry tests the 'interaction hypothesis' of school and district size. The interaction hypothesis expresses the possibility that the degree (i.e. strength or weakness) and directionality (positively or negatively) of the relationship of size to achievement is contingent on community SES (p 8).	1. Ohio proficiency tests in reading, writing, maths, citizenship and science with scores aggregated as percentage passing. Tests taken at grades 4, 6, 9 and 12. Both pass rates and advanced pass rates for grades 6 and 12 are presented. No overall descriptive statistics a=mean pass, b=mean advanced pass Grade 6 Larger schools: a. 55.8% b. 7.2% Smaller schools: a. 55.3% b. 6.2% Grade 9 Larger schools: a. 76.7% Smaller schools: a. 82.1% Grade 12 Larger schools: a. 64.9% b.13.5% Smaller schools: a. 64.9% b.12.7%	Regression analysis	Montana Office of Public Instruction (OPI) was contacted in August 1998 to obtain test score data.
Johnson *et al.* (2002)	Using data from Arkansas, the study explores the relationship between school and district size and SES, testing the notion that the best size for schools and districts - in terms of their capacity to cultivate academic excellence - depends on the poverty level in the communities they serve (p 4). This study replicates research (see, for example, Howley and Bickel) considering the relationship between academic achievement in elementary and high school, SES and school and district size.	1. School and district level achievement at grades 5, 7 and 10 (Stanford achievement test (SAT)) and grades 4, 8 (Benchmark maths and literacy tests) across two or three years (SAT '98, '99, '00 and Benchmark '99 and '00). The average score was used, computed from the grades across all the years. SAT mean percentile rank: Grade 7: 48.09 (SD 8.81) Grade 10: 46.89 (SD 7.63) Benchmark grade average mean: Grade 8 literacy: 23.69 (SD 12.79) Grade 8 math: 15.11 (SD 9.02)	Regression analyses	Data were used spanning the years 1998-2000.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Kirjavainen and Loikkanen (1998)	The purpose of this paper is to study the efficiency of Finnish senior secondary schools. Implicitly the study asks: 1. What is the efficiency of senior secondary schools in Finland? 2. How does this differ, depending on the inputs and outputs used in the DEA model? 3. How is this altered depending on the assumptions of constant returns to scale (CRS) and variable returns to scale (VRS)? 4. How is efficiency (calculated in the DEA models) related to number of variables of interest, such as class size and school size from the input side and matriculation exam results from the output side?	1. Inefficiency score: calculated using different input and output variables, including number of students who passed their grade, number of graduates, teaching hours per week, experience of teachers in the school. Average efficiency dependent on model used to calculate efficiency Model 3 CRS: Mean 81.9 Range 43.8-100 Model 3 VRS: Mean 84.1 Range 58.4-100 Model 4 CRS: Mean 91.3 Range 59.7-100 Model 4 VRS: Mean 93.7 Range 59.8-100	Data envelopment analysis (DEA) TOBIT analysis	Cross sectional data are used for 1988-91. Share of state grants is taken from the year 1992.
Lee and Smith (1997)	Using a national sample of high school students, the study (1997a paper) aims to identify an ideal high school size, defined in terms of student learning. A second objective is to define the optimal size in terms of the equitable distribution of learning within schools. A third objective is to identify whether the ideal size is constant across different types of high schools, defined in terms of the social background of the students they serve. The study (1997b and 1995 papers) aims to investigate how elements of high school organisation affect learning and its distribution, and whether organisational differences explain the educational advantages that have been found in previous studies that focused on structural practices in secondary schools. Paper 1 asks: 1. Which size high school is most effective for students learning? 2. In which size high school is learning most equally distributed? 3. Are size 'effects' consistent across high schools defined by their social composition? Paper 2 asks: 1. What are the differences in the social and academic organisation of high schools that report different types of restructuring practices? 2. Do the benefits of attending high schools that reported several	#1: 1. Achievement gains over 8-12th grade in maths and reading, from NELS test scores. #2: 1. Achievement gains over two spans 8-10 and 10-12 for maths and science, from NELS test scores. #3: 1. Achievement gains from 8-10 in maths, reading, and science and 2. engagement with school. Average unweighted gain scores (SD) in reading and maths based on school size category: Maths: <300: -0.87 (0.38), 301-600: -0.09 (0.24), 601-900: 1.37 (0.63), 901-1200: 0.61 (0.16), 1201-1500: 0.07 (0.19), 1501-1800: -0.16 (0.28), 1801-2100: -0.50 (0.22), >2100: -1.57 (0.67) Reading: <300: -0.34 (0.83), 301-600: 0.07 (0.80), 601-900: 0.52 (0.94), 901-1200: 0.48 (0.88), 1201-1500: 0.14 (0.99), 1501-1800: -0.08 (0.96), 1801-2100: -0.46 (0.81), >2100: -0.77 (0.92)	Hierarchical linear modelling	The study uses three panels of data from the NELS (1998, 1990 and 1992).

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
	non-traditional structural practices persist in later grades? 3. Do differences in the social and academic organisation of high school help explain the positive 'effects' of structural practices on learning and its equitable distribution?			
Lee and Burkam (2001)	In this study, we explore how high schools, through their structures and organisations may influence their students' decisions about whether to stay in school until graduation or drop out. 1. Within the students' high schools, which background factors are associated with the decision to drop out? 2. What features of high schools structure, curriculum and social organisation are associated with dropping out, once personal background and school demographics are accounted for? 3. Is the influence of school social organisation on dropout decisions contingent on school structure and, if so, what is the nature of these contingencies?	1. Student drop out between 10th and 12th grade. School average unweighted drop out described by school size category (no SD) <600: 5.3%, 601-1500: 7.0%, 1501-2500: 11.8%, >2500: 7.5%	Hierarchical linear modelling	Data were collected from 1990 to1992.
Leung and Ferris (2002)	To investigate the relationship between school size and youth violence, and a range of other potentially explanatory factors Asks if school size has an independent effect on youth violence when demographic characteristics are controlled for.	1. Violence: whether an individual reported participating in some form of violent behaviour over the year (1995). Proportion reporting participation in violent behaviour 0.45	Binary logit model	Data are used for 1995.
Ma (2001)	Aims to identify individual and school characteristics of victims and offenders of bullying in middle school, and to examine the victim-bully cycle, using cross-sectional data from the New Brunswick School Climate Study. Asks; (a) What are student-level characteristics that are associated differently with victims and bullies? (b) What are student-level characteristics that are associated differently with victims and bullies? (c) For those student-level characteristics in (b), are they associated more strongly with victims or bullies? A similar group of research questions was asked for school-level characteristics.	1. Students being bullied (termed 'victims') - the extent to which a student had been verbally and/or physically victimised. 2. Students bullying others (termed 'bullies') - whether a student had participated in bullying activities against others. Sample average for being bullied or bullying (based on a scale of 1-4) Grade 6: Bully: 1.89 (SD 0.91) Being bullied: 1.72 (SD 0.61) Grade 8: Bully: 1.75 (SD 0.84) Being bullied: 1.54 (SD 0.55)	Hierarchical linear models	Data were collected in 1996.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
McLaughlin *et al.* (2000)	The aim of this paper is to show the potential value of a linkage between SASS and data on student academic achievement. To achieve this aim, the approach is in two stages: first, matching1993-94 SASS data with state reading and mathematic scores for public schools in twenty sates; and second, by combining these data sources to identify school level correlates of student achievement in a broad sample of American public schools. Two major questions are addressed by these analyses: 1. Are organisational factors and school behavioural climate correlates of school mean assessment scores? 2. How is empirical evidence on the correlates of achievement affected by the removal of between-state variation from measures?	1. School average reading and maths score taken at grades 4, 8 and 11 (Note: ages are not particularly clear.) Inter-relationships between all variables studied, therefore results can also be drawn about size and school climate, teachers' self-perceptions of their influence, normative cohesion and class size. Average school maths score: Middle: 49.7 (SD 8.0) High: 49.3 (SD 8.4) Average school reading score: Middle: 48.7 (SD 7.7) High: 48.2 (SD 8.7) Other variables are composites for which averages are only given for individual items (scales used are unclear)	Partial correlations Ordinary least squares multivariate regression analyses Structural equation modelling	The study analyses data collected by the SASS in 1993-94, and the State NEAP carried out in 1994.
McMillen *et al.* (2000)	To understand better how school size relates to achievement and behaviour in North Carolina What is the relationship between school size and achievement? What is the relationship between school size and violence? What is the relationship between school size and dropout rate?	1. Student achievement scores in maths, reading, English and algebra. Scores from end-of-grade (EoG) and end-of-course (EoC) testing databases. Scale scores for different grades and courses were converted to standard scores with a mean of 50 prior to conducting the analysis. For grade 9-12 schools, the achievement data consisted of EoC scores for Algebra I, English I, History, Biology I and Economic, Legal and Political Systems. 2. Number of violent incidents per 100 students. Type of incidents included in this count were assaults, drug possessions, weapons possessions, robberies, sexual offences. 3. School dropout rates Mean standardized achievement scores presented in school size categories: Reading G6-8: <400: 50.75, 400-700: 49.87, >700: 49.56 Maths G6-8: <400: 50.92, 400-700: 50.00, >700: 49.48 English G9-12: <700: 50.08, 700-1000: 49.83, 1001-1500: 49.74, >1500: 49.67 Algebra G9-12, <700: 47.95, 700-1000: 48.25, 1001-1500: 48.22, >1500: 47.83 Achievement gains Reading G6-8: <400: 0.36, 400-700: -0.04, >700: -0.12	Partial correlations	Data were analysed from the 1997-98 and 1998-99 school years.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
		Maths G6-8: <400: 0.18, 400-700: 0.08, >700: -0.13 English G9-12: <700: 0.43, 700-1000: -0.03, 1001-1500: 0.13, >1500: -0.42 Algebra G9-12: <700: 0.65, 700-1000: 0.35, 1001-1500: 0.38, >1500: -0.06		
McNeely *et al.* (2002)	The central question this paper asks is whether school structure and environment are associated with shifts in the average level of school connectedness. Hypothesis specified related to school size: 1. As schools grow, they become more bureaucratic. Connections between individuals become less personal, and both students and staff feel less connected to the school. 2. In larger schools, students have fewer opportunities to participate in extra-curricular activities. Participation in extra-curricular activities is positively associated with student connectedness. 3. Larger schools have larger class sizes and larger classes make personal connections between students and teachers more difficult.	1. School connectedness (based on five-iitem Likert style scale) Mean school connectedness (5 point Likert scale) 3.64 (SD.25)	Hierarchical linear modelling	Data from students were collected in 1994-95; presumably, the school characteristics data are also from this period.
Silins and Mulford (2000)	The study examines the nature and strength of inter-relationships between 12 variables chosen to expand our understanding of the nature and processes of teacher leadership and organisational learning; and their impact on student participation in, and engagement with, the school.	12 variables (composite measures of individual survey items measured on a five-point Likert scale) considered in three categories, each measures is used as a dependent and independent variable: 1. School context variables: (i) school size (ii) SES. 2. Internal school variables: (i) resource availability; (ii) leadership style; (iii) satisfaction with leadership; (iv) response to the needs of the community, and ability to work with the local community; (v) collective teacher leadership; (vi) organisational learning; (vii) student perceptions of teachers' work 3. Student outcome variables: (i) student participation and (ii) student engagement All items in regression are composite measures; descriptive statistics are only given for individual items rather than composites.	Latent variable least squares path analysis	Data were collected during 1997-99.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Spielhofer *et al.* (2002)	Broad aims are to explore impact on performance of school size and single sex education. Asks about what is the impact of school size and single sex education on performance (especially at KS2 and KS4) and also the probabilities of entry to higher tiers at KS3 in maths and science; and probabilities of taking various subjects.	Multilevel modelling: 1. Total GSCE point score (points derived from A*=8, A=7, B=6 down to G=1) 2. Average GSCE point score 3. Number of GCSEs taken 4. Maths point score 5. English language point score 6. Total science score (e.g. grade CC for double science= 10 points) 7. Average science points score 8. Number of science GCSEs In logistic regression: Probabilities of entry to higher tiers at KS3: 1. In maths 2. In science Probabilities of taking various subjects/options at GSCE: 1. Double award science 2. Design and Technology - food, graphics, resistant materials 3. Physics 4. Chemistry 5. Biology 6. French plus German No descriptive means and standard deviation information	Multilevel modelling and logistic regression	Based on data from 1996 to 2001.
Stiefel *et al.* (2000)	The study explores the relationship between school size and budgets per student and per graduate to consider the possible impact of New York policy developments on cost of schooling. Asks: what is the impact of school size on budgets and performance in New York City high schools?	1. Budget per student (total general education plus part-time special education budget divided by the number of students registered in general education) 2. Budget per graduate (total budget per students multiplied by four, divided by the number of graduates who graduate in four years) Average budget per student in dollars: 6,790 (SD 940.3) Average budget per graduate in dollars: 65,559 (SD 43,621)	Least squares regression	Datasets were accessed for the 1995-96 school year.

Study	What are the study research questions and/or hypotheses?	Outcome variable and descriptive statistics	Methods of analysis	Date of data collection
Taylor and Bradley (2000)	To identify determinants of costs per student in English secondary schools. Focuses on mismatch between demand and supply for secondary school places on school costs. Uses school capacity utilisation rate to indicate short-run deviations from student capacity and student capacity as an indicator of school size to capture scale effects on costs per student. Estimate extent to which size of schools and their capacity utilisation rates affect student costs per student.	1. Costs per student (logged) (grant maintained only) 2. Staff hours per student (logged) (grant maintained only) 3. Teaching hours per student (logged) 4. Support hours per student (logged) Average teaching hours per student: 2.47 Average support hours per student: 0.54	Logarithmic and quadratic regression analyses	Data were collected for period 1993-97.
Welsh *et al.* (1999)	Drawing on control theory, school climate theory and social disorganisation theory, this study examined the relative influence of individual, institutional and community factors on misconduct in Philadelphia middle schools. Implicitly the paper asks: What is the relationship between student misconduct and individual level factors, and institutional and community level factors? What are the differences when considering the above as conceptualised as a local model (i.e. the area around the school) and an imported model (i.e. the area where the student actually resides)?	1. School disorder: measured by the student misconduct scale created from the effective school battery (ESB) student survey items (four items on a dichotomous scale 0=yes and 1=no – reverse scored low scores indicate higher levels of misconduct). Average misconduct scores 0=high and 4=low 2.53 (SD 1.23)	Hierarchical linear modelling	The survey was conducted during the period 1994-95. Secondary data are used from 1990 to 1993.
Welsh *et al,* (2000)	Guided by school climate and social disorganisation perspectives, this study attempts to assess the direct and indirect effects of community and school level variables on school disorder rates. The study uses path analysis to examine the goodness of fit of hypothesised causal models. The authors constructed variables that capture crime and socio-demographic measures of the census tract in which the schools are located (local model) and variables aggregated to the school from the communities in which students actually reside (imported model). The null hypothesis is that the models developed fit equally well.	1. School disorder: data on incidents occurring in or on school property reported to school police for the 1992-93 school year, as well as dismissal rates from each school during the 1990 school year. Average factor score Mean: 0.00 SD: 1.00 Minimum –2.06 Maximum 2.35	Path analysis	Data for this study were taken from 1990 to 1993.

Appendix 4.3: Sampling, student and school characteristics

Study/ Country	Sampling		Student characteristics		School characteristics
		Sample size	Age/grade Gender	SES/ Ethnicity /SEN	
Abbott *et al.* (2002) USA: Washington	Sampling frame: All public schools >10 Students Sampling: Criterion - All schools meeting above criteria	Total schools N = 417 (seventh grade) Total students: NS Mean students per school: 177 (SD: 112) Attrition: NA Missing data: NS	Age 12-13, grade 7, year 8 Mixed sex	School level data: Mean free school meals 33.27% with a standard deviation of 21.33. Ethnicity: NS SEN: NS	Funding: All public schools Organizational structure: NS Class size: NS Student/teacher ratio: NS Use of streaming: NS
Atkinson and Wilson (2003) England	Sampling frame: All secondary schools in England from the DFES exam datasets KS3 and CSE/GNVQ, for cohort 1997-99 and school census Sampling: Criterion-based All schools meeting above criteria as well as KS3 result and GCSE & GNVQ result needed	Total schools: 3,129 Total students: 517,695 Mean students per school: 1,042 (range: 50–2,361) Attrition: NA Missing data: Students for whom KS3 results could not be obtained were excluded.	Age 13-14, grade 8, year 9 Age 15-16, grade 10, year 11 Mixed sex 49.24% girls, 50.76% boys	Free school meals included as an explanatory variable but no explicit information is given. Ethnicity: NS SEN: All individuals in the sample had taken KS3 tests and GCSE exams.	Funding: Publicly funded schools only Organizational structure: Secondary schools Class size: NS Student/teacher ratio: NS Use of streaming: NS
Bedard K *et al.* (1999) USA: California	Sampling frame: All publicly funded schools in California Sampling: Criterion sampling 1. School configurations that were not K-5 or 6, 6 - 7 or 8 and 9 - 12 2. Charter schools, magnet schools	Total schools: 801 middle schools and 618 high schools Total students: NS Mean students per school: 353 (middle), 457 (high)	Age 13-14, grade 8, year 9 Age 15-16, grade 10, year 11	School level data are presented for: Families who receive aid for dependent children: 17% middle school, 14% high school Free school meals: 47% middle school, 31% high school	Funding: All publicly funded schools Organizational structure: High schools and middle schools Class size: NS

	Sampling		Student characteristics		
	and juvenile detention centres 3. Schools with fewer than 10 students writing a specific exam (for confidentiality excluded by department of California). Later on, to ensure outliers were not driving the results: 4. Elementary schools with more than 250 students, middle schools with more than 700 students and high schools with more than 900 students per grade were excluded.	Attrition = See below Missing data: Some schools were excluded due to unavailable data in the analyses. Figures can be seen in the tables at the back of the paper.	Mixed sex	% Limited English proficiency: 21% middle school, 15% high school Ethnicity, School level averages: % Black: 9% middle school, 8% high school % Hispanic: 38% middle school, 35% high school % Asian: 9% middle school, 10% high school % Pacific: 6% middle school, 6% high school % Philipino: 2% middle school, 3% high school % Native American: .8% middle school, 1% high school SEN: NS	Student/teacher ratio: Average 23:1 in middle schools and 24:1 in high schools. Use of streaming: Streaming in some subjects
Bickel and Howley (2000) USA: Georgia Reported across two papers (1999 and 2000)	Sampling frame: Georgia state dataset of schools in 1996 school year Sampling method = All schools and all students with test scores	Total schools (dependent on analysis) 1999: 367-371 (grade 8) 1999: 303-304 (grade 11) 2000: 367 (grade 8) 2000: 298 (grade 11) Total students: NS Mean students per grade: 280 Attrition = See below Missing = 1,626/1,800 (90.3%) public schools in Georgia in 1996-97 but schools with missing relevant test data at grades 3, 5, 8 or 11 were not able to be included.	Age 13-14, grade 8, year 9 Age 16-17, grade 11, year 12 Mixed sex	Grade 8: schools mean (SD) % students eligible for free and reduced cost lunch 45.3 (22.9) Grade 11: schools mean (SD) % students eligible for free and reduced cost lunch 33.5 (21.3) Grade 8: schools mean (SD) % students black: 37.3 (29.7) % students other racial or ethnic minority groups; 4.1 (5.4) Grade 11: schools mean (SD) % students black; 38.0 (29.8) % students other racial or ethnic minority groups; 3.8 (5.0) SEN: NS	Funding: Publicly funded schools only Organizational structure: High, Middle and K-12 schools Class size: NS Student/teacher ratio: 16.1 (1.5) for grade 8, 17.0 (2.3) for grade 11 Use of streaming: NS

	Sampling		Student characteristics		
Bickel *et al.* (2001) USA: Texas (Also reported in Bickel, 1999)	Sampling frame: All schools in Texas that report to Texas Department of Education maintain records. Sampling: All schools	Total schools: In 1999 paper: 1,441-1,448 (grade 8 schools) and 1,190-1,197 (grade 10 schools) In 2001 paper: 1,001 high schools Total students: NS Mean students per school: 877 Attrition: See below Missing: In 2001, 196 high schools were excluded as there were not values available for one or more of the variables used in the analyses.	Age 13-14, grade 8, year 9 Age 15-16, grade 10, year 11 Mixed sex	In the 2001 paper, 36.51% (SD 30.93) are recorded as having free or reduced price meals. In the 1999 paper a total average across all schools is given of 49.8% having free or reduced price meals. In 2001 paper only: percentage of Black students 11.07 (17.34) percentage of Hispanic students 27.73 (27.78) percentage enrolled in full time special education 13.54 (SD 6.08)	Funding: NS Organisational structure: High, Middle & K-12 schools Class size: NS Student/teacher ratio : for high schools: 13.24 (SD 3.15) Use of streaming: NS
Bowen *et al.* (2000) USA	Sampling frame: Not clear Sampling: Two-stage stratified sampling design (1) Selecting schools (2) Selecting students. It is stated the sampling design was specified to ensure adequate representation of students by gender, race, size of place, grade enrolment and region.	Total schools: 39 Total students: 945 Mean students per school: 689 (range: 70–1,393) Attrition: NA Missing: NS	Age 11-12, grade 6, year 7 Age 12-13, grade 7, year 8 Age 13-14, grade 8, year 9 Mixed sex 50.5% boys	31.9% respondents get free or reduced price school meals Adolescents of colour = 36.4% SEN: NS	Funding: Public school students Organisational structure: Middle schools grades 6-8 Class size: NS Student/teacher ratio: NS Use of streaming: NS
Bowles and Bosworth (2002) USA: Wyoming	Sampling frame: Unclear Paper states that 17 Wyoming school districts provided expenditure and enrolment data by school for four years, representing 80 schools. Sampling: Unclear	Total schools: approx. 80 (exact numbers not given) Total students: NS Range of students per school: 3–1,500 Attrition: NS Data collected over four years Missing: NS	Age 13-14, Grade 8, year 9 Age 16-17, grade 11, year 12 Mixed sex	SES: NS Ethnicity: NS SEN: NS	Funding: All publicly funded Organisational structure: All school types (including elementary) Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
Bradley and Taylor (1998) England	Sampling frame: All published data from the School Performance Tables and the Schools' Census for the years 1992-96 Sampling: All county, voluntary and grant-maintained schools were included. Two types of selective schools were excluded: those selecting students on the basis of special educational needs; and those selecting students on academic ability.	Total schools: 11-16 1992: 1,307 schools 1996: 1,350 schools 11-18 1992: 1,580 schools 1996: 1,514 schools Total students: NS Mean students per 11-16 school: 765 Mean students per 11-18 school: 1,010 Attrition: See below Missing: See below N varies from totals of 2,657 in 1992 and 3,094 in 1996. It is unclear whether this is related to missing data or growth in school numbers.	Age 15-16, grade 10, year 11 Mixed sex	Free school meals: 1992: 10.8% (average per sampled school) 1993: 17.2% 1994: 18.6% 1995: 19.1% 1996: 19.5% In 1996, on average, 10.5% of students in the sampled schools were from ethnic minorities in 1996. SEN (average per sampled school) 1992: 1.3% 1993: 1.5% 1994: 1.9% 1995: 2.1% 1996: 2.4%	Funding: Only publicly funded schools: county schools ranged from 74.7%-65.6% of the total over the five years; voluntary aided schools ranged from 13.1%-10.2%; voluntary controlled 3.8%-3.2%, special agreement 1.0-1.6%; grant maintained 6.8%-20.1%. Organizational structure: 11-16 and 11-18 schools Class size: authors assume that average class size is 30 Student/teacher ratio: 15.33 - 16.22 across the years. Use of streaming: NS
Bradley and Taylor (2003) England	Sampling frame: All published data from the School Performance Tables and the Schools' Census for the years 1992-2002 Sampling: All with available data	Total schools: 3,202 schools in the 1992 data set and 3,098 schools in the 2002 dataset. Total students: approx. 3 million students in 2002 Mean students per school: 1,004 Attrition: 104 schools Missing: Appear to be some missing data; looks like non-response to some questions, rather than complete non-response.	Age 13-14, grade 8, year 9 Age 16-17, grade 11, year 12 Mixed sex School by gender type 2002 (N=3098) 2,686 schools of mixed gender; 185 boys only; 227 girls only	The average percentage of students eligible for free school meals in the schools was 16%. The average percentage of students from ethnic minorities in the schools was 12% in 2002. The percentage of students with special needs in the schools was 16% in 2002.	Funding: Publicly funded schools only Organizational structure: 11-16 and 11-18 schools only Class size: NS Student/teacher ratio: average 16.9 (SD 1.6) Use of streaming: NS

	Sampling		Student characteristics		
Driscoll *et al.* (2003) USA: California	Sampling frame: California Department of Education Schools database Sampling: All schools for which relevant data available	Total schools: 5,525 Total students: NS Mean students per school: 526 (SD: 394) Attrition = See below Missing = 1,205 were excluded because of missing values	Average across grades Mixed sex	Free school meals, mean (%): 38.44, SD: 32.10, Parents who graduated from college: mean (%): 22.56, SD: 14.03 Median household income: mean $36,037, SD: 12,136. Ethnicity: NS SEN: NS	Funding: Unclear Organizational structure: High, Middle & K-12 schools Class size: average 7-12 class is mean, 27.92 with a SD of 3.53. Student/teacher ratio: NS Use of streaming: NS
Fetler (1997) USA: California	Sampling frame: All regular mandated California high schools Sampling: Criterion Alternative high schools were excluded.	Total schools: 805 Total students: approx 1.3 million per year Mean students per school: 1,983 (SD: 853) Attrition: NS Four-year averages Missing: 600 alternative high schools, serving 100,000 students, were excluded.	Ages 12-18, grades 7-12, years 8-13 Mixed sex	% of children in the school who are receiving Federal Aid to Families with Dependent Children Mean: 15.3% (SD 12.3) Ethnicity: Not stated unclear. Authors state, this is reflective of Californian high schools. SEN: NS	Funding: All publicly funded schools Organizational structure: Regular public high schools, mainly grades 9-12 (some 7-12) Class size: NS Student/teacher ratio: Average ratio of K-12 students to classroom teachers in California from 91 to 96 was 23 to 1, (includes elementary schools) Use of streaming: NS
Gill *et al.* (2002) England	Sampling frame: DFES register of schools in England in which 15 year olds were enrolled, in January 1999. Sampling: PISA study method. Multi-stage sampling strategy: criterion, stratified, random	Total schools: 155 Total students: 4,120 Quartiles of school sizes: <428 (25%tile) – >985 (75%tile) Attrition; See missing data below	Age 15-16, grade 10, year 11 Age 16-17, grade 11, year 12 Mixed sex	% of students in LEA schools eligible for free school meals: average 25% Highest level of parental education No qualifications 424 (11%) GCSE or A-level 1,507 (39%) Higher Education 1,920 (50%)	Funding: 91% of students in publicly funded schools, 9% in independent schools Organizational structure: 11-16 & 11-18 schools only Class size: NS

	Sampling		Student characteristics		
		Missing: Response rate of 59% of those schools initially selected; 82%, after including replacements for those schools that refused to take part. Response rate amongst students: 81% .	Female 2048 (50%) Male 2013 (49%)	English second language 181 (4.5%); 6% students from outside the UK SEN: Unclear, but were eligible to take part	Student/teacher ratio: NS Use of streaming: NS
Heck (1993) USA: One Western state	Sampling frame: Not clear Sampling: All elementary, intermediate, and high schools in one western state of the USA All principals and a random sample of between 20%-40% of teachers at each school Two classes of parents were randomly selected in each school.	Total schools: 235 Total students: NS Mean students per school: NS Attrition: See below Missing: Response rates: 75% of principals, 74% of teachers, 45% of parents	Ages 11-18, grade 6-12, years 7-13 Mixed sex	SES: Not reported descriptively Ethnicity: Not reported descriptively SEN: NS	Funding: NS Organizational structure: All school types (inc. elementary) Class size: NS Student/teacher ratio: NS Use of streaming: NS
Howley (1996) USA, West Virginia,	Sampling frame: All West Virginia schools with a third grade, sixth grade, a ninth grade or an eleventh grade. As identified through the Common Core Data, NCES School District Data Book and the West Virginia SEA. Sampling: All with available data	Total schools: 508 (grade 6), 196 (grade 9), 106 (grade 11) Note: These are not mutually exclusive. Students N= NS Students per school N= NS Attrition = See below Missing = Outliers identified through box plots were removed from analyses. Missing data were imputed by taking average scores for similar schools in area.	Age 11-12, grade 6, year 7 Age 14-15, grade 9, year 10 Age 16-17, grade 11, year 12 Mixed sex	SES: Not reported descriptively. Ethnicity: NS SEN: Children with special educational needs are not tested with the CTBS, therefore it can be inferred that the sample did not include children with SEN.	Funding: NS Organizational structure: High, middle & K-12 schools Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
Howley (1999a) USA: Montana	Sampling frame: All schools in Montana Sampling: Criterion based. All those except those with missing data, and those without free schools meals programmes (N=112, average enrolment 36)	Total schools: 220 (grade 8), 168 (grade 11) Note: It is not clear which schools are mutually exclusive (e.g. how many schools supplied both grades 8 and 11 data). Total students: NS Mean students per school: 165 (grades 6-8), 275 (grades 9-12) Attrition: See below Missing: Unclear	Age 13-14, grade 8, year 9 Age 16-17, grade 11, year 12 Mixed sex	SES: NS Ethnicity Within the state 13% of students are from ethnic minorities SEN: NS	Funding: NS Organisational structure: High, middle & K-12 schools Class size: NS Student/teacher ratio: NS Use of streaming: NS
Howley (1999b) USA: Ohio	Sampling frame: Whole state dataset obtained from Ohio state Department of Education Sampling: Criterion everything except those missing data and excluded 'special circumstances' districts (N=5) and schools within these districts from the analyses (e.g. very small districts include those districts not offering high school instruction).	Total schools: 1,314 (grade 6), 811 (grade 9), 650 (grade 12) Note: It is not clear which schools are mutually exclusive (e.g. how many schools supplied both grade 8 and 11 data). Students: NS Mean students per school: 165 (grades 6-8), 275 (grades 9-12) Attrition: See below Missing: Unclear	Age 11-12, grade 6, year 7 Age 14-15, grade 9, year 10 Age 17-18, grade 12, year 13 Mixed sex	SES: Not reported descriptively Ethnicity: 18% of students in Ohio are from ethnic minorities. SEN: NS	Funding: NS Organisational structure: High, Middle & K-12 schools Class size: NS Student/teacher ratio: NS Use of streaming: NS
Johnson *et al.* (2002) USA: Arkansas	Sampling frame: All schools and districts in the state of Arkansas Sampling: Full dataset used but achievement figures for those children with IEP (SEN) or limited English proficiency were excluded from the analyses.	Total schools: 309 (grade 7), 321 (grade 10), 326 (grade 8 literacy), 307 (grade 8 maths). Unclear how many of these are mutually exclusive. 'Roughly' 57 students per grade Attrition: NS	Age 12-13, grade 7, year 8 Age 13-14, grade 8, year 9 Age 15-16, grade 10, year 11	SES: Not reported descriptively Ethnicity: Not reported descriptively SEN: Not stated	Funding: NS Organisational structure: High, middle & K-12 schools Class size: NS Student/teacher ratio: NS

	Sampling		Student characteristics		
		Missing: Unclear	Mixed sex		Use of streaming: NS
Kirjavainen and Loikkanen (1998) Finland	Sampling frame: All senior secondary schools in Finland Sampling: All schools for whom data could be collected	Total schools: 291 Total students: NS Mean students per school: 186 (SD: 99.3); range: 31– 623 Attrition = See below Missing: Of 450 schools in total, data could be obtained for 291.	Age 17-18, grade 12, year 13 Mixed sex	SES: Unclear Ethnicity: NS SEN: NS	Funding: Majority publicly funded. 16 private schools in sample. Organisational structure: High/ secondary school only Class size: Average 19.9 (SD 3.32) range 5.6 - 31.4 Student/teacher ratio: NS Use of streaming: NS
Lee and Burkam (2001) USA	Sampling frame: Students who had been selected to participate in the 1988 high school effectiveness supplement of the National Educational Longitudinal Study (was originally a stratified random sample). Sampling: Criterion urban and suburban areas (rural schools were eliminated) that enrolled at least five of the original cohort. Students were also eliminated where there were missing data on ethnicity, gender and SES as well as test scores, transcript information and dropout status.	Total schools: 190 Total students: 3,840 Range of students: schools with <300– >2,100 students Attrition: NS Missing: Original NELS sample traced 21,126 at eighth grade in approx. 1,000 middle schools. Original NELS sample traced in 1990 17,424 in 1,508 high schools.	Between 15-18 years, grades 8-12, years 9-13 Out of 179 dropouts, 49.4% were female; out of 3,661 in schools, 47.3% were female.	SES: Z-score standardized to a mean of 0 and SD 1 SES mean for drop outs -0.57 (SD) 1.01 SES mean for in school children 0.03 (SD) 0.99 Ethnicity: % Asian in (i) population of dropouts 0.7 (ii) school population 6.1 % Hispanic in (i) dropouts 13.7 (ii) school 11.8 % Black in (i) dropouts 22.6 (ii) school 3.6 SEN: NS	Funding: 131 public, 31 catholic 28 independent Organisational structure: High, Middle & K-12 schools Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
Lee and Smith (1997a) #1 Lee and Smith (1997b) #2 Lee and Smith (1995) #3 USA	Sampling frame: Students who had been selected to participate in the 1988 high school effectiveness supplement of the National Educational Longitudinal Study (was originally a stratified random sample) Sampling criteria 1. There had to be full cognitive test score data on the students across all three waves. 2. There had to be data from their high school and their teachers. 3. Student had to be in public, Catholic or elite private high schools. 4. They must have been attending high schools with at least five NELS sampled students in grade 10. 5. They had to be in the same high school in grades 10 and 12.	Total schools: 789 (#1), 789 (#2), 820 (#3) Total students: 9,812 (#1), 9,631 (#2), 11,794 (#3) Range of schools: schools with <600– >2,500 students Attrition = See below Missing = Original NELS sample 21,126 at eighth grade in approx. 1,000 middle schools Original NELS sample traced in 1990 17,424 in 1,508 high schools	Age 13-14, Grade 8, year 9 Age 15-16, grade 10, year 11 Age 17-18, grade 12, year 13 #1: Average % of females in schools ranged from 47.9% to 52.8%	SES: Z-score standardized to a mean of 0 and SD 1 #1: School <300: -0.12 School 301-600: 0.07 School 601-900: 0.11 School 901-1,200: 0.05 School 1,202-1,500: 0.03 School 1,501-1,800: 0.08 School 1,801-2,100: -0.04 School >2,100: -0.06 Proportion of ethnic minority for schools of that size between 143.3% and 21.5% SEN: NS	Funding: 88.5% publicly funded, 6.5% Catholic, 5% privately funded Organisational structure: High and middle schools Class size: NS Student/teacher ratio: NS Use of streaming: NS
Leung and Ferris (2002) Canada	Sampling frame: Data from the Montreal longitudinal Study that started in 1984 when the subjects were in kindergarten (Tremblay *et al* 1994) and which used inclusion criteria of male; subjects had parents who were born in Canada with French as mother tongue; and came from a family with low SES. Used school size data and other school characteristics from Studies in Education Policy (Crowley and Marceau, 2000) to identify size and characteristics of schools the subjects attended.	Total schools: 107 Total students: 616 Mean students per school: 1,149 Attrition = See below Missing: The Montreal Longitudinal Study began in 1984. The original sample size is not given. In 1995, sample size of Montreal Longitudinal Study was 1,047.	Age 17-18, grade 12, year 13 Males only	SES: Low income: 249 subjects were from schools where the average family income was below 30,000 dollars, and 367 above 30,000. High income: 71 boys were from schools where the average family income was over 50,000 dollars, and 545 below 50,000. Subjects had both parents born in Canada with French as mother tongue. SEN: NS	Funding: NS Organisational structure: NS Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
	Sampling: In 1995, subjects who were 17 yrs old were included who were still participating in the study.				
Ma X (2001) Canada	Sampling frame: Student data from the original New Brunswick School Climate Study. The NBSCS collected data from all grade 6 and grade 8 students but only from the English speaking sector. Sampling: All students	Total schools: 148 (sixth grade), 92 (eighth grade) Total students: 6,883 sixth-graders and 6,868 eighth-graders Students per school N = NS Attrition= See below Missing = Not clear	Age 11-12, Grade 6, Year 7 Age 13-14, Grade 8, year 9 Mixed sex	SES: Estimated based on student's reports of education-related possessions at home, and their participation in social, cultural activities. Authors state this is representative of New Brunswick. Ethnicity: English education sector; those in the French sector were not included. SEN: NS	Funding: Not clear Organisational structure: 'Middle schools' Class size: NS Student/teacher ratio: NS Use of streaming: NS
McLaughlin *et al.* (2000) USA	Sampling frame: Schools and Standards Survey (SASS) dataset (nationally representative sample of schools), Common Core of Data (CCD) dataset, National Assessment of Educational Progress (NAEP) dataset. 2. Publicly available school level state assessment scores. Sampling: Criteria (1) Public schools (2) 39 States participated in the 1994 State NEAP fourth grade reading assessment (3) 23 provided the American Institutes for Research with 1993-94 school-level assessment scores in conjunction with a study of the State NAEP sampling procedures. (4) Scores from three of these states were not useable due to lack of correlation with the NAEP, not being an easily readable medium,	Total schools: 1,123 public elementary schools 496 public middle schools 595 public high schools across 20 states in USA Total students: NS Mean students per school: NS Attrition: See below Missing: The 20 states for which data could be used included 3,785 of the 8,767 SASS public schools. 2,916 of these had students enrolled in grades corresponding to the State assessment and 2,627 were identified as having both SASS and State assessment information. Of these, 66 has no teacher data so the final file included 2,561 school records.	Age 13-14, Grade 8, year 9 Age 16-17, grade 11, year 12 (not explicit) Mixed sex	In the sample of middle school students, 44% (SD 31) were eligible for a lunch programme; whilst in the secondary sample, 28% (SD 25) were enrolled in a lunch programme. In the middle school sample, 34% (SD 32) were ethnic minority students, whilst in the high school sample 30% (SD 31) were ethnic minority students. SEN: NS	Funding: All publicly funded Organisational structure: High and middle schools Class size: Class size: middle 25.4 (SD 7.0) Class size: secondary 23.1 (SD 6.1 Student/teacher ratio: Middle 18.0 (SD 4.9) Student teacher ratio: secondary 16.9 (SD 4.8) Use of streaming: NS

	Sampling		Student characteristics		
	or test usage varying between districts.				
McMillen *et al.* (2000) USA: North Carolina	Sampling frame - All Schools in the state of North Carolina State education databases for the period 1997-98 and 1998-99 Sampling: Criteria All school excepts those excluded from sample on following basis: Alternative, special, ungraded, atypical grade level configurations (i.e. K-12) excluded from all analysis. For analysis of school dropout rates: Only in public high schools serving grades 9-12 (schools with grades 9-12 but not in this configuration excluded)	Total schools: 308 (grades 6-8), 292 (grades 9-12) Total students: NS Mean students per school (years 1996-97): Elementary/middle: 550 Secondary: 956 Attrition = See below Missing = Dropout rate: 273 out of 434 Attainment and violence: Unclear	Outcomes are end of course or end of programme scores; ages not given. Mixed sex	SES: NS Ethnicity: NS SEN: NS	Funding: NS Organisational structure: High, middle & K-12 schools Class size: NS Student/teacher ratio: NS Use of streaming: NS
McNeely *et al.* (2002) USA	Sampling frame: All high schools in the USA that included an 11th grade and at least 30 students. Sampling: Stratified random sample of high schools (basis of stratification not given) A feeder middle school was 'selected' for each high school	Total schools: 127 Total students: 71,515 Mean students per school: 642 (SD: 765.1); range: 42–5,422 Attrition = See below Missing: The total eligible number of schools and students in the sample is not given. 10.9% of sample dropped because of missing responses to the school connectedness scale questions. 'Missing' 10.9% were younger, twice as likely to be old for their grade, had	Age 12-18, grades 7-12, years 8-13 Mixed sex 52% female	SES: NS Ethnicity: % Black - school 14, individual 15 % Latino - school 11.4 ,individual 12.2 SEN: NS	Funding: 83% publicly funded Organisational structure: High, and middle schools Class size: Average class size is 23 Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
		lower grade point average, and were more likely to be male, Latino and/or Black.			
Silins and Mulford (2000) Tasmania and Australia	Sampling frame: Unclear Sampling: Data were from the larger LOLSO project but not clear how selected for that study or for this particular analysis.	Total schools: 96 Total students: 3,500 Total teachers: 2,503 Mean students per school: 631.94 (SD: 283.23) Attrition = See below Missing = Not clear	Age 14-15, grade 9, year 10 Mixed sex	SES: Presented as constructs not easily interpretable Ethnicity : NS SEN: NS	Funding: NS Organisational structure: High school only Class size: NS Student/teacher ratio: NS Use of streaming: NS
Spielhofer *et al.* (2002) England	Sampling frame: DfES Secondary School National Value Added Database 2001, NFER School Database Sampling: All schools	Total schools: 2,954 Total students: 369,341 Range of students per year 11: 25–405 Attrition = 200 students (2.5% of the sample) Missing = See above	Age 13-14, grade 8, year 9 Age 15-16, grade 10, year 11 Mixed sex	Average level of eligibility for free school meals was 16% Ethnicity: NS SEN: NS	Funding: Public schools Organisational structure: 11-16 and 11-18 schools only Class size: NS Student/teacher ratio: NS Use of streaming: NS
Stiefel *et al.* (2000) USA: New York	Sampling frame: Board of education database of school-based budgets, cohort graduates per high school and high school report cards Sampling: Criteria Whole dataset except: 1. Those that were classified as 'programmes' as these frequently do not serve all grades or served	Total schools: 121 Total students: NS Mean students per school: 2,030 (SD: 1,191) Attrition = See below Missing = A total of 201 entries in the database of which 121 met criteria	No student outcome - predicted budget cost per graduate Mixed sex	Of the sample: 45.4% were eligible for free school meals (SD 21.8) 13.4% had limited English proficiency (SD 11.6%) 5.4% were registered for part-time special education (SD 4.8).	Funding: All publicly funded Organisational structure: High schools only Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
	very specialised populations. 2. Those schools that did not serve all grades. 3. Those that did not provide data across all the variables required for the study.	for inclusion in the sample.			
Taylor and Bradley (2000) England	Sampling frame: DfES – Annual Schools performance tables, Annual Schools Census and Funding Agency for Schools Sampling: All secondary schools in England (except for those designed specifically for students with special needs)	Total schools 1993: 2,034 1995: 3,030 1996: 3,014 1997: 3,087 Students: NS Mean students per school: 912 Attrition = See below Missing: Numbers used in each analysis differ so it is inferred that there may be some missing data and exclusions as a result of this.	No student outcomes Mixed and single sex schools used (mean values) between 6% and 7% of schools were boys only; 7% were girls only; and between 86 and 87% were mixed.	Mean values for students on free school meals (%) between 17.2 and 19.5 Mean values for students with English as second language (%) 1996: 7.2 1997: 7.4 Mean values for students with special educational needs within all other schools (%) 1993: 1.5 1997: 2.5	Funding: All publicly funded Organisational structure: 11-16 and 11-18 school only Class size: NS Student/teacher ratio: NS Use of streaming: NS
Welsh *et al.* (1999) USA: Philadelphia	Sampling frame: All public middle schools in Philadelphia Sampling: Schools were chosen in consultation with Philadelphia School district officials in the basis of four criteria: 1. Schools were chosen that had both high, medium and low levels of disorder. 2. Schools were chosen on the basis of their levels of poverty to represent high, medium and low levels. 3. Selection was conditioned by regional representation with	Total schools: 11 Total students: 6,693 Mean students per school: 966 (SD: 218); range: range 605–1,288 Attrition = See below Missing = 57% response rate (after exclusions)	Ages 11-14, grades 6-8, years 7-9 Mixed sex	SES: % of children in the school who were receiving Federal Aid to Families with Dependent Children: mean 77.78 (SD 15.64) with a range of 48.2-93.8 Ethnicity: Majority were non-white Mean 0.78 with SD of 0.41 where 0 = white and 1 = non-white SEN: NS	Funding: All publicly funded Organisational structure: Middle schools only Class size: NS Student/teacher ratio: NS Use of streaming: NS

	Sampling		Student characteristics		
	schools included across the entire city gaining representation of 6 of 7 districts. 4. The principal had to be willing to take part in the survey and participate in an interview.				
Welsh *et al.* (2000) USA: Philadelphia	Sampling frame: The population of middle schools in Philadelphia Sampling: All with data	Total schools: 43 Total students: N = approximately 34,000 Mean students per school: 817.24 (SD: 305.20); range: 223-1,288 Attrition = See below Missing: Some missing data; for example, between 41 and 43 schools provided data to calculate the community stability, school stability and community poverty scales.	Ages 11-14, grades 6-8, years 7-9 Mixed sex	Local community (i.e. that the schools were in): The median household income across the schools was $23,146 with a SD of $8,719 and a range of $9,275 to $46,608. Imported community (i.e. that where the students actually resided): The median household income across the schools was $20,521 with a SD of $6,191 and a range of $10,559 to $33,647. 69% of enrolments were African-American, 2% Asian, 11% Latino and 10% White. SEN: NS	Funding: All publicly funded Organisational structure: Middle schools only Class size: NS Student/teacher ratio: NS Use of streaming: NS

NS; Not stated; NA; Not applicable

Appendix 4.4: Details of results of the studies in the in-depth review

Study	Variables	Results	Conclusions
Abbott *et al. (*2002) Hierarchical linear modelling School: N=417 District N=255 R^2= NR	**Dependent variable** 4 and 7 grade student Washington assessment of student learning (WASL) scale scores in reading and maths aggregated to the school level **Explanatory variables** (i) School size as measured by span size (e.g. number of students divided by number of grades), (ii) percentage receiving free and reduced lunch as a measure of SES, (iii) district size as measured by the total district enrolment School level averages were nested in district averages to investigate the interactions between schools and districts.	Model 1 (school size, district poverty and the product of school size and district poverty) 1. School size 2. Interaction Unstandardized regression coefficient (SE) Grade 7: Maths 1. 0.04 (0.04) 2. –0.25 (0.23) Grade 7: English 1. 0.01 (0.01) 2. –0.05 (0.07)	**Reviewers' interpretations** No significant associations were found for the 'effect' of school size on achievement (grade 7 maths and reading achievement), or for the interaction between school size and district poverty. An increase in *span size* of 100 is associated with an increase in grade 7 maths achievement of four marks, and grade 7 English achievement of one mark (NS).
Atkinson and Wilson (2003) Multiple regression analysis KS3 mean N=484,569, R^2=0.18 English KS3 N=471,209, R^2=0.15 Math KS3 N=472502, R^2=0.16 Science KS3 N=469,776, R^2=0.16 GCSE points N=476,403, R^2=0.17 Maths GCSE N=470,877, $R^2$0.14 Science GCSE N=468,933, R^2=0.13 Value added N=476,403, R^2=0.05 5 GCSE C or above N=484,569, R^2=NR	**Dependent variables** (i) KS3 mean and subject level, result (Maths, English, Science) (range 1-9 average 5), (ii) GCSE total points and subject level (Maths, English, Science) (range 0 (ungraded) -8 (A*) for each GCSE), (iii) value added scores (between KS3 and GCSE) at student level, (iv) 5 GCSEs A* to C **Independent variables** (i) School size/10,000 and school size^2/100,000, (ii) gender, (iii) school type (funding status and admissions policy), (iv) age (within-year variable to identify older and younger students), (v) free school meals, (vi) single sex establishment, (vii) religious denomination	Unstandardized coefficients (robust t statistic) 1. Values for size^2/100,000 2. Values for size/10,000 Overall KS3 Mean: (1) -0.02 (5.54**), (2) 5.12 (7.98**) English KS3: (1) -0.03 (5.62**), (2) 7.69 (7.36**) Maths KS3: (1) -0.01 (3.97**), (2) 4.31 (6.26**) Science KS3: (1) -0.01 (2.42*), (2) 3.00 (4.41**) Total GCSE: (1) -0.28 (4.76**), (2) 92.66 (7.11**) English GCSE: (1) -0.02 (4.93**), (2) 7.75 (7.05**) Maths GCSE: (1) -0.03 (3.83**), (2) 8.58 (5.91**) Science GCSE: (1) -0.02(3.00**), (2) 5.95 (4.62**) Value added: (1) -0.11 (2.40*), (2) 33.33 (3.40**) 5 GCSEs at Grade C or above (1) -0.02 (5.58**) (2) 5.93 (7.95**).	**Authors report** The significance of the size and size^2 terms 'confirms the non-linearity of the relationship between all the dependent variables and school size' (p 8). 'The optimal size of school ranges from 1280-2155' (dependent on subject) (p 8).

Study	Variables	Results	Conclusions
Bedard *et al.* (1999) Standard and modified ordered probit Maths Middle: N=801 High: N=618 Reading Middle: N=800 High: N=598 Language: Middle: N=805 High: N=606 Science High: N=621 Social science: High: N=623 Spelling: Middle: N=803 R^2= NR	**Dependent variable** The distribution of test scores (maths, reading, language, social science, science and spelling) within the schools in terms of whether the schools distribution is poor, average or good, and across grades 3, 8 and 10. **Explanatory variables** (i) School size (both as average enrolment per grade and then as total school size), (ii) student-teacher ratio, (iii) % Hispanic, Asian, Pacific, Philipinno, Native American, (iv) computers per student, (v) % families on aid for dependent children, (vi) % eligible for free school meals, (vii) % limited English proficiency, (viii) year-round school, (iv) Urban, Suburban. Sub-analyses were completed but are not presented here.	Standard ordered probit Unstandardized probit coefficient (SE) Maths: Middle –0.0002 (0.0006) High –0.0008 (0.0005) Modified ordered probit Unstandardized probit coefficient (SE) Maths: Middle: 0.0005 (0.0006) High (0.0012 (0.0005) ~ Reading: Middle: 0.0013 (.0006)~ High: 0.0015 (.0007)~ Language: Middle: 0.0015 (.0006)~ High: 0.0004 (0.0005) Science: High: 0.0009 (.0005)~ Social science: High: 0.0005 (.0004) Spelling: Middle: 0.0016 (0.0006)~	Probit coefficients are not easily interpretable; the coefficient is how much difference a unit change in the independent (e.g. size of school) makes in terms of the cumulative normal probability of the dependent variable (e.g. of being categorised as poor, average or good). **Reviewers' interpretations** School size is not statistically significant as an explanatory variable when a standard ordered probit model is used. The modified ordered probit models for high schools indicate that smaller schools are less likely to be in the 'poor performance' category than larger schools. This result holds for Maths, Reading and Science, but only at a $p<0.1$ level. At the middle school level, the result holds for Reading, Language and Spelling, but only at the $p<0.1$ level.
Bickel and Howley (2000) Multilevel model using ordinary least squares methods Grade 8 N=367 schools N=158 districts Grade 11 N=298 schools N=155 districts Grade 8 School level variance 26% Group level 31% Interactions 22%	**Dependent variables** (i) Iowa test of basic skills at grade 8 mean school-level percentile scores for seven subtests a composite score, (ii) Georgia high school graduation test (grade 11) school-level % of students passing the first administration of the GHSG. **Explanatory variables** (i) Number of students per grade level in 1,000-student units, (ii) % students eligible for free and reduced cost lunch, (iii) % students black, (iv) % students other racial or ethnic minority groups, (v) student-teacher ratio as proxy for class size. Measured at both school and district level with school level variables nested in district level variables and interaction terms considering interaction between levels	Results presented here are only for the model with a greater number of variables specified. Single level model with only size, SES and interaction show significant positive 'effects' for span size on achievement and significant negative 'effects' for the interaction between size and SES. Weighted regression results Unstandardized (standardized) regression coefficients Interaction 1 refers to school span size and district span size. Interaction 2 refers to school span size and school SES. Interaction 3 refers to school span size and district SES. Grade 8: composite	**Reviewers' interpretations** An increase in *span size* of 1,000 is associated with a decrease in mean school composite achievement (grade 8) by six percentile scores (NS). An increase in *span size* of 1,000 is associated with an increase in mean school composite achievement (grade 11) by three percentile scores (NS). This can be compared with the model of lower specification (1999 paper) which shows that: an increase in *span size* of 1,000 is associated with an increase in mean school composite achievement (grade 8) of 24 percentile scores ($p<0.01$); an increase in *span size* of 1000 is associated

Study	Variables	Results	Conclusions
Grade 11 School level variance 29% Group level 10% Interactions 10%	**Interaction terms** 1. School SPANSIZE and school FREEPCT 2. District SPANSIZE and district FREEPCT 3. District SPANSIZE and school FREEPCT 4. School SPANSIZE and district FREEPCT	Individual level -6.401 (-.050) Interaction1 –308.619**(-.167) Interaction2 0.141 (.024) Interaction3 –1.046*** (-.237) Grade 11: composite Individual level 3.688 (.027) Interaction1 –133.985 (-.100) Interaction2 0.281 (-.075) Interaction3 –1.357*** (-.456)	with an increase in mean school composite achievement (grade 11) of 36 percentile scores (p<0.001).
Bickel *et al.* (2001) Regression analyses N=1,001 Reading R^2=.40 Math R^2=.30 Writing R^2=.40 Expenditure1 R^2=.51 Expenditure2 R^2=.53 Composite R^2=.43 R^2= adjusted	**Dependent variables** (i) Achievement; measured at grade 10 from reading, writing, maths, and as a composite score of the Texas Assessment of Academic Skills, (ii) expenditure per student (dollars) **Explanatory variables** (i) Size: expressed in thousand student units, or in logarithms of single student units, (ii) poverty: percentage of students eligible for free or reduced cost lunch, (iii) ethnicity: percentage black, and percentage Hispanic, (iv) English proficiency: percentage of students classified as having limited English proficiency, (v) student to teacher ratio, (vi) expenditure per student (only in achievement analyses), (vii) percentage of total budget allotted for instruction, (viii) percentage of students enrolled in a full-time career and technical education programme, (iv) percentage of students enrolled in a full-time special education programme, (v) percentage of students classified as gifted, (vi) single unit schools/number of grade levels, (vii) number of high schools in a district.	Unstandardized (standardized) coefficients Size = direct effects of school size Size x = interaction effects school size and SES Reading: Size: 0.177 (.065) Reading: SES x Size: -0.035 (-0.143) ** Maths: Size: 0.019 (.040) Maths: SES x Size: -0.060 (-0.144)** Writing: Size: 0.052 (.025) Writing: SES x Size: -0.033 (-0.171)*** 1. Expenditure: Size: -254.415 (-0.199)*** Expenditure: Unit x Size: -730.195 (-0.172)*** 2. Expenditure: Size: -290.519 (-0.227)*** Expenditure: High schools x Size: -114.038 (-0.076) Expenditure: Levels x Size: -48.445 (-0.108)** Composite: Size: 0.218 (.079) Composite: High schools x size: 0.534 (0.166)*** Composite: Levels x Size: 0.050 (.051) Composite: Size x SES: -0.034 (-0.116)*	**Reviewers' interpretations** No significant 'effects' of school size on achievement are reported for grade 10. An increase in *span size* of 1,000 is associated with an increase in reading, maths and writing scores of 0.2, 0.02, 0.05 respectively (NS). An increase in *span size* of 1,000 is associated with an increase in composite achievement score of 0.2 (NS). An increase in *span size* of 100% is associated with a decrease in expenditure per student of 254.4 dollars (model 1) (p<0.001). An increase in *span size* of 100% is associated with a decrease in expenditure per student of 290.5 dollars (model 2) (p<0.001).
Bowen *et al.* (2000) Bivariate analysis of variance N=945 Students R^2=NA	**Dependent variables** (i) School satisfaction: five-item dichotomous scale, (ii) teacher support: eight-item dichotomous scale, (iii) school safety: ten-item scale; responses: three-item Likert scale **Explanatory variables** (i) School size = total number of students enrolled in the school, (ii) gender, (iii) race/ethnicity, (iv) poverty % eligible for FSM	No numerical data are presented for the analyses completed for school size and SES.	**Reviewers' interpretation** It is inferred that no significant interactions were found between school size and SES for measures of school satisfaction, teacher support and school safety.

Study	Variables	Results	Conclusions
Bowles and Bosworth (2002) Least squares dummy variables Least squares dummy variables with time effects Simultaneous equation modelling N=80 1. R^2=.77 2. R^2=.95 3. R^2=.95 4. R^2=.89 5. R^2=.89 6. R^2=.48	**Dependent variable** (i) Per student cost average across four years, (ii) average student test scores **Explanatory variables** (i) School size, (ii) average teacher salary in the district for each type of school, elementary, middle and high, (iii) Percentage of children not eligible for federal free or reduced lunch, (iv) standardized test results mean figure across reading, writing and maths **Dummy variables** (i) School type (elementary, middle and high school), (ii) district school is in, (iii) time period of data-collection	Unstandardized coefficients 1. LSDV –0.1998 2. LSDV corrected for heteroscedaticity –0.2046 3. LSDV with time effects with correction –0.2013 4. Pooled model with correction –0.2316 5. Simultaneous equations –0.2052 The above a reported as being statistically significant 6. Student performance -1.0900 The above reported as not being statistically significant	**Reviewers' interpretations** Across all models, an increase in school size of 10% is associated with a decrease in student costs of approximately 2% (reported as statistically significant). No statistically significant association was found between school size and test scores; an increase in school size of 100% is associated with a decrease in mean student test score of 1. Interpretation of this is not straightforward as there is no differentiation between age groups (e.g. elementary, middle and high school age).
Bradley and Taylor (1998) Ordered logit model 11-16 1992 N=1,307 Count R^2= 0.58 Pseudo R^2= 0.32 11-16 1996 N=1,350 Count R^2= 0.61 Pseudo R^2= 0.35 11-18 1992 N=1,580 Count R^2= 0.59 Pseudo R^2= 0.32 11-18 1996 N=1,514 Count R^2= 0.62 Pseudo R^2= 0.38	Dependent variable: (i) Exam performance (GCSEs A*-C) of all secondary schools in England covering the period 92-96, (ii) change in exam performance **Explanatory variables** (i) Number of students per school 11-16 and 11-18, (ii) type of school, (iii) Selection policy, (iv) school gender; single sex establishment, (v) school age range and course types for post-16 students, (vi) student- teacher ratio, (vii) staff hours per student, (viii) % of qualified teaching staff, (ix) ratio of part-time to full-time teachers, (x) support staff hours per student, (xi) student characteristics: % with SEN, % on FSM, % ESL, % ethnic minority Two models: 11-16 schools and 11-18 schools	Ordered logit model Unstandardized coefficients (SE) 1. Size/100 2. Size/100^2 Exam Performance 11-16 schools (1992) 0.55 (4.16)*** –0.024 (-3.11)* 11-16 schools (1996) 0.38 (3.43)*** –0.015 (-2.55)* 11-18 schools (1992) 0.55 (5.72)*** –0.021 (-4.62)*** 11-18 schools (1996) 0.56 (5.70)*** –0.019 (-4.34)*** Authors calculate size at which exam performance is maximised: For 11-16 schools range is 1090-1230 For 11-18 school range is 1600-1350	Logit coefficients are not easily interpretable. Marginal effects are not calculated by the authors for school size variable. **Authors' report** 'The positive coefficient on the number of pupils and the negative coefficient on the square of this variable indicate a non-linear relationship: exam performance rises as school size increases at a decreasing rate as school size increase, but at some point the relationship is reversed, thus implying the existence of a school size at which exam performance is maximized' (p 310). 'The relationship (between school size and exam performance) is 'flat topped' for both groups of schools, an increase in school size above 900 (but under 1,500) pupils has very little effect on exam performance; and the same is true for 11-18 schools with above 1,200 (but under 1,800) pupils' (p 316).

Study	Variables	Results	Conclusions
Ordinary least squares regression N=2,881 R^2=0.06 (model 1), 0.08 (model 2) Note R^2 not the same a OLS R^2		**Ordinary least squares regression** Unstandardized coefficients (SE) Change in exam performance Model 1 0.007 (4.41)*** –0.0006 (-1.64) Model 2 0.007 (4.13)*** –0.0007 (-1.89)	'The estimated coefficient on the change in school size indicates that an increase in school size by 100 pupils is associated with an increase of 0.7 percentage points in exam performance' (p 316).
Bradley and Taylor (2003) Regression analyses N=3,098 1. R^2= 0.78 2. R^2=0.87 3. R^2=0.34 4. R^2=0.34 5. R^2=0.53 6. R^2=0.15	**Dependent variables:** (i) Value added KS3 to KS4, (ii) Average score in KS4 exams (GCSE/GNVQ - year 11), (iii) Change in school exam performance (based on proportion of students receiving 5 or more GCSE results A*-C) 1993-2002, (iv) Truancy rate (%) and absence rate (%). **Explanatory variables:** (i) School size (number of students) and school size squared, (ii) Other school characteristics: presence of 6th form; ratio f/t:p/t teachers; student/teacher ratio, (iii) Initial attainment (estimated by KS3 test scores), (iv) School governance and selection policy: type and coeducational policy, (v) School specialism, (vi) Student characteristics incl. % with special needs; % with authorised and unauthorised absence; % eligible for school meals; % ethnic minority, (vii) Competition including scores for other local schools.	Unstandardized regression coefficients (SE) 1. KS4 results: Size: 0.569 (.093)*** Size^2: -0.017 (.004)*** 2. KS4 results including score for initial achievement: Size: 0.170 (0.078)* Size^2: -0.005 (0.003) 3. Value added KS3-KS4 Size: 0.143 (0.067)* Size^2: -0.005 (0.003) 4. Truancy rate Size: -0.043 (0.032) Size^2: 0.002 (0.001) 5. Overall absence rate Size: -0.198 (0.054)*** Size^2: 0.007 (0.002)** 6. Change in exam performance 1993-2002 Size: 0.010 (0.001)***	**Authors' report** 'An increase of 100 pupils is associated with an increase in exam performance of 1 percentage point' (p<0.001) (p 27).
Driscoll *et al.* (2003) Instrumental variable estimation technique Middle: N=753 R^2=0.68 High: N=747 R^2=0.71	**Dependent variables** (i) 1999 California academic performance index (API) based on a weighted average of SAT **Explanatory variables** (i) School size, (ii) class size: K-3 average, 4-6 average 7-12 average, (iii) district size, (iv) SES: free or reduced price meals, (v) parental income: median household income tied to the school zip code, (vi) percentage of schools' parents who have graduated from college, (vii) population density: per acre by city, town or place name, (viii) share of	1. Unstandardized regression coefficients 2. Standardized coefficients Middle school: 1. –0.0021 2. –0.0073 High school: 1. –0.0027 2. –0.015	**Reviewers' interpretations** No significant associations between school size and achievement An increase in school size of 100 was associated with a decrease in middle school API of 0.21 (NS), and a decrease in high school API of 0.27 (NS)

Study	Variables	Results	Conclusions
R^2= Adjusted	children enrolled in private schools.		
Fetler (1997) Stepwise backwards multiple regression N=805 R^2=0.50	**Dependent variable** (i) School average dropout rate **Explanatory variables:** (i) Enrolment, (ii) annual dropout rate (iii) % of students covered by Federal Aid to Families of Dependent Children (AFDC), (iv) location (e.g. urban, rural), (v) annual growth in the number of employed teachers, (vi) % of new first-time teachers, (vii) average years of experience, (viii) % of teachers with only a Bachelors degree, (ix) average number of teacher years of education	Parameter estimate 0.001 (p<0.001) Standardized Beta 1.6	**Reviewers' interpretations** An increase in 1,000 students is associated with an increase in dropout rate (percentage of students leaving the courses over a given year) of 1% (sig p<0.001).
Gill *et al.* (2002) Multilevel modelling N=4,120 Students R^2=NR	**Dependent variable** Reading literacy: The only outcome that is reported by school size is reading literacy. **Explanatory variables:** (i) School size, (ii) gender, (iii) year group, (iv) language spoken at home, (v) country of birth, (vi) socio-economic status, (vii) level of parental education, (viii) number of parents at home, (ix) birth order, (x) number of siblings, (xi) whether school is maintained by LEA or independent, (xii) coeducational policy	No numerical data presented for school size	**Authors' report** 'School size was one of three factors initially considered in the model that were found not to be associated with reading literacy, because other factors with which they were highly correlated were more important' (p 77). 'Two other factors were excluded from the model: class size and percentage eligible for FSM as data were not available for private schools' (p 77).
Heck (1993) Stepwise forced entry regression analyses N=235/174 Math R^2=0.56 Reading R^2=0.58 Attendance R^2=0.45 Suspensions R^2=0.41	**Dependent variables** (i) Reading and maths attainment scores on Stanford achievement test, (ii) average daily attendance, % (iii) number of student suspensions for significant offences. **Explanatory variables** (i) School size, (ii) other school characteristics: i. % students reduced/ free lunch, ii. % students limited English-language speakers, iii. grade level, iv. number of principals in last five years, v. percentage of special education students, (iii) school academic indicators: data collected using an 18-item questionnaire sent to parents, teachers, administrators, (iv) level of teacher	Standardized Beta coefficients Maths -0.12* Reading -0.16** Attendance -0.28** Suspensions -0.03	**Reviewers' interpretations** An increase in school size of one standard deviation unit is associated with a decrease in 0.12 standard deviation units of SAT achievement. An increase in school size of one standard deviation unit is associated with a decrease in SAT score of 0.16 standard deviation units. An increase in school size of one standard deviation unit is associated with a decrease in daily attendance of 0.28 standard deviation units.

Study	Variables	Results	Conclusions
R^2=adjusted	professionalism (experience)		
Howley (1996) Backwards Regression analysis Grade 6 N=508 R^2=0.17 Grade 9 N=196 R^2=0.13 Grade 11 N=106 R^2=0.09	**Dependent variable** Achievement in grades 3, 6, 9 and 11 in the comprehensive test of basic skills **Explanatory variables:** (i) Schools size: autumn enrolment in the grade level cohort that is the subject of the analysis, (ii) SES as measured as the proportion of students at each school receiving free or reduced lunches in autumn 1990, (iii) interaction: the product of the two **Ancillary analysis** District size: total enrolment in autumn 1990 of the subject cohort in the county district (reported in second paper)	1. Unstandardized regression coefficients (SE) 2. Standardized coefficients Grade 6 interaction: 1. –0.094 (0.04)* 2. –0.10 Grade 9 interaction: 1. –0.072 (0.03)* 2. –0.15 Grade 11 enrolment: 1. 0.026 (0.01)* 2. 0.33 Grade 11 interaction: 1. -0.084 (0.04)* 2. –0.21	**Reviewers' interpretations** An increase in Grade 11 *span size* of 100 is associated with an increase in achievement of 2.6 marks on the comprehensive test of basic skills (sig p<0.05).
Howley (1999a) Regression analyses Grade 8 N=220 R^2=0.18 Grade 11 N=168 R^2=0.30 R^2= adjusted	**Dependent variable** (i) Student tests at grades 4, 8 and 11, using comprehensive test of basic skills, Iowa test of basic skills (ITBS) and the Stanford achievement test. (Note: not everybody uses the same test) **Explanatory variables** (i) Size (computed as grade cohort enrolment, logged), (ii) rates of free and reduced price meals as an indicator of SES, (iii) the product of the two **Sub-analyses** completed on the most common tests; ITBS at grade 8 and the Stanford at grade 11 **Additional analyses** were completed using the following as explanatory variables: 1. District size 2. Class size 3. Proportion of Indian Americans	1. unstandardized regression coefficients (SE) 2. standardized coefficients Grade 8 enrolment: 1. .061 (.442) 2. .009 Grade 8 interaction: 1. –0.380 (.205) 2. –0.128 Grade 11 enrolment: 1. –1.123 (.612) 2. -.161 Grade 11 interaction: 1. –0.034 (.308) 2. –0.010 Sub-analyses Grade 8 enrolment: 1. –0.796 (.539) 2. –0.153 Grade 8 interaction: 1. –0.845 (1.675)** 2. –0.766 Grade 11 enrolment 1. –5.374 (1.675)** 2. –0.766 Grade 11 interaction 1. –1.467 (0.728) 2. –0.464	**Reviewers' interpretations** No statistically significant 'effects' were found between school size and achievement or between the product of size and SES and achievement in the overall analyses. Overall: An increase in *span size* of 100% is associated with an increase in grade 8 achievement of .06 marks (NS), and a decrease in grade 11 achievement of 1.1 marks (NS). In sub-analyses: An increase in *span size* of 100% is associated with a decrease in grade 11 achievement of 5.4 SAT marks (sig p<.01). An increase in *span size* of 100% is associated with a decrease in grade 8 achievement of 0.8 marks in the ITBS (NS).

Study	Variables	Results	Conclusions
Howley (1999b) Regression analyses PASS Grade 6 N=1,314 R^2=0.47 Grade 9 N=811 R^2=0.43 Grade 12 N=650 R^2=0.37 Advanced PASS Grade 6 N=1,309 R^2=0.42 Grade 12 N=650 R^2=0.46 R^2= adjusted	**Dependent Variable:** (i) Ohio proficiency tests in reading, writing, maths, citizenship and science with scores aggregated as percent passing. Tests taken at grades 4, 6, 9 and 12. Both pass rates and advanced pass rates for grades 6 and 12 are presented. **Explanatory variables** (i) Size (computed as grade cohort enrolment), (ii) aid to dependent children (as measure of SES), (iii) the product of the two **Additional analyses** were completed using as explanatory variables: 1. District size 2. Class size 3. Location; rural or small town 4. Proportion of African American students	1. Unstandardized regression coefficients (SE) 2. Standardized coefficients PASS Grade 6 enrolment: 1. –0.869 (0.514) 2. –0.034 Grade 6 interaction: 1. –2.723 (0.419)*** 2. –0.132 Grade 9 enrolment: 1. –4.033 (0.613)*** 2. –0.183 Grade 9 interaction: 1. –2.680 (0.480)*** 2. –0.184 Grade 12 enrolment: 1. 2.065 (0.614)*** 2. 0.114 Grade 12 interaction 1. –0.915 (0.482) 2. –0.073 ADVANCED PASS Grade 6 enrolment: 1. 0.142 (0.037)*** 2. 0.081 Grade 6 interaction: 1. –0.153 (0.030)*** 2. –0.109 Grade 12 enrolment: 1. 0.223 (0.047)*** 2. 0.148 Grade 12 interaction 1. –0.162 (0.037)*** 2. –0.156	**Reviewers' interpretations** An increase in *span size* of 100% is associated with a decrease in percent passing (at grade 6) of 0.87 (NS) An increase in *span size* of 100% is associated with a decrease in percent passing (at grade 9) of 4.0 (p<0.001) An increase in *span size* of 100% is associated with an increase in percent passing (at grade 12) of 2.1 (p<0.001) An increase in *span size* of 100% is associated with an increase in the percentage receiving advanced pass marks (at grade 6) of 0.14 (p<0.001) An increase in *span size* of 100% is associated with an increase in the percentage receiving advanced pass marks (at grade 12) of 0.22 (p<0.001)
Johnson *et al.* (2002) Multiple regression analyses R^2 Grade7=0.34 Grade10=0.37 Grade 8 lit=0.10, math=0.19 Grade 7 N=309 Grade 10 N=321 Grade 8 lit N=318	**Dependent variable** (i) School and district level achievement at grades 5, 7 and 10 (Stanford achievement test (SAT)) and grades 4 and 8 (Benchmark maths and literacy tests) **Explanatory variables** (i) School size was the ratio of total school enrolment to number of grade levels (logged), (ii) SES was that proportion of children receiving subsidised meals. (iii) Product of SES and size (logged) as an interaction term.	(a) = Unstandardized (SE) (b) = Standardized (i) Size Grade 7: (a) –0.444 (0.453) (b) –0.047 Grade 10: (a) –2.26 (0.510) (b) –0.274*** Grade 8 lit: (a) –1.508 (.816) (b) –0.106 Grade 8 math: (a) –1.100 (.576) (b) –0.108 (ii) Size x SES Grade 7: (a) –0.062 (0.024) (b) –0.122* Grade 10: (a) –0.067 (0.020) (b) –0.180*** Grade 8 lit: (a) –0.107 (0.043) (b) –0.138** Grade 8 math: (a) –0.131 (0.030) (b) –0.242***	**Reviewers, interpretations** An increase in *span size* of 100% is associated with a decrease of 0.44 (NS) (grade 7) and 2.26 (p<0.001) (grade 10) marks in the SAT tests. An increase in *span size* of 100% is associated with a decrease of 1.5 (NS) (literature) and 1.1 (NS) (maths) marks in the benchmark test.

Study	Variables	Results	Conclusions
Grade 8 maths N=299			
Kirjavainen and Loikkanen (1998) Tobit model N=291 3A R^2=0.36 3B R^2=0.36 3C R^2=0.36 4A R^2=0.25 4B R^2=0.23 4C R^2=0.23	**Dependent variable** Inefficiency score; based on efficiency score calculated in stage 1 (maximum efficiency is 100). **Explanatory variables** (i) Class size, class size^2, (ii) school size, school size^2, (iii) share of state grants: taken from 1992, (iv) proportion of private schools, (v) proportion of female students, (vi) heterogeneity of student body, (vii) municipality, (viii) parents' education Six different Tobit models are calculated, based on the different efficiency scores obtained from using different variables to calculate the efficiency score.	Parameter estimates (normalized coefficients) 1. school size 2. school size^2 Tobit 3A 1. 0.0006 2. –0.000003 Tobit 3B 1. 0.0016 2. -0.000004 Tobit 3C 1. 0.0016 2. –0.000004 Tobit 4A 1. 0.00009 2. –0.000002 Tobit 4B 1. 0.0023 2. –0.000004 Tobit 4C 1. 0.0023 2. –0.000005	The parameter estimates resulting from a Tobit regression are not easily interpretable. Coefficients need to be decomposed into two parts: the 'effect' of the probability being above zero, and the 'effect' on the mean given that it is above zero. **Authors' report** No statistically significant associations were found between school size and inefficiency in any of the six tobit models (p 391).
Lee and Smith (1997) Hierarchical linear model N=9,812 students N=789 schools	**Dependent variables** (i) Achievement gains over 8th-12th grade in maths and reading. (ii) Academic engagement from grades 8 to 10 **Explanatory variables** **Student level:** (i) Ethnicity, (ii) gender, (iii) ability (composite measure of achievement at 8th grade), (iv) SES **School level:** (i) Enrolment size, (ii) SES, (iii) minority concentration, (iv) sector (public, Catholic, elite private) **Additional analysis** SES/gains slope to investigate equitable distribution of learning	All 'effects' are presented in a standardized effect size metric, weighted with size as a categorical variable. Reference category is a school of size 1,200-1,500 students (i) maths (ii) reading <300 (i) –0.931*** (ii) –0.532* 301-600 (i) –0.089 (ii) 0.149 601-900 (i) 1.512*** (ii) 0.539* 901-1200 (i) 0.589*** (ii) 0.290 1501-1800 (i) –0.152 (ii) –0.254 1801-2100 (i) –0.415** (ii) –0.455* >2100 (i) –1.842*** (ii) –0.911*** **SES/gains slope** <300 (i) –1.187 (ii) –2.161 301-600 (i) –0.985*** (ii) –3.153* 601-900 (i) –0.667 (ii) –2.156* 901-1,200 (i) –0.123 (ii) –0.487 1,501-1,800 (i) 0.984** (ii) 2.115* 1,801-2,100 (i) 1.481*** (ii) 3.795** > 2,100 (i) 1.264** (ii) 3.876**	**Reviewers' interpretations** Schools with fewer than 300, between 1,801 and 2,100, and greater than 2,100 students are associated with significantly lower achievement gains in maths (0.9, 0.4, 1.8SD) and reading (0.5, 0.4, 0.9SD) than schools with 1,200-1,500 students. No significant differences in achievement gains were found in schools with between 301 and 600 or with 1,501-1,800 students compared with schools with 1,200-1,500 students. Schools with between 601 and 900 students are associated with increased achievement gains in maths (1.5SD) and reading (0.5SD) when compared with schools with 1,200-1,500 students. Schools with between 901 and 1,200 students are associated with increased achievement gains in maths (0.5SD) when compared with schools with 1,200-1,500 students.

Study	Variables	Results	Conclusions
		Academic engagement -0.19* Interaction effect based on SES/gain slope 0.10	**Authors' report** Equity effects: 'In virtually all schools the relationship between SES and achievement growth is positive; higher SES students learn more. Thus by definition, negative effect sizes are more equitable, as they reflect a decreased relationship between SES and learning…learning is distributed more equitably in smaller schools' (p 213).
Lee and Burkam (2001) Hierarchical linear modelling For dichotomous data N= 3,840 students in 190 schools Variance explained = NR	**Dependent variable** Students drop out between 10th and 12th grade. (log odds) **Explanatory variables** **Individual level:** (i) Gender, (ii) ethnicity, (iii) SES, (iv) over age in year, (v) academic maths courses, (vi) maths achievement grade 10, (vii) maths GPA grades 9 and 10. **School level:** (i) School structure; size, (ii) school structure: sector, (iii) school demographic composition; (a) school SES, (b) high minority enrolment, (c) average achievement, (d) average 9th grade GPA, (iv) academic organisation; (a) offering calculus, (b) number of courses offered below algebra 1, (v) social organisation; school-based social capital (student questionnaire) **Interaction effects** of student teacher relations by type school and size of school are also included in the school-level model.	Log-odds of dropping out. Reference category is set as medium school: 601-1,500 students Small 0.75 (p<0.10) Large 1.32 (p<0.001) Very large 0.76 (p<0.01) Interaction effects: S-T relations by large 2.50 (p<0.01) S-T relations by very large 2.65 (p<0.01)	Results presented in log odds metric authors convert to odds ratios for the narrative summary. **Authors' report** 'Compared to medium sized schools, large and very large schools have higher drop out rates among schools of average student teacher relations. This is particularly strong for large schools (coefficient 1.32 or nearing 300% increase in the odds of dropping out p<0.001). Small schools also have higher drop out rates than medium sized schools (coefficient of 0.75, or more than 100 percent increase in the odds, p<0.10)' (p 22). 'As school size increases, the impact of positive student-teacher relations also disappears (for large schools: -1.96 + 2.50 = .54, a non significant difference; for very large schools: -1.96 + 2.65 = .69, also a non significant difference (referenced to public or Catholic schools of small or medium size)' (p 23).
Leung and Ferris (2002) Binary Logit model N= 616 students	**Dependent variables** Violence: whether an individual reported participating in some form of violent behaviour over the year (1995).	1. Unstandardized coefficients 2. Marginal effects School size continuous 1. 0.35*	Logit coefficients are not easily interpreted; interpretations below are based on marginal effects calculated by author. **Authors' report**

Study	Variables	Results	Conclusions
Continuous Count R^2=0.60 McFadden R^2=0,15 Categorical Count R^2=0.60 McFadden R^2=0.15 Note these R^2 are not the same as OLS R^2	**Explanatory variables** (i) School size: (a) continuous variable; number of students in 1,000s. (b) categorical variable: <1000, 1,000-1,499, 1,500-1,999, >2,000, (ii) dropout: if the student had dropped out of school by 17 yrs of age, (iii) if average family income in the subject's school was (a) below $30,000, or (b) above $50,000, (iv) two parents: if subject lived with both biological parents at age of 16 years, (v) friend: best friend arrested by police at the age of 16 years, (vi) adult: if student knew adults who were criminal(s) at age of 16, (vii) gangmen: if student admitted to be a gang member at age of 16, (viii) college: if mother had college education, (ix) college: if father had college education	2. 0.09* School size categorical Size: 1000-1499 1. 0.27 2. 0.068 Size: 1500-1999 1. 0.22 2. 0.05 Size: >2000 1. 0.884** 2. 0.219**	An increase in school enrolment of 1,000 is associated with an increase in self-reported violence of 9% (p<0.05) (p 13). Students attending schools of 1,000-1,999 students are not statistically significantly more likely to engage in violent behaviour than students attending schools with fewer than 1,000 students (NS) (p 13). Students attending schools with over 2,000 students are statistically significantly more likely (22%) to report engaging in violent behaviour than students attending schools with fewer than 1,000 students (p<0.01) (p 13).
Ma (2001) 6th grade N=6,883 students N=147 schools 8th grade N=6,868 students N=92 schools No variance scores reported considered unreliable as random effects used	**Dependent variables** (i) Students being bullied (termed 'victims'), (ii) students bullying others (termed 'bullies') **Explanatory variables** **School level:** (i) school size, (ii) other school level variables: (a) school average SES, (b) disciplinary climate (the extent to which students internalise the norms of the school and conform to them), (c) academic press (the extent to which school staff value academic achievement and hold high expectations of students, (d) parent involvement (the extent to which parents communicate with their children about schoolwork, or volunteer in school). **Individual level:** (i) Gender, (ii) SES (iii) number of parents, (iv) number of siblings, (v) academic condition, (vi) affective condition, (vii) physical condition	Unstandardized coefficient (SE) Results based on transformed true scores Grade 6: Base -0.00 (.00) Contrast 0.01* (.00) Grade 8: Base –0.01*** (.00) Contract .00 (.00) Significant base effect shows an association that is the same for both victims and bullies, significant contrast shows an association that is significantly different between victims and bullies. Results based on original outcomes Grade 6: Victims: –0.00 (0.00) Bullies: –0.01* (0.00) Grade 8: Victims: –0.01** (.00) Bullies: –0.01** (0.00)	Dependent variable based on factors scores for being and bully or a victim **Authors' report** At grade 6, school size shows a partial association with victims that is significantly different from that of bullies (p<0.05) (p 361). At grade 8, school size shows an equivalently shared 'effects' on victims and bullies that was statistically significant (p<0.001) (p 362). At grade 6, school size was more strongly related to being a bully than a victim. The negative sign means that students in small schools were more likely to bully (p<0.05) (p 363). At grade 8, school size showed an equivalently shared 'effect' on victims and bullies. As above, the negative sign means that students in small schools were more likely to bully and be bullied (p<0.01) (p 365).

Study	Variables	Results	Conclusions
McLaughlin *et al.* (2000) Partial correlations Least squares ordinary regression Simultaneous equation modelling Middle: N=496 1.a. R^2=0.55, b. 0.87 2.a. R^2=0.58, b. 0.62 3.a. R^2=0.15, b. 0.26 4.a. R^2=0.02, b. 0.03 5.a. R^2=0.28, b. 0.42 High: N=595 1.a. R^2=0.51, b. 0.82 2.a. R^2=0.57, b. 0.68 3.a. R^2=0.20, b. 0.45 4.a. R^2=0.03, b. 0.09 5.a. R^2=0.38, b. 0.51	**Dependent variable** School average reading and maths score taken at grades 4, 8 and 11 (Note: ages are not particularly clear.) Inter-relationships between all variables studied therefore results can also be drawn about size and school climate, teachers' self perceptions of their influence, normative cohesion and class size. **Explanatory variables** (i) Log of enrolment, (ii) % of students eligible for free or reduced lunch, (iii) teachers' perception of poverty being a problem, (iv) % of minority students (v) teachers' perception of racial tension (vi) % of students with limited English proficiency, (vii) % of students in an ESL program, (viii) % of teachers with a Masters degree, (ix) average years of teaching experience, (x) teachers note student behaviour problems (two measures), (xi) average class size, (xii) student-teacher ratio, (xiii) teachers satisfied with classroom size, (ix) teachers' sense of influence over classroom affairs, (xv) teachers' sense of influence over school policies, (xvi) teachers' perceptions of clarity of norms, (xvii) teachers' perception of co-operation among staff	Standardized coefficients a. OLS, b. SEM 1. Student achievement: Middle: a. –0.07 b. 0.06 High: a. 0.18* b. 0.32* 2. School climate Middle: a. –0.15* b. –0.14 High: a. –0.23* b. –0.28* 3. Teacher perceptions of influence Middle: a. –0.15* b. –0.06 High: a. –0.21* b. –0.20* 4. Normative cohesion Middle: a. –0.01 b. –0.11 High: a. –0.10* b. –0.18* 5. Class size Middle: a. 0.51* b. 0.57* High: a. 0.55* b. 0.63*	Measures are based on composite measures and are not easily interpretable quantitatively. **Authors' report** 'The SEM and OLS results both… indicate that there is higher achievement in larger secondary schools, but that there is no significant relation between school size and achievement among elementary and middle schools' (p 38). 'School size also appears to be a correlate of school climate – smaller schools tend to have better behavioural climates, as perceived by teachers. However according to SEM analysis this relationship is only statistically significant at the secondary school level. A similar finding holds for teachers' perceptions of control over their classrooms and influence on school policies' (p 38).
McMillen *et al.* (2000) Partial correlations Middle: N=308 High: N=292 R^2= NA	**Dependent variables** (i) Student achievement scores in maths, reading, English and algebra. Scores from End-of-Grade (EOG) and End-of-Course (EoC) testing databases. For grade 9-12 schools, the achievement data consisted of EoCo scores for algebra I, English I, history, biology I and economic, legal and political systems, (ii) number of violent incidents per 100 students, (iii) school dropout rates. **Explanatory variables** (i) School size, (ii) ethnicity (percentage of non-white students), (iii) percentage eligible for free school meals, (iv) percentage whose parents had no formal education beyond high school	No numerical results presented	**Authors' report** 'A statistically significant interaction with the "larger size = lower achievement" connection being magnified in schools where a large percentage of children were eligible for free or reduced-price lunch. In subsequent analyses, however, this finding was nullified when parent education level was taken into consideration' (p 20).

Study	Variables	Results	Conclusions
McNeely *et al.* (2002) Hierarchical linear modelling Individual N=71,515 School N=127 Individual level model explains 10.5% of variance School level model 6 explains 41.8% of the variance	**Dependent variable** 1. School connectedness (based on five-item Likert style scale) **Explanatory variables*** **School level: (i) School size measured in 100s, (ii) demographic (Black, % Latino, % from two-parent families),** (iii) % teachers with a masters degree, (iv)% teachers in their first year of teaching at the school, (v) **discipline policies,** (vi) other structural school characteristics i. class size, ii. public schools, iii. urbanicity, (vii) Student participation and classroom management **Individual level: (i) Age, (ii) two-parent family, (iii) ethnicity i. Black, ii. Hispanic, (iv) gender, (v) grade score, (vi) no participation in extracurricular activities, (vii) classroom management score, (viii) skipping school more than twice in last year**	Six models are presented but results are only presented for the final model Weighted HLM Coefficient Model 6 -0.089*** *Bold variables were those remaining in the final model	**Authors' report** (based on coefficient in the model 4 (-0.084)) 'The strength of this association (school size and school connectedness) is fairly weak. An increase of 500 students in schools size – a change of fairly major economic significance to a school district is associated with a very small decline in school connectedness (.04 units on a 0-4 scale)' (p144).
Silins and Mulford (2000) Path analysis N=2,503 Resource: R^2=0.09 Leader: R^2=0.53 Staff valued: R^2=0.61 Leadership satisfaction: R^2=0.77 Community focus: R^2=0.71 Teacher learning: R^2=0.57 Organisational learning: R^2=0.90 Teachers work: R^2=0.08 Participation: R^2=0.40 Engagement: R^2=0.83	12 variables (composite measures of individual survey items measured on a five-point Likert scale) considered in three categories: 1. School context variables: (i) school size, (ii) SES. 2. Internal school variables: (i) resource availability; (ii) leadership style; (iii) satisfaction with leadership; (iv) response to the needs of the community, and ability to work with the local community; (v) collective teacher leadership; (vi) organisational learning; (vii) student perceptions of teachers' work 3. Student outcome variables: (i) student participation and (ii) student engagement Eleven models presented; each model has one of the 12 variables as the dependent variable (socio-economic achievement is not included as a dependent variable), variables that exerted an 'effect' are included in each model. School size is included in each of the models.	Path coefficients (1) Total effects, (2) indirect effects, (3) direct effects Resource: (1) –0.30, (2) No effect (3) –0.30 Leader: (1) –0.22, (2) –0.22, (3) No effect Staff valued: (1) –0.21, (2) –0.21, (3) No effect Leadership satisfaction: (1) –0.20, (2) –0.20, (3) No effect Community focus: (1) –0.18, (2) –0.18, (3) No effect Teacher leadership: (1) –0.15, (2) –0.15, (3) No effect Organisational learning: (1) –0.23, (2) –0.23, (3) No effect Teachers' work: (1) –0.06, (2) –0.06, (3) No effect Participation: (1) –0.39, (2) –0.03, (3) –0.36 Engagement:	Coefficients are based on factor scores and the results are not easily interpretable quantitatively. **Authors' report** (relating to organisational learning) 'Significant indirect effects were exerted from two other variables: school size (*i*=-0.23) and SES (*I*=-0.18)' (p 9). 'Teachers' work and school size directly influenced participation (p=0.51 and –0.36). The negative path from school size indicates that students' participation is higher in the smaller schools' (p 9). 'Indirect influences on Engagement worthy of note were organizational learning (*I*=0.25), resource (*I*=0.19), leader (*I*=0.16) and school size (*I*=-0.16)' (p 9).

Study	Variables	Results	Conclusions
		(1) –0.16, (2) –0.16, (3) No effect	
Spielhofer *et al.* (2002) Multilevel model Logistic regression N=369,341 students N=2,954 schools Proportion of variance explained by each level is unclear	**Multilevel model** **Dependent variables:** (i) Total GSCE point score (points derived from A*=8, A=7, B=6 down to G =1), (ii) average GSCE point score, (iii) number of GCSEs taken, (iv) maths point score, (v) English language point score, (vi) total science score (e.g. grade CC for double science = 10 points), (vii) average science points score, (viii) No. of science GCSEs **Explanatory variables** **School level:** (i) number of year 11 students in the school, with number of year 11 students squared, in hundreds, (ii) grammar or comprehensive, (iii) % students eligible for free school meals (FSM), (iv) boy's or girl's only school; mixed school, (v) average size of one teacher classes in school, (vi) whether school has a sixth form, (vii) size of sixth form **LEA level:** % of students entering grammar schools **Student level:** (i) Prior attainment (level achieved at KS2 in maths, English and science in 1996), (ii) sex (boy or girl), (iii) ages in years and months **Logistic regression** **Dependent variables** (i) Probabilities of entry to higher tiers at KS3 in: (a) maths, (b) science (ii) probabilities of taking various subjects at GSCE, (a) double-award science, (b) design and technology - food, graphics, resistant materials, (c) physics, (d) chemistry, (e) biology, (f) French plus German **Explanatory variables** **School level:** (i) Small school (up to 180 students), medium school (181-230) or large school (231 or more) based on number of students in year 9, (ii)	Main results for size and size^2 are presented additional interaction results are shown. Unstandardized coefficients (SE) 95% CI (i) size (ii) size^2 1. Total GCSE score: (i) 3.874 (0.668)* 2.57 to 5.18 (ii) –1.124 (0.172)* –1.46 to –0.79 2. Average GSCE point score (i) 0.334 (0.054)* 0.23 to 0.44 (ii) –0.093 (0.014)* –0.12 to –0.07 3. Number of GCSEs taken (i) 0.590 (0.117)* 0.36 to 0.82 (ii) **–**0.152 (0.030)* **–**0.21 to **–**0.09 4. Maths point score (i) 0.291 (0.065) * 0.16 to 0.42 (ii) **–**0.078 (0.017)* **–**0.11 to **–**0.05 5. English language point score (i) 0.440 (0.074)* 0.30 to 0.58 (ii) **–**0.118 (0.019)* **–**0.16 to **–**0.08 6. Total science score (i) 0.650 (0.187)* 0.28 to 1.02 (ii) **–**0.176 (0.048)* **–**0.27 to **–**0.08 7. Average science score (i) 0.240 (0.071)* 0.10 to 0.38 (ii) **–**0.067 (0.018)* **–**1.10 to **–**0.03 8. Number of science GCSEs (i) 0.117 (0.034)* 0.05 to 0.18 (ii) **–**0.027 (0.009)* **–**0.04 to **–**0.01 Logistic regression results Base case is a medium-sized mixed comprehensive school. Odds ratios (significance) 1. Entry to a higher tier in maths small school 1.03 (0.2%) small grammar school 0.56 (0.7%) 2. Entry to a higher tier in science size NS	**Authors' report** 'School size was found to have a non-linear effect on all outcomes. It was found that each variable tended to increase with school size until an optimum year size was reached, after which increasing year size related to decreases in the outcome variables. For most outcomes, the results indicated an optimum size of between 170-190 pupils in year 11' (p 56). 'The optimum size for a school with low numbers of FSM children is much lower than for a school with a high number. For example, if there are no FSM children in the school, the optimum size is around 150 pupils in year 11. With 16% FSM (the average level) this rises to 175 pupils. With 50% FSM, the optimum size is 215' (p 56). 'There is a significant positive interaction between year 11 size and prior attainment. That is the greater the prior attainment of a child the larger the optimum size for this child' (p 57). **Reviewers Interpretation:** Students in small schools were comparatively more likely to be entered into KS3 higher tier in maths (odds ratio 1.026), but not for science, than those attending middle sized schools. Students in small schools were less likely to be entered for either double science (0.895) or single chemistry (0.772), physics (0.0747) and

Study	Variables	Results	Conclusions
	grammar or comprehensive, (iii) % students eligible for free school meals (FSM), (iv) boys' or girls' only school or mixed school, (v) average size of one teacher classes in school, (vi) whether school has a sixth form **Student level:** (i) Prior attainment (level achieved at KS2 in maths, English and science in 1996, average level, and an indicator for students with average level below 3), (ii) sex (boy or girl), (iii) ages in years and months	3. Entry to double science small school 0.895 (0.0%) large school 1.08 (.0%) 4. Entry to food option small school 1.14 (0.0%) small grammar school 1.44 (0.0%) 5. Entry to graphics option small school 0.93 (0.0%) small grammar school 1.38 (0.0%) 6. Entry to resistant materials option small school 1.17 (0.0%) small grammar school 2.28 (0.0%) 7. Entry to physics GCSE small school 0.75 (0.0%) large school 1.07 (0.3%) 8. Entry to chemistry GCSE small school 0.77 (0.0%) large school 1.09 (0.0%) 9 Entry to biology GCSE small school 0.78 (0.0%) large school 1.07 (0.5%) small grammar school 0.89 (4.8%) 10. Entry to French and German GCSE Large school 0.91 (0.0%) Small grammar school 1.17 (2.5%)	biology (0.0784), whilst students in large schools were more likely to be entered (1.077, 1.085, 1.070, 1.065 respectively). Students in small schools were more likely to be entered for food technology (1.136) and resistant materials (1.168) GCSEs, but less likely to be entered into graphics (0.928) than students in medium-sized schools. Students in large schools were less likely to be entered for both French and German GCSE (0.905) than those in medium-sized schools.
Stiefel *et al.* (2000) Ordinary least squares multiple regression N=121 (i) R^2=0.60 (ii) R^2=0.70	**Dependent variables** (i) budget per student (logged), (ii) budget per graduate (logged) **Independent variables** (i) school enrolment (logged), (ii) school type (dummy), (iii) poverty, (iv) part time special educational needs, (v) proportion passing maths competency test **Interaction terms (dummy)** (i) articulated school x size (ii) transfer school x size (iii) vocational school x size Reference category academic and specialised school	Unstandardized regression coefficient (SE) (i) Budget per student –0.096 (0.015)*** (ii) Budget per graduate –0.140 (0.048)** Interaction effects between type and size (i) budget per student (ii) budget per graduate articulated x size (i) –0.005 (0.008) (ii) –0.018 (0.026) transfer x size	**Reviewers' interpretations** An increase in school size of 10% is associated with a decrease in budget per student of 0.96% (p<0.001). An increase in school size of 10% is associated with a decrease in budget per graduate of 1.4% (p<0.01). An articulated school that increases its size by 10% is associated with a decrease in budget of 1% (per student), 1.6% (per graduate) (NS). A transfer school that increases its size by 10% is associated with an increase its budget of 0.02% (per student), 0.07% (per graduate)

Study	Variables	Results	Conclusions
		(i) 0.012 (0.006)** (ii) .147 (0.019)*** vocational x size (i) 0.005 (0.003) (ii) 0.001 (0.011)	(p<0.01 and 0.001). A vocational school that increases its size by 10% is associated with a decrease its budget of 0.9% (per student), 1.4% (per graduate) (NS).
Taylor and Bradley (2000) Ordinary least squares regression equations 1993 N=2,034 (TH) R^2=0.53 (SH) R^2=0.56 1995 N=3,030 (TH) R^2=0.44 (SH) R^2=0.44 1996 N=3,014 (TH) R^2=0.50 (SH) R^2=0.55 1997 N=3,087 (TH) R^2=0.52 (SH) R^2=0.57 Change N=2,041 (TH) R^2=0.28 (SH) R^2=0.27 R^2=adjusted	**Dependent variable** (i) Costs per student (logged) (grant maintained only), (ii) staff hours per student (logged) (grant maintained only), (iii) teaching hours per student (logged), (iv) support hours per student (logged) **Explanatory variables** (i) Student capacity of school, (logged), (ii) recurrent expenditure per student, (logged), (iii) capacity utilisation rate of school, (iv) school type, (v) admission policy of school, (vi) gender of students, (viii) proportion of students taking A - levels or vocational courses, (iv) proportion of students with special educational needs, (v) family background of students, % free school meals, ethnic minority backgrounds, English second language, (vi) influence of LEA on unit costs of school	Unstandardized regression coefficients (SE) 1. Student capacity 2. Student capacity utilisation rate (iii) Teaching hours 1993 1. –0.096 (0.006)*** 2. –0.199 (0.010)*** 1995 1. –0.075 (0.005)*** 2. –0.136(0.008)*** 1996 1. –0.074 (0.005)*** 2. –0.134 (0.008)*** 1997 1. –0.069 (0.005)*** 2. –0.140 (0.009)*** (iv) Support hours 1993 1. –0.234 (0.018)*** 2. –0.420 (0.028)*** 1995 1. –0.185 (0.016)*** 2. –0.330 (0.026)*** 1996 1. –0.195 (0.015)*** 2. –0.335 (0.023)*** 1997 1. –0.198 (0.016)*** 2. –0.335 (0.024)*** Changes 1993-97 (iii) Teaching hours 1. –0.072 (0.019)*** 2. –0.159 (0.016)*** (iv) Support hours 1. –0.383 (0.045)*** 2. –0.404 (0.040)***	**Reviewers' interpretations** A 10% increase in student capacity is associated with a decrease in teaching hours per student of 0.69%-0.96%. A 10% increase in student capacity utilisation rate is associated with a decrease in teaching hours per student of 1.9%-1.4%. A 10% increase in student capacity is associated with a decrease in support hours per student of 1.9%-2.3%. A 10% increase in student capacity utilisation rate is associated with a decrease in support hours of 3.3%-4.2%. A 10% change in student capacity of 10% is associated with a decrease in teaching hours of 0.72%, and in support hours of 3.8%. A 10% change in student capacity utilisation rate is associated with a decrease in teaching hours of 1.6% and in support hours of 4%. (p<0.001, in all analyses)

Study	Variables	Results	Conclusions
Welsh *et al.* (1999) Hierarchical linear modelling N= 7,583 students in 11 schools Individual level model: 16% of total variance Local community, between school model: 5% of total variance Imported community, between school model: 4% of total variance	**Dependent variable** School disorder: measured by the student misconduct scale **Explanatory variables** **Level 1:** (i) Commitment to conventional goals (a) school effort, (b) school rewards, (ii) attachment to pro-social others, (iii) involvement in conventional activities, (iv) belief about conventional rules, (v) age, (vi) race, (vii) gender **Level 2:** (i) School size: measured using school district data as total number of students enrolled in school, (ii) school attachment, (iii) community crime, (iv) community poverty: measured using a. single parent households, b. minority composition and c. median income and d. household size, (v) community stability: measured, using household size and residential stability, minority composition and residential stability Two models were examined; one with variables based on data of the areas in which the schools were situated and one using data from where the students actually lived	(Low scores indicate higher levels of misconduct.) Unstandardized coefficients (standard error) Local community models: 1. With stability variable: 0.00024 (0.0008) 2. Without stability variable 0.00026 (0.0004) Imported community model: 0.00004 (0.0004)	**Reviewers' interpretations*** No statistically significant associations were identified between school size and student self reported misconduct. An increase in school size of 100 is associated with an increase of approximately 0.02 in scores of student self reported misconduct (range 0-4). An increase in school size of 100 is associated with an increase of 0.004 in scores of student self-reported misconduct (range 0-4). ***Note:** Lower scores indicate higher levels of misconduct; therefore a positive increase means that there are lower levels of self-reported misconduct.
Welsh *et al.* (2000) Regression analysis N=43 Local R^2=0.64 Imported R^2=0.67 Path analysis N=43 Local	**Dependent variable** School disorder based in school incident data and dismissal rates **Explanatory variables** (i) School size: (exogenous) total student enrolment, (ii) community poverty: (exogenous) based on single parent families, minority compositions, household income and household size, (iii) community stability: (exogenous) household size, residential stability, minority composition, (iv) community crime: (endogenous), (v) school stability: (endogenous) index of school culture calculated from average daily attendance	Initial results are presented (1) in a multiple regression, (2) in a path model and (3) correlations are decomposed into total, direct and indirect effects. 1. Multiple Regression: Unstandardized regression coefficients (SE) (i) Local community -3.4935^{-04} (3.6540^{-04}) (ii) Imported community -4.7054^{-04} (3.6515^{-04}) 2. Path coefficients (taken from standardized regression coefficients) (i) Local community School size exerted an influence on school	Coefficients are based on factor scores and so are not easily interpretable. No significant associations were found between school size and school disorder (see multiple regressions). Authors' report 'School size exerts a substantial indirect effect on school disorder through school stability. However its total effect on school disorder is considerably less than that of imported community poverty' (p 268).

Study	Variables	Results	Conclusions
Q=0.95 Imported Q=0.90	rate, and percentage of non-graduating students from the previous school year who did not re-enrol. Two models used (1) based on 'local model' (school surroundings) and (2) based on 'imported' model, based on data where the students actually resided. Factor analysis was used to identify constructs relating to the independent variables for both local and imported models.	stability (P=-0.237) (NS) (ii) Imported community School size exerted an influence on school stability (P=-0.321) (sig p<0.05) 3. Decomposition of correlations (a) Total, (b) Direct and (c) Indirect effects on school disorder (i) Local community: (a) 0.08, (b) -0.11, (c) 0.19 (ii) Imported community: (a) 0.14, (b) -0.14, (c) 0.28	The influence of school size on school stability was only statistically significant in the imported community model.

Appendix 4.5: Threats to validity: meta-summaries

Table 4.5.1: Threats to validity and direction of 'effect' – Achievement without control for prior attainment

Indicators that threat to validity met (for studies in this review)	Abbott *et al.*, 2002	Bradley and Taylor, 1998	Bedard *et al.*, 1999	Bickel and Howley, 2000	Bickel *et al.* 2001	Bowles and Bosworth, 2002	Driscoll *et al.*, 2003	Gill *et al.*, 2002	Heck, 1993	Howley, 1996a	Howley, 1999a	Howley, 1999b	Johnson *et al.*, 2002	Mclaughlin *et al.*, 2000	McMillen *et al.*, 2000
Normal distribution/Outliers (or corrected)	x	√	√	√	√	√	x	x	x	√	√	√	√	x	x
Independent observations (or corrected)	√	x	x	√	x	x	x	√	X	x	x	x	x	x	x
Confirmation that data are linear (or corrected)	x	√	√	√	√	√	x	x	X	x	√	√	√	√	x
Data are homoscedastic (or corrected)	x	√	NA	√	√	√	√	x	X	x	x	x	x	x	x
Specification of the model: 1 = Adjusted R^2, 2 = R^2	NR	0.59 to 0.62*	NR	26% to 29%	0.51 to 0.53[1]	0.77 to 0.95 [2]	0.68 to 0.71 [2]	0.56 to 0.58 [1]	0.41 to 0.45[1]	0.09 to 0.17[2]	0.18 to 0.30[1]	0.37 to 0.47 [2]	0.19 to 0.37 [2]	0.51 to 0.57 [2]	NA
F Test for model fit	x	x	x	x	x	x	x	x	√	x	x	x	x	√	NA
Analysis of residuals	x	x	x	√	x	x	x	x	X	x	x	x	x	x	NA
Scale reliability coefficients:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Inter-rater reliability	x	x	x	x	x	x	x	√	X	x	x	x	x	x	x
Random selection of sample (at school level or weighted data)	NA	NA	NA	NA	NA	NA	NA	√	NA	NA	NA	NA	NA	x	NA
Check for multicollinearity (IVR Technique)	√	x	x	√	√	NA	√	√	√	x	√	√	√	√	x
Adequate sample size *	√	√	√	√	√	x	√	√	√	√	√	√	√	√	NA
Multiple dependent variables for the same construct	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Measure dependent variable in more than one way	x	√	x	x	x	x	x	x	X	x	x	x	x	x	√
Longitudinal data	x	√	x	x	x	√	x	x	√	x	x	x	√	x	√
Use of gain or value-added scores?	x	x	x	x	x	x	x	x	X	x	x	x	x	x	x
Class size/student-teacher ratio	x	√	√	√	√	x	√	√	X	x	√	√	x	√	x
Funding: public/private	NA	NA	NA	NA	X	NA	x	√	X	x	x	x	x	NA	NA
Sensitivity analysis/weighting	x	x	x	X	X	x	x	√	X	√	x	x	x	√	x
Use of path analysis	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Assessment of simultaneity (SEM)	x	x	x	x	x	√	x	x	x	x	x	x	x	√	x
Direction of 'effects' **	+	∩	+*	-/+	+	-	-	NR	-	+/-	-/+	+/-	-	+	NR
WoE A High, Medium, Low	M	H	M	M	M	L	H	H	M	M/L	M/L	M/L	M/L	H	L
WoE B High, Medium, Low	M	M	M	M	M	L	M	L	L	M	M	M	M	M	L
WoE C High, Medium, Low	M	H	M	M	M	L	M	L	L	M	M	M	M	H	M
WOE D High, Medium, Low	M	H	M	M	M	L	M	L	L	M/L	M/L	M/L	M/L	H	L

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m. **Direction of 'effects' - negative, + positive, ∩∪ = Quadratic
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.2: Threats to validity and direction of 'effect' – Achievement with control for prior attainment

Indicators that threat to validity met (for studies in this review)	**Atkinson and Wilson, 2003**	**Bradley and Taylor, 2003**	**Lee and Smith, 1997**	**Spielhofer *et al.*, 2002**
Normal distribution/outliers (or corrected)	√	x	√	√
Independent observations (or corrected)	x	x	√	√
Confirmation that data are linear (or corrected)	√	√	√	√
Data are homoscedastic (or corrected)	x	x	x	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	0.05 to0.18[2]	0.15 to 0.87 [2]	Math 67% Read 48% variance explained	NR
F Test for model fit	x	x	x	x
Analysis of residuals	x	x	√	x
Scale reliability coefficients	NA	NA	α0.8 to 0.94	NA
Inter-rater reliability	x	x	0.7 & 0.56	x
Random selection of sample (at school level or weighted data)	NA	NA	√	NA
Check for multicollinearity (IVR Technique)	x	x	√	x
Adequate sample size*	√	√	√	√
Multiple dependent variables for the same construct	√	√	√	√
Measure dependent variable in more than one way	√	√	√	√
Longitudinal data	x	√	√	√
Use of gain or value-added scores?	√	√	√	√
Class size/student-teacher ratio	x	√	x	√
Funding: public/private	NA	NA	√	NA
Sensitivity analysis/weighting	x	x	√	x
Use of path analysis	x	x	x	x
Assessment of simultaneity (SEM)	x	x	x	x
Direction of 'effects' - negative, + positive, ∩/∪ = Quadratic	∩	∩	∩	∩
WoE A High, Medium, Low	H	H	H	H
WoE B High, Medium, Low	M	M	M	M
WoE C High, Medium, Low	H	H	H	H
WoE D High, Medium, Low	H	H	H	H

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.3: Threats to validity and direction of 'effect' – dropout and absence

Indicators that threat to validity met (for studies in this review)	Bradley and Taylor, 2003	Fetler, 1997	Heck, 1993	Lee and Burkam, 2001	McMillen *et al.,* 2000
Normal distribution/outliers (or corrected)	x	x	x	√	x
Independent observations (or corrected)	x	x	x	√	x
Confirmation that data are linear (or corrected)	√	x	x	√	x
Data are homoscedastic (or corrected)	x	x	x	√	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	0.34 to 0.53 [2]	0.50 [2]	0.18 to 0.30 [1]	NR	NA
F Test for model fit	x	x	√	x	NA
Analysis of residuals	x	x	x	x	x
Scale reliability coefficients	NA	NA	NA	NA	NA
Inter-rater reliability	NA	NA	NA	NA	x
Random selection of sample (at school level or weighted data)	NA	NA	NA	√	NA
Check for multicollinearity (IVR Technique)	X	√	√	√	X
Adequate sample size*	√	√	√	√	NA
Multiple dependent variables for the same construct	X	x	X	x	X
Measure dependent variable in more than one way	x	x	x	x	x
Longitudinal data	√	√	x	x	X
Use of gain or value-added scores?	NA	NA	NA	NA	NA
Class size/student-teacher ratio	√	x	x	x	x
Funding: Public/Private	NA	NA	x	√	NA
Sensitivity analysis/weighting	x	NA	x	√	x
Use of path analysis	x	x	x	x	x
Assessment of simultaneity (SEM)	x	√	x	x	x
Direction of 'effects' - negative, + positive , ∩/∪ = Quadratic	∩	+	-	∩*	NR
WoE A High, Medium, Low	H	M	M	H	L
WoE B High, Medium, Low	M	M	L	M	L
WoE C High, Medium, Low	H	M	L	H	M
WoE D High, Medium, Low	H	M	L	H	L

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.4: Threats to validity and direction of 'effect' – Attitudes towards school and perceptions of school

Indicators that threat to validity met (for studies in this review)	Bowen *et al.,* 2000	Lee and Smith, 1997	McNeely *et al.,* 2002	Silins and Mulford, 2000
Normal distribution/outliers (or corrected)	x	√	x	x
Independent observations (or corrected)	x	√	√	x
Confirmation that data are linear (or corrected)	√	√	x	x
Data are homoscedastic (or corrected)	NA	x	x	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	NA	Math 67% Read 48% variance explained	41.8% between school variance	0.09 to 0.90 [2]
F Test for model fit	NA	x	x	√
Analysis of residuals	NA	√	x	x
Scale reliability coefficients	α0.67 to 0.89	α0.8 to 0.94	α0.79	α0.79 to 0.83
Inter-rater reliability	NA	0.7 & 0.56	NA	NA
Random selection of sample (at school level or weighted data)	√	√	√	x
Check for multicollinearity (IVR Technique)	x	√	x	x
Adequate sample size*	x	√	√	x
Multiple dependent variables for the same construct	√	√	x	X
Measure dependent variable in more than one way	√	x	x	√
Longitudinal data	x	x	x	x
Use of gain or value-added scores?	NA	NA	NA	NA
Class size/student teacher ratio	x	x	√	x
Funding: public/private	NA	√	√	x
Sensitivity analysis/weighting	√	√	x	x
Use of path analysis	x	x	x	√
Assessment of simultaneity (SEM)	x	x	x	x
Direction of 'effects' - negative, + positive, ∩/∪ = Quadratic	NA	-	-	-
WoE A High, Medium, Low	M	H	M/H	L/M
WoE B High, Medium, Low	L	M	M	M
WoE C High, Medium, Low	L	H	M	M
WoE D High, Medium, Low	L	H	M	L/M

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.5: Threats to validity and direction of 'effect' – behaviour

Indicators that threat to validity met (for studies in this review)	Heck, 1993	Leung and Ferris, 2002	Ma, 2001	McMillen *et al.*, 2000	Welsh *et al.*, 1999	Welsh *et al.*, 2000
Normal distribution/outliers (or corrected)	x	√	√	x	x	x
Independent observations (or corrected)	x	x	√	x	√	x
Confirmation that data are linear (or corrected)	x	√	x	x	x	x
Data are homoscedastic (or corrected)	x	x	x	x	x	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	0.18 to 0.30 [1]	0.60*	NR	NA	20%of total variance	0.64 to 0.67 [2]
F Test for model fit	√	√	x	NA	x	√*
Analysis of residuals	x	x	x	x	x	x
Scale reliability coefficients	NA	NA	α0.61 to 0.78	NA	α0.54	NA
Inter-rater reliability	NA	NA	NA	NA	NA	NA
Random selection of sample (at school level or weighted data)	NA	x	NA	NA	x	NA
Check for multicollinearity (IVR Technique)	√	x	x	x	√	√
Adequate sample size*	√	x	√	NA	x	√
Multiple dependent variables for the same construct	x	x	√	X	x	x
Measure dependent variable in more than one way	x	x	x	x	√	√
Longitudinal data	x	x	X	X	X	X
Use of gain or value-added scores	NA	NA	NA	NA	NA	NA
Class size/student-teacher ratio	x	x	x	x	x	x
Funding: public/private	x	x	x	NA	NA	NA
Sensitivity analysis/ weighting	x	x	x	x	x	NA
Use of path analysis	x	x	x	x	x	√
Assessment of simultaneity (SEM)	x	x	x	x	x	x
Direction of 'effects' - negative, + positive , ∩/∪ = Quadratic	-	+	-	+/NR	+	-
WoE A High, Medium, Low	M	M/L	H	L	M	M
WoE B High, Medium, Low	L	M	M	L	L	L
WoE C High, Medium, Low	L	M	H	M	M/L	M
WoE D High, Medium, Low	L	M/L	H	L	M/L	M

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.6: Threats to validity and direction of 'effect' – school organisation

Indicators that threat to validity met (for studies in this review)	Spielhofer *et al.,* 2002	McLaughlin *et al.,* 2000
Normal distribution/outliers (or corrected)	x	x
Independent observations (or corrected)	x	x
Confirmation that data are linear (or corrected)	x	√
Data are homoscedastic (or corrected)	x	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	NR	0.42 to 0.51[2]
F Test for model fit	x	√
Analysis of residuals	x	x
Scale reliability coefficients	NA	NA
Inter-rater reliability	NA	NA
Random selection of sample (at school level or weighted data)	NA	x
Check for multicollinearity (IVR Technique)	x	√
Adequate sample size*	√	√
Multiple dependent variables for the same construct	√	√
Measure dependent variable in more than one way	√	√
Longitudinal data,	x	NA
Use of gain or value-added scores?	NA	NA
Class size/student-teacher ratio	√	√
Funding: public/private	NA	NA
Sensitivity analysis/weighting	x	√
Use of path analysis	x	x
Assessment of simultaneity (SEM)	x	√
Direction of 'effects' - negative, + positive , ∩/∪ = Quadratic	+/-	+
WoE A High, Medium, Low	H	H
WoE B High, Medium, Low	M	M
WoE C High, Medium, Low	H	H
WoE D High, Medium, Low	H	H

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.5.7: Threats to validity and direction of 'effect' – teacher perceptions of school climate and organisation

Indicators that threat to validity met (for studies in this review)	Silins and Mulford, 2000	McLaughlin *et al.*, 2000
Normal distribution/outliers (or corrected)	x	x
Independent observations (or corrected)	x	x
Confirmation that data are linear (or corrected)	x	√
Data are homoscedastic (or corrected)	x	x
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	0.09 to 0.90 [2]	0.02 to 0.62[2]
F Test for model fit	√	√
Analysis of residuals	x	x
Scale reliability coefficients	α0.79 to 0.83	a.=0.42 & 0.20, b.= 0.5 & 0.7, c.=NR*
Inter-rater reliability	x	a.=0.46, b=0.75, c.=0.62*
Random selection of sample (at school level or weighted data)	x	x
Check for multicollinearity (IVR Technique)	x	√
Adequate sample size*	x	√
Multiple dependent variables for the same construct	√	√
Measure dependent variable in more than one way	√	√
Longitudinal data	NA	NA
Use of gain or value-added scores?	NA	NA
Class size/student-teacher ratio	x	√
Funding: public/private	x	x
Sensitivity analysis/weighting	x	√
Use of path analysis	√	x
Assessment of simultaneity (SEM)	x	√
Direction of 'effects' - negative, + positive, $\cap/\cup$ = Quadratic	-	-
WoE A High, Medium, Low	L/M	H
WoE B High, Medium, Low	M	M
WoE C High, Medium, Low	M	H
WoE D High, Medium, Low	L/M	H

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Table 4.6.8: Threats to validity and direction of 'effect' – economic outcomes

Indicators that threat to validity met (for studies in this review)	Bickel *et al.,* 2001	Bowles and Bosworth, 2002	Kirjjavainen and Loikkanen, 1998	Stiefel *et al.,* 2000	Taylor and Bradley, 2000
Normal distribution/outliers (or corrected)	√	√	√	√	√
Independent observations (or corrected)	x	x	x	x	x
Confirmation that data are linear (or corrected)	√	√	√	√	√
Data are homoscedastic (or corrected).	√	√	x	x	√
Specification of the model: Adjusted R^2 (range for each dependent outcome reported) 1= Adjusted R^2 , 2= R^2	0.51 to 0.53 [1]	0.77 to 0.95 [2]	0.23 to o.36 [2]	0.60 to 0.70 [2]	0.27 to 0.56 [2]
F Test for model fit	x	x	x	x	x
Analysis of residuals	x	x	x	x	x
Scale reliability coefficients	NA	NA	NA	NA	NA
Inter-rater reliability	NA	NA	NA	NA	NA
Random selection of sample at school level or weighted data)	NA	NA	x	NA	NA
Check for multicollinearity (IVR Technique)	√	√	x	√	x
Adequate sample size*	√	x	√	√	√
Multiple dependent variables for the same construct	x	x	√	√	√
Measure dependent variable in more than one way	x	x	√	x	√
Longitudinal data,	x	√	x	x	√
Use of gain or value-added scores?	NA	NA	NA	NA	NA
Class size/student teacher ratio	√	x	√	x	√
Funding: public/private	NA	NA	√	NA	NA
Sensitivity analysis/weighting	x	x	x	x	x
Use of path analysis	x	x	x	x	x
Assessment of simultaneity (SEM)	x	√	x	x	x
Direction of 'effects' - negative, + positive, ∩/∪ = quadratic	-	-	∩	-	-
WoE A High, Medium, Low	M	L	H/M	M	H
WoE B High, Medium, Low	M	L	M	M	M
WoE C High, Medium, Low	M	L	M	M	H
WoE D High, Medium, Low	M	L	M	M	H

*Regression analysis: N >= 104 + m, where m = number of independent variables; Stepwise regression: N >= 40m.
NA = Not applicable. NR = Not reported; Where random selection of sample = NA means that 'all' schools (in city/state/country) used.

Note: In economic analysis, negative (-) direction of 'effect' is a 'good outcome' (i.e. costs fall as school size increases.

First published in 2004 by the
The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR
Tel: 020 7612 6367

http://eppi.ioe.ac.uk/

www.ioe.ac.uk/ssru

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the funder. All errors and omissions remain those of the authors.

This report is available to download from:

http://eppi.ioe.ac.uk/EPPIWebContent/reel/review_groups/School_size/s_s_rv1.pdf

ISBN: **0-9548415-1-4**