

What do we know about effects of school resources on educational results?

Jan-Eric Gustafsson*

Summary

■ The main tradition of research on effects of resources on results investigates educational production functions in which input factors are related to output in statistical models. Many such studies have been conducted but the results tend to be inconsistent, and conclusions from reviews conflict. However, several recent meta-analytic integrations of the estimates from different studies indicate positive effects of resources such as per pupil expenditure, class size, teacher education and teacher experience. Methodological limitations of the production function approach imply, however, that these results should be interpreted with caution. In a more recent approach, randomized experiments have been conducted, with a special focus on the effects of class size. In the largest and most thoroughly analyzed study (the STAR experiment), smaller classes were found to have an advantage at the primary level. Meta-analytic summaries of the effects of class-size also support the conclusion that smaller classes are beneficial in early grades, and particularly so for students with less favorable socio-economic backgrounds. Alternative interpretations of the class size effect are discussed, and it is suggested that it may be due to a more effective socialization to the school environment in smaller classes. Another line of research has established large differences between teachers in terms of the achievement of their students. The research evidence also suggests that teacher education, teacher experience, and in-service training influence teacher competence. It is concluded that class size may be of importance for certain categories of students, while teacher competence appears to be the single most important resource factor.■

JEL classification: I20.

Keywords: Educational production functions, quality of schooling, class size, teacher competence.

*Jan-Eric Gustafsson is Professor at the Department of Education, Göteborg University.

What do we know about effects of school resources on educational results?

Jan-Eric Gustafsson

Systematic research on effects of school resources on student achievement goes back at least to the 1960s, and literally hundreds of empirical studies have been reported. During the 1980s, reviews of the research were published by Hanushek (1981, 1986), which quite clearly showed that economic resources have little or no effect on student achievement. Burtless (1996, p. 3) observed that as late as in the mid 1990s, Hanushek's conclusion "... probably remains the prevailing view among economists who study school resources and educational achievement."

However, Hanushek's conclusion has been challenged by new results. These were achieved when new methods were employed to synthesize research results, and by new empirical research, based on other methods than have previously been used. The consensus about the conclusion that resources do not matter has thus been replaced by considerable controversy concerning theory, empirical results and methodology. These controversies indicate that exciting progress has been made in this area of research. The purposes of the present paper are to describe and discuss some of the basic issues in these controversies, try to assess what we currently know, and indicate what seems to be the most promising areas of further research.

1. Studies on educational production functions

The earliest attempts to understand effects of school resources on educational achievement were inspired by research on so-called industry production functions, which have been used in economic research for a long time. The basic idea is that to produce an output, inputs such as labor and capital are transformed by technology into products and services. To maximize profit, inputs must be used as efficiently as possible, and the optimal use of resources may be determined by es-

timating functions which relate input factors to output. Schools also transform inputs such as teachers, facilities, and support personnel to produce output in the form of, among other things, student achievement. This suggests that it should be possible to estimate educational production functions with statistical methods, such as multiple regression analysis, that determine relations between input factors and outputs.

1.1. Empirical estimates of educational production functions

The Coleman et al. (1966) report is a landmark study in the estimation of such educational production functions. The purpose of the study was to measure the extent of racial segregation of American schools, but the large amount of data collected allowed the researchers to address other, more general, issues as well. They collected data from a representative sample of over 570 000 students, 60 000 teachers and 3 000 schools, including third-, sixth-, ninth-, and twelfth-grade classes. Among the variables measured were student achievement in different domains, family background, characteristics of teachers, class size, and characteristics of the schools and communities. Using multiple regression analysis at the school level, the researchers tried to determine the relative importance of these categories of variables to explain variation in student achievement.

The major finding was that family background characteristics and community level variables accounted for variance in student achievement at the school level, while school resource variables, such as pupil/teacher ratios, per pupil expenditures, or teacher characteristics accounted for no or little variance.

It may, of course, be that the Coleman report failed to identify effects of resources on achievement because of limitations in the study. The design and analysis of the Coleman et al. study have been extensively discussed, numerous critiques have been published and a large number of reanalyses have been performed (e.g., Bowles and Levin, 1968; Ehrenberg and Brewer, 1995; Mosteller and Moynihan, 1970). It would carry too far to discuss this criticism in detail here, but there are reasons to identify some of the more fundamental problems encountered when attempts are made to estimate the educational production function from survey data of the kind upon which the Coleman report is based.

One of the most fundamental problems that is met when causal inferences are made about the effects of a certain resource variable is

that the validity of the inference is threatened by the problem of omitted variables. When investigating the effects of variations in class size, say, there is no guarantee that comparable groups of students were assigned to small and large classes, which implies that any effect that the class size variable may have is mixed up with pre-existing differences between students. It is, in principle, possible to make causal inferences from non-experimental data through statistical modeling, but one fundamental assumption that must be fulfilled for this approach to allow valid inference about causal effects is that all the relevant variables are included in the analysis, i.e., there must not be any omitted variables which are correlated with the independent variables and the residual of the dependent variable. It is quite obviously impossible to guarantee that a particular study has included all the relevant variables, but it is often easy to see that a study has failed to include variables which have been demonstrated to be important determinants of the dependent variable in other studies.

In this respect, the Coleman report suffered from the obvious limitation that the variables were measured at a single point in time, which made it impossible to control for what is known to be the most important determinant of student achievement, namely previous student achievement. This also made it necessary to make the assumption that only the level of resources available at the time when the measures were taken is important, while the history of resources does not matter. As has been argued by Ehrenberg, Brewer, Gamoran and Willms (2001), this is a very strong assumption, which is likely to be incorrect. Suppose, for example, that class size in third grade is related to gain in achievement, and that a relation is found. But students who are in a small class in grade three are likely to have been in a small class in the first and second grades as well. If there is a prolonged effect on achievement in third grade of having attended a small class in first and second grades this will, incorrectly, be attributed to class size in third grade, unless measures of class size in first and second grades are also included in the study.

Another problem that is met when educational production functions are estimated from data is that errors of measurement in the independent variables will tend to cause bias in the estimates (Wansbeek and Meijer, 2000). Even a seemingly simple variable such as class size is quite difficult to define and measure. Sometimes, pupil/teacher ratio and class size are used interchangeably, for which there is little conceptual and empirical justification (see Gustafsson and Myrberg,

p. 61-65). Class size also often varies over subject matters and time, which makes it difficult to measure this variable with precision in survey studies. This implies that there is often a substantial amount of both random and non-random errors of measurement in the variables. It is well known that random errors cause the effects of a variable to be underestimated in multiple regression analysis.

After the Coleman report was published, a large number of educational production function studies have been conducted. These studies have relied on basically the same approach as that used in the Coleman report, even though more care has been taken to include measures of previous achievement, and to investigate change in achievement rather than achievement at a single point in time (Hanushek, 1979). These studies present a bewildering array of findings, there being little consistency of results over the studies. This is not unusual in any area of research, however, and since the 1970s, so-called meta-analytic techniques have been developed for the purpose of integrating empirical results obtained in different studies into a single unified estimate.

1.2 Reviews of results from educational production function studies

Reviews of the literature have been published by Hanushek (e.g., Hanushek, 1979, 1981, 1986, 1997), and as already mentioned, his conclusion is that economic resources are of no importance for the achievement of schools and students. Analyzing basically the same set of empirical studies, but using other techniques for integrating results over studies, Hedges, Laine and Greenwald (1994; see also Greenwald, Hedges and Laine, 1996a,b; Hedges and Greenwald, 1996) instead concluded that there is quite a strong relationship between economic resources and educational results. They concluded that global resource variables, such as per-pupil expenditure, are important, as are also more specific categories of resources, such as smaller schools and smaller classes. They also concluded that variables that attempt to describe the quality of teachers, such as teacher ability, teacher education and teacher experience show very strong relations with achievement.

Hedges et al. have argued that one reason why Hanushek arrives at the conclusion that economic resources do not matter, is that he has applied a weak method of meta-analysis, which only looks at whether a significant effect was obtained or not. The method does not take

into account size of effects, and if the power of each individual study is low, this so-called “vote-counting” method will fail to detect any effect of resources on achievement. Krueger (2003) has formulated a similar kind of criticism, observing that Hanushek often has included several estimates from the same empirical study, each estimate being computed for a subgroup of the cases. But dividing the sample into subgroups causes the power of each such analysis to be low, making it difficult to find any significant effect. As is demonstrated by Krueger (2003), other results are obtained if a single estimate is computed from each study.

It would seem that the more powerful meta-analytic techniques employed by Hedges et al yield more trust-worthy and dependable results than the simpler technique used by Hanushek. This provides a basis for concluding that the generalization that economic resources are important for educational achievement is more valid than the negative conclusion arrived at by Hanushek. For several reasons, this must be regarded as a very tentative conclusion, which must be supported by further evidence.

One reason for this is that each educational production function study suffers from severe limitations. As was pointed out above, there is reason to believe that every single study that has been conducted has omitted variables that should have been included in order to obtain unbiased estimates of the effects of resource variables. If variables are omitted more or less randomly over studies, a meta-analytic synthesis of findings may still yield correct results. However, there is reason to believe that variables are not omitted randomly, but that certain important variables tend to be more frequently omitted than others. Examples of such variables are the entry achievement level of the students, and their resource history. This may cause systematic bias in the results from educational production function research.

A related problem is that the analyses frequently include inappropriate variables that may obscure relations that are in the data. Furthermore, the models sometimes include more than one measure of what amounts to basically the same resource variable (e.g., expenditure per pupil and student-teacher ratio) which causes the effect for all the overlapping variables to be underestimated (see Krueger, 2003).

The production function paradigm suffers from other limitations and problems as well. One problem is that it treats the educational process as a “black box”, which makes it difficult to understand any particular outcome, and develop integrated theory. The so-called

frame-factor theoretic tradition, which was established by Dahllöf (1967), stresses the importance of investigating the intervening educational process when relating input-factors to educational achievement. This research tradition may provide useful theoretical tools and methods in research which aims at opening up the “black box”. Similarly, Monk (1992) has argued that the production function approach reduces the complexity of the educational process so far as to have limited explanatory value.

A related problem is that the production function approach tends to disregard the hierarchical, or multi-level, character of educational data. The data structure is such that students are nested within classrooms, which are nested within schools, which are nested within districts, and so on. This problem has typically been dealt with by aggregating the observations on lower-level units (e.g., students) to higher-level units (e.g., schools). Such aggregation may, however, change the meaning of variables and introduce bias in the estimates of parameters. What is even more important is that the approach disregards essential features of the educational phenomena, and makes it impossible or difficult to investigate certain questions, such as differential effects of resources on different categories of students. During the 1990s, multi-level regression techniques (e.g., Bryk and Raudenbush, 1998, 1992) have become available which allow ways of taking full advantage of multi-level data (see Ehrenberg et al. 2001).

Wenglinsky (1998) has reported a study using such methods of analysis. He investigated effects of resources at the school district level on the mathematics achievement of grade 12 students and found essentially no general effects at the school level of categories of resources, such as per pupil expenditure on instruction. But the multi-level analysis disclosed an interaction effect such that in school districts with a low level of resources, there was a stronger relationship between the socio-economic background of students and achievement than was the case in school districts with a high level of resources. Wenglinsky interpreted this in the following way:

...when schools lack sufficient funds, their capacity to educate all students toward a common yardstick is reduced. Students enter high school with different levels of preparation, depending on their SES and various other factors. To reduce these inequalities in preparation to the point that both low- and high-SES students become proficient in the requisite subject matter requires the active intervention of the school; when the school lacks adequate funds, its ability to intervene is compromised and, as a result, students will be more likely to advance based on their past preparation -- that is, a situation of within-school inequity. (s. 279)

There is much more that could be said about the educational production function approach, but there is now reason to return to the question of what conclusions may be drawn from the meta-analytic syntheses of findings. I have already concluded that there is reason to put more faith in the conclusions drawn by Hedges et al. than by Hanushek, but that the limitations from which the individual studies suffer make this a very tentative conclusion indeed. To reach stronger conclusions, we need a more solid foundation of research than is furnished by the educational production function studies. Fortunately, a wealth of interesting research has recently been published in which those resource factors which seem most important according to the results achieved by Hedges et al. have been investigated in greater detail. One of these is class size, and the other is teacher competence. Below, this research is reviewed, but before doing so, I will briefly mention another recent line of research which is of particular interest, both because of the amount of attention it has received, and because it has investigated earnings rather than achievement as an outcome of schooling.

2. Quality of schooling and earnings

In the early 1990s, Card and Kreuger (1992) published an investigation in which they combined income data for all American men born 1920 to 1949 with information about school quality of the state in which the individuals went to school. In particular, the analysis took advantage of the information about earnings of men who had moved to another state. The results showed quite strong effects of the quality of education within different states on the amount of earnings in adult life. Longer years of schooling, more teachers per pupil and higher teacher salaries proved to be significant determinants of earnings.

Other studies have, however, yielded different results. Betts (1995) used data from a U. S. longitudinal study which included a representative sample of persons aged 14 to 24 years at the first wave of measurement (National Longitudinal Survey of Youth, NLSY). Among other variables, the survey included information about the average class size of the high school the persons had attended. The sample was followed until they had reached ages 25 to 35, when earnings data were collected as well. Betts could thus perform essentially the same analyses as did Card and Kreuger, except that he had access to infor-

mation from the schools actually attended, rather than the state-wide average.

The analysis failed to show any significant relation between school quality measures and earnings, even though it was found that earnings depended upon which school the men had attended. But interestingly enough, Betts found that when the school-level variables were replaced with the statewide average variables, several significant relationships emerged, including a highly significant relation between the state-level pupil-teacher ratio and earnings.

There are several possible interpretations of this finding. One possibility, which was suggested by Betts (1996, p. 242), is that the state-level teacher-pupil ratio reflects other differences between states, such as school quality in lower grades, which may have a stronger effect on subsequent earnings than the high-school quality measures used. Some other possible interpretations suggested by Betts is that the results obtained with the state-level measure is a statistical artifact caused by the aggregation; that there is a smaller difference in school quality in these newer data; or a diminishing return of school quality on earnings. Another possibility, which was emphasized by Card and Kreuger (1996), is that there are errors of measurement in the school-level measures of teacher-pupil ratio. This is because school quality was measured at a single point in time, which was not necessarily aligned with the time the individual in the sample attended the high-school. Card and Kreuger (1996) note that the resources of a school vary from time to time, and each individual in the sample may, furthermore, have attended more than one high-school. This causes random fluctuations, or errors, in the measure of the school quality measure, which in turn causes the estimate of the effect of school quality to be biased downwards. Since the state-level measure is an average measure of school-quality it is not so much affected by these random errors because they cancel. Card and Kreuger (1996) also suggest other possible explanations for the different results obtained by them and by Betts (1995) but discussing these here would carry too far.

At a more general level it seems, however, that these studies do illustrate a very fundamental problem in the study of natural variation of school resources, namely that the variation of the independent variable is captured very crudely and is only distantly related to the actual instructional experiences of the individuals.

There is, first of all, the problem that an administrative measure like teacher-pupil ratio is relied upon, rather than class size. The latter measure refers to the actual number of pupils taught together at a particular time, while the former measure is based on full-time equivalents of teachers at the school, whether they teach regular classes, or fulfill functions such as administrators, assistants, special educators, librarians, or even are on sick leave. These two measures need not be particularly highly correlated, and while the teacher-pupil ratio measure is strongly related to the amount of money spent on each student, the class size measure is more likely to be relevant from a psychological and instructional point of view. When investigating effects of resources on educational achievement, the class size measure seems preferable. This measure is not easily obtained, however. Not only is class size a number which varies over time, but also over different activities and subject areas. Thus "... one would like to have a measure of the actual class size experienced by every pupil during every school day, over the school year" (Ehrenberg et al., 2001, p. 2). Little is known about the reliability of the class size measures used in actual research, but in the few cases when this measure is at all available, it is likely to include a sizeable proportion of error variance. Such error variance causes the estimates to be downwards biased.

After this digression on problems of measurement, there is reason to point out that other problems with the Card and Krueger (1992) study have been identified by Heckman et al. (1996). In a reanalysis of the data, Heckman et al. (1996) replicated several of the main findings of the original study but they also arrived at partially different results, showing less of a clear pattern of relations between school resources and earnings. One of the reasons for this is that they took into account non-linearities in the relation between schooling and income which are caused by the effects of taking a degree (so-called "sheepskin" effects), showing a strong positive effect of completing college (see also Card and Krueger, 1996). Heckman et al. conclude by saying:

The evidence in this chapter, like the evidence in the literature that precedes it, is not decisive on the question whether schooling inputs can increase earnings. All we have done is to raise considerable doubts about the reliability of the evidence on schooling quality based on aggregated quality data (Heckman et al., 1996, s. 254)

Thus, the results on the relation between quality of schooling and outcomes of schooling which seemed so well established 10 years ago,

now appear to be more uncertain. Further research will be needed to resolve these issues.

Before leaving this field of research, there is reason to emphasize, however, that a large body of literature has established a relation between school resources and greater educational attainment in terms of number of years of schooling. It also is a well established fact that there is a relation between the level of educational attainment and earnings (see Card and Kreuger, 1996). Thus, school resources transformed into more years of education imply higher earnings (see also Betts, 1996).

3. Effects of class size

We may conclude that empirical studies of naturally occurring variation are fraught with problems when the aim is to make causal inferences. The problems caused by omitted variables are difficult to solve, and the statistical models used to correct for selection bias are highly demanding, both with respect to data and assumptions related to the method of statistical analysis. Problems of obtaining precise and accurate measures of the variables involved are another great challenge. This concerns both the independent variables under study, and the variables used as control variables.

One way to solve these, and other, problems is to use randomized experiments instead of trying to capture naturally existing variation. The randomization implies that comparable groups of students receive well defined treatments, which are delivered by comparable groups of teachers. The problem of selection bias is solved by the randomization, and the fact that the independent variable is under control by the researcher provides a solution to the problem of errors of measurement in the independent variable. Control of the independent variable also makes it possible to study a wider range of variability than is the case if naturally occurring variation is investigated.

In spite of the great advantages associated with an experimental approach, experiments are only rarely conducted within the field of education. There are several reasons for this. Ethical concerns may make it impossible to actually carry out experiments which involve an unequal distribution of resources to different groups; costs may be prohibitively large; problems of differential attrition may cause selection bias to appear in spite of initially comparable groups; and the impossibility to achieve the ideal of a double-blind experiment may

cause diffusion of treatments and treatment groups and create so-called Hawthorne and John Henry effects. These and other problems explain why large-scale field experiments in education are rare, and to the extent that the experimental approach is used at all, it is typically restricted to small scale experiments of short duration.

However, during the 1990s, the situation has improved. Results have been reported from a large-scale experiment on the effects of class size, which has not only had profound effects on our knowledge about the effects of resources, but which also has stimulated further experimentation. This study is the so-called STAR-experiment (Student/Teacher Achievement Ratio), which Mosteller (1995) described as “one of the greatest education experiments in United States history”.

3.1. The STAR experiment

The experiment, which started in 1985, had three treatment groups: small classes with 13-17 students; regular classes with 22-26 students; and regular classes with an assistant teacher. For a school to be included in the study, it had to be large enough to have at least one class of each type. Some 80 schools participated, with more than 100 classes of each type. During the first year of the study, about 6 000 students were included, and throughout the four years that the study comprised, almost 12 000 students were involved, because of the addition of new students. Within schools, both students and teachers were randomly assigned to the three treatments. Most of the students entered the study either in kindergarten or grade 1, while a few entered in grades 2 or 3. In the first phase of the study, the students were followed till the end of grade 3, with measures of achievement being made at the end of each grade.

At the end of grade 3, the results showed quite a striking advantage for the students who had been assigned to the small classes, while there was no clear difference between the results achieved in regular classes with one teacher, and regular classes with an assistant teacher as well (Finn and Achilles, 1990, 1999). The results were particularly strong for reading, and it was found that the small class advantage was larger for students who came from socio-economically and ethnically disadvantaged groups (Finn and Achilles, 1990; Kreuger, 1999). The results showed the effect of class-type to be strongest for grade 1, while the difference remained more or less constant over grades 2 and 3.

The participants of the STAR-experiment have been followed up through their continuing education as well. These follow-up studies have shown some quite remarkable findings. After the experiment ended in grade 3, all students were put in normal classes, but a lasting benefit of having attended a small class has been demonstrated. Achievement as measured by standardized tests is higher for those who attended a small class in the STAR-experiment (Nye, Hedges and Konstantopoulos, 1999); they have a lower rate of class repetition (Pate-Bain et al., 1997), and a higher tendency to take the SAT and other tests which give access to higher education in the U. S. (Krueger and Whitmore, 2001). These results are even more remarkable given that the usual pattern of results in intervention studies is that any effect on knowledge and achievement that may be observed initially tends to disappear after the treatment is discontinued. That was, for example, the case in the Head Start program (e.g., Brody, 1992, p. 174-175). Krueger and Whitmore (2001) did, however, find that some of the achievement advantage of small-class children faded out in the year they returned to regular-size classes.

It may be added that in addition to the original reporting of results from this study, the STAR-project data have been reanalyzed by several researchers who have used more advanced methods of analysis than were available for the original analysis. These analyses have generally supported the general conclusions drawn in the original analysis, but at a more specific level, there are some quite notable differences in the results. To take one example, all studies that have investigated interactions between the class size treatment variable and ethnicity have found stronger effects of small classes for minority students than for non-minority students. Finn and Achilles (1999), for example, report an effect size for reading achievement in grade 3 of .17 for non-minority students and of .40 for minority students, while the corresponding estimates for mathematics are .16 and .30. The effect sizes are thus two or three times as large for minority students as for non-minority students. Nye, Hedges and Konstantopoulos (2000) also found the point estimates of treatment to be higher for minority students than for non-minority students, although the difference was somewhat less pronounced. In contrast to other researchers, they did not find the effects to be statistically significant. One reason for this is that they used multilevel analysis procedures, which correctly take into account the intra-school covariation in student performance. Another reason may be that in their regression models, they simultaneously

included two highly correlated variables (minority status and socioeconomic status) which may have caused the standard errors to escalate. This illustrates that, even with such a large sample as in the STAR experiment, power may be too limited to investigate interactions in complex models.

Another source of disagreement between analysts is whether the mean performance difference due to the class size effect, which was observed after grade 1, keeps constant over the following grades, or if it increases. Hanushek (1999) has argued that if there is a class size effect, the achievement difference between small- and regular-sized classes should increase over grades, because the small classes should keep adding a resource-related performance advantage. According to Finn and Achilles (1999), this is indeed the case, which is seen if the scores on the standardized achievement tests are rescaled to express “grade equivalents”. However, the traditional effect size measures are more or less constant over the years, which provides a basis for the conclusion that there is no additional class size effect after the first year. So far, this issue seems to be unsettled. These conflicting results illustrate, however, that the scale chosen for reporting the results may influence the outcome. Different studies use different scales (e.g., percentile ranks, raw scores, grade equivalents) so this may be a source of conflicting results, which should be given some attention in further research.

Another topic of discussion among critics and defenders of the STAR experiment is whether any bias in the results may have been caused by the impossibility to carry out a four-year field study with strict randomization (Hanushek, 1999; Hoxby, 2000). There was attrition from the experiment because students moved, came into the experiment successively over the years, and there was also considerable switching between small and large classes within the experiment. Only 48 percent of the initial experiment group remained for the entire four years. It seems, though, that attrition patterns were similar across small and large classes, and the only tendency towards differential attrition that has been found is that students dropping out of small classes had somewhat higher achievement than students dropping out of large classes (Nye, Hedges and Konstantopoulos, 1999). Thus, differential attrition does not seem to account for the small-class advantage in the STAR study (see also Krueger, 1999).

It is not possible to evaluate all evidence concerning the internal validity of the conclusions from the STAR study. But it may be noted

that it is impossible to carry out a large-scale field-experiment strictly according to the methodological rules. Even though initial assignment may be random, this can not be strictly kept throughout the duration of the study because of movements. Another problem with this kind of experiment is that those involved are not only aware that they participate in an experiment, but also in which treatment condition they participate. This may cause teachers and students to modify their teaching and learning activities in order to produce an outcome that they favor, or it may sensitize teachers of small classes to more fully take advantage of the instructional possibilities offered by such an environment (Hoxby, 2000). It is also conceivable that teachers and students (and their parents) in regular-size classes put in an extra effort to overcome a perceived disadvantage (a so-called John Henry effect).

It is obvious that even a randomized experiment like the STAR-study is sensitive to threats to the internal validity. In theory, at least, these threats may be severe enough to reject interpretations in terms of effects of class size but so far, little evidence has been presented that provides obvious threats to the main conclusions from the study. I will therefore tentatively accept the findings as showing a real effect of class size on achievement.

3.2. Replications and implementations

The results from the STAR experiment have, indeed, been widely accepted, even though there are also skeptics (e.g., Hanushek, 1999, 2003; Hoxby, 2000). In particular, the results have inspired further experimentation and they have had a noticeable impact on educational policy at different levels of the U. S. educational system. Let me just briefly mention a few of these.

In Tennessee, there has been a follow up of the STAR-study in the form of an implementation project (Project Challenge) in which the 17 poorest school districts were funded to cut class size to 15 in grades K-3 (Finn, 1998). When the project started in 1990, these school districts ranked 99 out of 138 in reading, but in 1993, the rank had improved to 78. In mathematics, the rank was 85 in 1990, which improved to 57 in 1993. These results indicate that class size reduction may have a substantial practical impact. It should be stressed, though, that from a scientific point of view, these results should be regarded as providing light-weight evidence on the class size issue,

since the improvement may have been due to regression effects and/or Hawthorne effects.

In 1996 a trial program similar to STAR, but with a special focus on disadvantaged students, was started in Wisconsin. The so-called SAGE (Student Achievement Guarantee in Education) program was focused upon grades K-3 in school districts where at least 50 percent of the students were living below the poverty level. The program involved, among other things, reduction of class size to 15 or fewer students. Comparisons between the schools participating in the intervention and similar schools from the same districts with normal class sizes indicate effects on achievement that are roughly comparable to those from the STAR project (Molnar et al., 1999). Interestingly enough, the effect sizes for African-American students were greater than for white students, thereby replicating the interaction between class size and minority status in the STAR experiment.

The STAR results have also inspired several other U. S. states to implement class size reduction programs. In 1996, California started a massive state-wide class size reduction program in grades K-3, cutting class size from an average of 28.8 to a maximum of 20 (CSRP, "Class Size Reduction Program"). In contrast to the STAR and SAGE programs, the program was under-funded and conditions were less than optimal also in other respects: there was a growing enrollment of students, a shortage of qualified teachers, and a lack of adequate facilities. The evaluation of CSRP also was less carefully planned than the evaluation of STAR, there being no control groups. However, at the end of third grade, classes with 20 or fewer students have been compared with classes with more than 20 students, using school means at grade 4 to control for pre-existing differences. The results indicate a weak positive effect, with an effect size of about 0.05-0.10 (Stecher and Bohrnstedt, 2000). In contrast to the findings in STAR and SAGE, the effects were not stronger for minority students than for non-minority students.

As has already been mentioned, there are several possible reasons why smaller effects were achieved in CSRP than in STAR. One is that CSRP class sizes were reduced to about the same level that was regarded a regular-sized class in STAR, rather than to 15. The program also made it necessary to hire many new teachers without credentials or experience, and to discontinue other programs in order to fund the class size reduction. Furthermore, poor school districts and school districts with a high percentage of minorities tended to lose their

qualified teachers to more affluent districts, so that the program widened the resource gap among schools (Ogawa and Huston, 1999). This may be one reason why the effects of the class size reduction did not prove to be stronger for educationally disadvantaged students in the CSR program. The experiences from CSR clearly indicate that class size reduction does not automatically lead to improvement of achievement.

Even in light of the CSR experiences, it seems that the experimental studies of effects of class size indicate that smaller classes are beneficial for learning. It must be emphasized, though, that these results are based upon studies of class size effects in grades K-3, and that the studies have typically involved samples with an overrepresentation of educationally disadvantaged students. The other research that has been done on the effects of class size tends to support these findings. Robinson (1990) has reported a large meta-analysis, which comprises more than 100 studies of class size. The results indicate that the effects of class size interact with the age of the students. For grades K-3, Robinson found a positive effect of small classes. For grades 4-8, a weak positive effect was found "... but the evidence is not nearly as strong as in grades K-3" (Robinson, 1990, p 84). Studies conducted on students from grades 9-12 provide no support for the hypothesis that class size has an effect on achievement. Robinson (1990, p. 85) also concluded that:

[t]he research rather consistently finds that students who are economically disadvantaged or from some ethnic minority perform better academically in smaller classes.

3.3. Natural variation in class size

As has already been pointed out, several researchers have argued that the fact that those involved in class size experiments are aware of this may cause the participants to modify their actions and activities, so that a certain outcome is favored (Hanushek, 1999; Hoxby, 2000). Given that double-blind experimentation is impossible in this field of research, this is a limitation of the experimental approach.

A few studies have been conducted in which this problem is instead solved by investigating natural variation in class size. Angrist and Lavy (1999) conducted a study in Israel, where there is a quite strict rule for the maximum number of students in a class, and which implies splitting one large class into two smaller classes. The variation in class size close to the maximum class size is essentially unrelated to

factors such as the socio-economic status of the area from which the school is recruiting, which implies that the variation in class size caused by the splitting rule can be used to estimate effects of class size. Using so-called instrumental variable estimation on data from some 2000 fourth and fifth grade classes, class size was found to significantly affect the achievement in reading and mathematics in grade 5, and in reading in grade 4, with smaller classes producing the better results. The effect sizes were somewhat lower than those found in the STAR experiment. Just like in the STAR experiment, Angrist and Lavy (1999) also found that there was an interaction between socio-economic background and class size, the benefits of small classes being larger in schools with a large proportion of students from a disadvantaged background.

Hoxby (2000) took advantage of the fact that natural variation in population size influences class size, and this variation causes random variation in class size which is not associated with any other variation, except perhaps achievement. She also used a similar approach as did Angrist and Lavy (1999), investigating the abrupt changes in class size caused by rules about maximum class size. Using data from 649 elementary schools covering a period of 12 years, both these approaches to estimating the effects of class size were used. In no case was a significant effect of class size found, in spite of the fact that Hoxby demonstrates that power was sufficient to detect class size effects as small as those found in the experimental research. Hoxby (2000) suggests that the differences between the results obtained in the experimental studies and her studies of natural variation may be interpreted as being due to the fact that teachers in the experimental studies tried to make good use of small classes, because an outcome showing an advantage for small classes would be favored by the teachers.

Lindahl (2000) has used a longitudinal approach to investigate effects of natural variation in class size. He administered the same mathematics test to 556 students in 16 schools on three occasions. The first measurement occasion was in the Spring semester of grade 5, the second in the Autumn semester of grade 6, and the third in the Spring semester of grade 6. In the analysis of data, Lindahl's idea was to take advantage of the fact that between the first and second measurements, there was a summer break, when no effect of class size is to be expected, while between the second and third measurements, there was almost a full academic year, when effects of class size may be expected. In the analysis of data, he investigated change between the

second and third measurements, controlling for change between first and second measurements. The results showed smaller classes to yield significantly better results than large classes.

So far, the studies investigating natural variation in class size have thus produced conflicting results. The research base is limited, though, and the conflicting results will hopefully spur further research efforts.

3.4. Why may smaller classes be better?

With the exception of the Hoxby (2000) study, the empirical results present a fairly simple and clear-cut picture, according to which small classes are beneficial in the first grades of school, particularly so for educationally disadvantaged children. According to the STAR findings, the beneficial effect of attending a small class also seems to be a lasting one. It is, however, not at all clear how this effect may be interpreted.

If we can observe an effect, it also is necessary to propose a theory which explains the mechanism through which the effect is created. There are three reasons for this (Ehrenberg et al., 2001). The first is that if we can explain the effect, this increases our confidence that the effect is a real one, rather than an artifact. The second is that typically, effects of a resource variable are contingent upon other factors, which reduce or enhance the effect of the resource. If we can explain the mechanism, we are also in a better position to identify and understand such interactive effects. The third reason why we need to understand mechanisms is that the factors we investigate in empirical research are typically crude approximations of the optimal design. Understanding the mechanism may help us improve the use of resources.

Three tentative theories, or categories of theories, which try to account for the class size effect may be identified. The first theory basically says that achievement in smaller classes is higher because instructional quality is higher in smaller classes than in larger classes (e.g., Achilles, 1999; Smith and Glass, 1980). The higher instructional quality could be realized because teachers take advantage of the smaller number of students and adopt other teaching strategies. They could, for example, use more frequent assignment of writing tasks, more small-group work, more discussion and problem solving, more individual diagnostics and help, and other ways of furthering the development of each individual student. But the instructional quality theory runs into problems when confronted with empirical results. Research

on the relation between class size and instructional strategies indicates that class size has little effect on teaching. Ehrenberg et al. (2001) concluded on the basis of a review of the literature that:

Overall, the weight of the evidence tilts strongly toward a conclusion that reducing class size, by itself, does not typically affect the instructional activities that occur in classrooms. (p. 23)

It must, of course, be admitted that there may be subtle differences in instructional quality between small and large classes which have gone undetected in these studies, which somewhat weaken the strength of this argument. But there are other problems with the instructional quality theory as well. If instructional quality is directly or indirectly affected by class size, one would expect the effect of class size to be more or less independent of student age and background. Class size does not seem to be an important factor after fourth grade, but there is certainly no reason to believe that instructional quality would cease to be important after fourth grade. The explanation in terms of instructional quality also fails to account for the long-lasting class size effects observed in the STAR follow up studies. Intervention studies designed to increase quality of instruction may have substantial effects on basic skills of young children, but these effects typically vanish over time. It seems that the instructional quality theory fails to account for the empirical patterns of results.

The second category of theories to be discussed focusses on the classroom environment and student conduct in classes of different size. The most elegant and elaborate theory is Lazear's (2001) disruption model of educational production. The basic idea of this theory is that there is a probability that any given student at any moment in time will not cause a disruption (p), for example by misbehaving or asking a question which is of relevance to this student only. Thus, p expresses the proportion of time that a student does not prevent all students from learning. This implies that in a class of size n , the probability that a disruption will occur at any time is $1 - p^n$. It may be observed that even when p is as high as .98, a class of 25 students causes disruption to occur 40 per cent of the time. With a smaller number of students in the class, the probability of disruption is lower for a given value of p and with a larger number of students, the probability is higher. The fact that there is a cost associated with class size, primarily for teachers and facilities, implies that there is an optimal class size, which varies as a function of p and cost. As is shown by Lazear (2001), more well-behaved students (i.e., a higher p) imply a larger op-

timal class size, as does a higher cost for each class. The model also predicts that there will be a tendency for more well-behaved students to be in larger classes, which tends to create the impression that larger classes are conducive to achievement. The fact that class size tends to be smaller for younger students may be explained referring to the fact that p tends to be lower for younger students.

This model can account for several of the findings in research on class size. It implies that class size reduction will have stronger effects on groups of students with a lower p than those with a higher p , which may explain why class size has a stronger effect in grades K-3 than in higher grades. Assuming that the value of p is lower for minority students than for non-minority students, this also explains why the effect size is larger for minority students than for non-minority students. It seems, however, that Lazear's disruption model fails to account for the lasting effects of class size found in the STAR experiment. The positive effects on learning for young students that are predicted in the model are not in themselves sufficient to explain the long-lasting effects. While the disruption theory is a powerful model for explaining a wide array of results from research on class size, it does not succeed in fully explaining them.

The third category of theories to account for the beneficial effects of class size in the early grades focuses on the better possibilities of a teacher with a small group of students to socialize the children to the school setting. Biddle and Berliner (2002) said:

In the early grades, students first learn the rules of standard classroom culture and form ideas about whether they can cope with education. Many students have difficulty with these tasks, and interactions with a teacher on a one-to-one basis—a process more likely to take place when the class is small—help the students cope ... Learning how to cope well with school is crucial to success in education, and those students who solve this task when young will thereafter carry broad advantages—more effective habits and positive self-concepts—that serve them well in later years of education and work. (p. 20)

This school socialization theory (see also Krueger, 1999) emphasizes the acquisition of skills and habits which makes it possible for students to cope with the requirements of life in school. This may be expected to have both short-term and long-lasting effects on achievement. It is, furthermore, reasonable to expect that the socialization effect is stronger for educationally disadvantaged groups, which accounts for the higher effect for minority students than for non-minority students.

The school socialization theory accounts for all major findings of research on class size, but it does not have the elegance of Lazear's disruption theory. These theories do not seem to be mutually exclusive, however, and they may both be true. Expressed in the language of the disruption theory, the school socialization theory may be said to account for the class size effect through a direct influence on p . This suggests that it may be an interesting extension of the disruption theory not to regard p as a fixed parameter, but instead as a dependent variable that may be influenced by different factors, including class size.

4. Teacher competence

Hanushek (2003) reviewed studies which estimate differences in teacher quality through the use of fixed effects models of student performance, in which entry performance and other relevant factors are taken into account (e.g., Hanushek, 1992; Hanushek, Kain and Rivkin, 1998; Murnane and Phillips, 1981). These studies have uncovered important variations in teacher quality, with estimates of effect sizes that are very large as compared to those associated with other factors, such as class size. According to Hanushek (2003), a conservative estimate implies that a one standard deviation change in quality leads to a .11 standard deviation increase in achievement. According to another estimate, students of the best teachers gain 1.5 grade level equivalents for a single academic year, while students of the least well-performing teachers only gain 0.5 grade level equivalents in the same time.

These results establish variations in teacher competence as the single most important resource factor in determining student achievement, and there does seem to be little disagreement about this. There is, however, considerable controversy about whether such unmeasured variability in teacher quality may be influenced by factors such as teacher education, and if it is possible to find indicators that measure teacher quality. Hanushek (1986, 1997) found no evidence that factors such as teacher experience or teacher education are systematically related to student achievement. However, in their meta-analyses, the Hedges group (Greenwald, Hedges and Laine, 1996a, b; Hedges and Greenwald, 1996; Hedges, Laine and Greenwald, 1994) found that variables that describe the quality of teachers, such as teacher ability, teacher education and teacher experience show very strong relations

with achievement. It may also be noted that Hanushek (2003) with his “vote-counting” technique found that teacher experience tended to have a positive relation with achievement in a fairly large proportion of studies, even when controlling for study quality. Given the great impact of the teacher quality factor, there is reason to take a somewhat closer look at the research on different aspects of teacher competence.

Teacher education has been intensely debated, and one of the questions that has been focused upon is whether subject matter competence or pedagogical competence, if any, matters most. Several studies support the importance of subject matter knowledge for successful teaching, particularly in higher grades (e.g., Darling-Hammond, 1999; Monk, 1994; Ferguson and Womack, 1993). It seems that the relationship is curvilinear, with diminishing returns from education at the Master’s level (Monk, 1994; Hanushek, Kain and Rivkin, 1998). The amount of pedagogical education seems to be equally important, however (Darling-Hammond, 1999; Evertson, Hawley, and Zlotnick, 1985; Ferguson and Womack, 1993). This suggests that the quite popular idea that anyone who knows a subject matter content can also teach it is incorrect.

Darling-Hammond (2000) discussed what makes teacher education effective and she concluded:

One of the great flaws of the “bright person myth” of teaching is that it presumes that anyone can teach what he or she knows to anyone else, people who have never studied teaching or learning often have a very difficult time understanding how to convey material that they themselves learned effortlessly and almost subconsciously. ... Furthermore, individuals who have had no powerful teacher education intervention often maintain a single cognitive and cultural perspective that makes it difficult for them to understand the experiences, perceptions, and knowledge bases that deeply influence the approach to learning of students who are different from themselves. The capacity to understand another is not innate; it is developed through study, reflection, guided experience, and inquiry. (p. 170-171)

Teacher experience, measured in terms of number of years of teaching practice, is also related to student achievement according to several studies (Murnane and Philips, 1981; Klitgaard and Hall, 1974; Hanushek, Kain and Rivkin, 1988). Here too, the relation is curvilinear, with a diminishing return of additional years. According to Darling-Hammond (1999), there is little experience from contribution beyond five years, while Hanushek, Kain and Rivkin (1998) found that after two years, there was little improvement.

It must, of course, be emphasized that these studies of effects of teaching experience suffer from methodological difficulties. The results may be influenced by bias because the less successful teachers leave the teaching profession, or because more skilled teachers tend to be recruited by schools which are situated in affluent areas. As we have seen, it is difficult to control for such omitted variable bias with statistical methods, so results from these studies must be regarded as tentative.

Measures of different aspects of teacher ability have also been shown to correlate with student achievement. This has been found for teachers' verbal ability (Ehrenberg and Brewer, 1996) and measures of teachers' knowledge and skills (Carroll, 1975; Strauss and Sawyers, 1986). One such study has been reported by Strauss and Sawyers (1986), who investigated high schools in about 100 school districts. Among other measures of resources, such as per pupil expenditure and pupil-teacher ratio, they had access to a measure from the National Teacher Evaluation of teachers' knowledge. They found a relation between teacher competence and student achievement, but above all, they found a strong effect on the drop out rate. They concluded:

Of the inputs which are potentially policy-controllable (teacher quality, teacher numbers via the pupil-teacher ratio and capital stock) our analysis indicates quite clearly that improving the quality of teachers in the classroom will do more for students who are most educationally at risk, those prone to fail, than reducing the class size or improving the capital stock by any reasonable margin which would be available to policy-makers. The size of this differential impact among inputs is enormous ... teachers matter far more than has been previously documented by other researchers in the field. (Strauss and Sawyer, 1986, p 47)

In a similar study, Ferguson (1991) investigated effects on achievement of different kinds of resource measures, such as pupil-teacher ratio and school size, in 900 school districts in Texas. The study included scores on a teacher certification test measuring both subject matter and pedagogical knowledge. The results indicated a strong effect of teacher competence, which at the school district level was even stronger than the effect of socio-economic background. For lower grades, effects were found of pupil-teacher ratio and school size, but these were small compared to effects of teacher competence. Another interesting finding was that investments in in-service teacher training had effects on student achievement.

These studies rely on natural variation, so there is a risk that the results are influenced by omitted variable bias. It may, for example, be

that teachers whose students are performing well, are allowed to participate in in-service training as a kind of gratification. Angrist and Lavy (1998) have conducted an experimental study in Israel of the effects of in-service teacher training. This study took advantage of the fact that in 1995, some schools in Jerusalem obtained funds earmarked for on-the-job training, which allowed a matched control-group design. The in-service training program was designed to improve the teaching of language skills and mathematics, and had its focus on pedagogical skills. The experiment involved both religious schools and non-religious schools. The treatment effects on students' test scores were analyzed by several different methods, and for the non-religious schools a quite strong effect, with effect sizes between .2 and .4 standard deviations was found. For the religious schools, the results were less clear-cut, which according to Angrist and Lavy may be due to the fact that the program started later there and was implemented on a smaller scale. On the basis of the results obtained in the non-religious schools, Angrist and Lavy compared costs and expected outcome of three different methods for improving student achievement: reducing class size, lengthening the school day, and in-service training. They conclude that in service training is a less expensive method for raising student achievement than reducing school size or adding school hours.

The brief review of studies presented here suggests that there are important relations between different indicators of teacher competence and student achievement. This seems to be true for teacher education, experience, measured knowledge and skills, and in-service training. Most of the studies have investigated natural variation, and there are few well-controlled experiments, so there is reason to be somewhat cautious and conduct further research in this field.

Having concluded that teacher competence is very important indeed, there is reason to briefly consider what is the nature of teacher competence. I have already quoted Darling-Hammond's (2000) statement that the main function of teacher education is to learn to take the perspective of the learner. In a more general analysis of the nature of teacher competence, Darling-Hammond summarized the research in the following way:

... teachers who are able to use a broad repertoire of approaches skillfully (e.g., direct and indirect instruction, experience-based and skill-based approaches, lecture and small group work) are typically most successful. The use of different strategies occurs in the context of "active teaching" that is purposeful and diag-

nostic rather than random or laissez faire and that responds to students' needs as well as curriculum goals (Darling-Hammond, 1999, p. 14)

She also observed that it does seem reasonable that teachers' abilities to handle these tasks are likely to be influenced by factors such as verbal ability, subject matter knowledge, understanding of teaching and learning, and experience in the classroom, as has been found in empirical research.

5. Discussion and conclusions

So far, the discussion has been focussed on whether or not there is evidence that different resource factors affect student achievement. In the final discussion, I will briefly discuss policy implications and interrelations among different kinds of resources.

The results indicate that there is an effect of class size on achievement in lower grades, which is stronger for educationally disadvantaged students than for other students, and which seems to be long-lasting. For the former category of students, the effect size due to a reduction in class size from 22 to 15 seems to be larger than .20 standard deviations, while for the latter category, effect sizes seem a bit lower than 0.20. This result has been replicated several times, and good theoretical explanations may be provided for the class size findings. In Tennessee, and at other sites, it has been successfully demonstrated that the implementation of class size reduction policies does have the anticipated effects on student achievement. But all attempts at implementation have not been successful. The CSR project in California indicates that unless there is adequate funding, availability of trained teachers, and access to facilities, the implementation process may be frustrating and the results may be disappointing. It should also be added that the effect sizes of even quite large class size reductions are relatively small, and that class size reduction is associated with considerable cost. This suggests that careful consideration of alternative ways of spending schooling funds should be made.

The results indicate that among the resource factors, teacher competence is the single most powerful factor in influencing student achievement, and the effect sizes seem to be substantially larger than those associated with class size. As a policy variable teacher, competence is considerably more difficult to manipulate than is class size, and except for the Angrist and Lavy (1998) study of teacher in-service training, I am not aware of any study that pits class size reduction

against increasing teacher competence. Given the strength of effects associated with teacher competence, it would seem that investments in teacher competence would have a higher likelihood of paying off in terms of student achievement than would other investments, which was also the conclusion drawn by Angrist and Lavy (1998).

Murnane (1995) concluded that salaries and working conditions are the two most important factors determining teacher supply, but also that "... almost no studies shed light on the quality of the teaching force" (p 318). It may, nevertheless, be assumed that factors which cause the teaching profession to be attractive will not only affect teacher quantity, but also teacher quality.

The supply of teachers is, among other things, affected by the demographics of the population. As observed by Murnane (1995), trends in birthrates influence the number of potential teachers, and the likelihood of finding a teaching position. The supply of teachers is, furthermore, determined by the career decisions of prospective, current and former teachers, and by the health status and activity level of current teachers.

There is considerable evidence that salaries are important influences on decisions to become a teacher (e.g. Dolton, 1990), and in determining the attrition rate of novice teachers (Murnane et al., 1991). However, because the effect of salaries is based upon comparisons with the available alternatives, what is a competitive salary in one subject field, need not be so in another.

Little is known about the impact of working conditions on career decisions and activity levels in the field of teaching. Murnane (1995, p. 318) stated that :

... teachers care about difficult-to-measure variables such as the availability of materials, and the quality of administrative support. As a result, there is almost no solid evidence on the impact of working conditions on teacher supply in industrialized countries.

However, while it may be true that there is little solid evidence about the impact of working conditions on teacher supply, the generalization about what teachers care about may be challenged on the basis of research on teachers' perceptions of effects of class size on working conditions. On the basis of a review of the literature, Granström (1998) concluded that larger classes involve more hours of work each week, that they are associated with higher levels of stress, and a larger incidence of burnout syndromes. These results indicate that class size is an important factor in determining working condi-

tions for teachers. Furthermore, since class size is such an easily identifiable variable, it is likely to efficiently communicate the working conditions of the teaching profession to potential teachers.

Class size may thus be an important policy variable which affects the supply and quality of teachers. While the direct effects of class size on student achievement may be too weak to justify class size reductions, the indirect effects via the influence on teacher competence may provide a justification for class size reduction. At present, this is, of course, nothing but a hypothesis but given the available empirical evidence, it would seem to be an interesting and important topic for further research.

References

- Achilles, C. (1999), *Let's Put Kids First: Finally Getting Class Size Right*, Thousand Oaks, Corwin Press, California.
- Angrist, J.D. and Lavy, V. (1998), Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools, NBER Working Paper 6781, National Bureau of Economic Research.
- Angrist, J.D. and Lavy, V. (1999), Using Maimonides' rule to estimate the effect of class size on scholastic achievement, *Quarterly Journal of Economics* 114, 533-575.
- Betts, J. (1995), Does school quality matter? Evidence from the National Longitudinal Survey of Youth, *Review of Economics and Statistics* 77, 231-250.
- Betts, J. (1996), Is there a link between school inputs and earnings. Fresh scrutiny of an old literature, in G. Burtless (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, D.C.
- Biddle, B.J. and Berliner, D.C. (2002), Small class size and its effects, *Educational Leadership* 59, 12-23.
- Bowles, S. and Lewin, H. (1968), The determinants of school achievement: An appraisal of some recent evidence, *Journal of Human Resources* 3, 3-24.
- Brody, N. (1992), *Intelligence*. Wiley, New York.
- Bryk, A.S. and Raudenbush, S.W. (1988), Toward more appropriate conceptualization of research on school effects: A three-level hierarchical linear model, *American Journal of Education* 97, 65-108.
- Bryk, A.S. and Raudenbush, S.W. (1992), *Hierarchical Linear Models. Applications and Data Analysis Methods*, Sage Publications, Inc. Newbury Park, California.

WHAT DO WE KNOW ABOUT EFFECTS OF SCHOOL RESOURCES ON
EDUCATIONAL RESULTS?, Jan-Eric Gustafsson

- Burtless, G. (ed.) (1996), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, D.C.
- Card, D. and Krueger, A.B. (1992), Does school quality matter? Returns to education and the characteristics of public schools in the United States, *Journal of Political Economy* 100, 1-40.
- Card, D. and Krueger, A.B. (1996), Labor market effects of school quality: Theory and evidence, in G. Burtless (Ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, D.C.
- Carroll, J.B. (1975), *The Teaching of French as a Foreign Language in Eight Countries*. John Wiley and Sons, New York.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartlant, J., Mood, A.M., Weinfeld, F.D. and York, R.L. (1966), *Equality of Educational Opportunity*, US Government Printing Office, Washington, D.C.
- Dahllöf, U. (1967), *Skoldifferentiering och Undervisningsförlopp*, Almqvist and Wiksell, Stockholm.
- Dahllöf, U. (1999), Det tidiga ramfaktorteoretiska tänkandet, *En tillbakablick*, *Pedagogisk Forskning i Sverige* 4, 5-29.
- Darling-Hammond, L. (1999), Teacher quality and student achievement: A review of state policy evidence, Working Paper, Center for the Study of Teaching and Policy, University of Washington.
- Darling-Hammond, L. (2000), How teacher education matters, *Journal of Teacher Education* 51, 166-173.
- Dolton, P. (1990), The economics of UK teacher supply: The graduate's decision, *Economic Journal* 100, 95-104.
- Ehrenberg, R.G. and Brewer, D.J. (1995), Did teachers' verbal ability and race matter in the 1960s? Coleman revisited, *Economics of Education Review* 14, 291-299.
- Ehrenberg, R.G., Brewer, D.J., Gamoran, A. and Willms, J. D. (2001), Class size and student achievement, *Psychological Science in the Public Interest* 2, 1-30.
- Evertson, C.M., Hawley, W.D. and Zlotnik, M. (1985), Making a difference in educational quality through teacher education, *Journal of Teacher Education* 36, 2-13.
- Ferguson, R.F. (1991), Paying for public education: New evidence on how and why money matters, *Harvard Journal on Legislation* 28, 465-498.
- Ferguson, P. and Womack, S.T. (1993), The impact of subject matter and education coursework on teaching performance, *Journal of Teacher Education* 44, 55-63.

WHAT DO WE KNOW ABOUT EFFECTS OF SCHOOL RESOURCES ON
EDUCATIONAL RESULTS?, Jan-Eric Gustafsson

- Finn, J.D. and Achilles, C.M. (1990), Answers and questions about class size: A statewide experiment, *American Educational Research Journal* 28, 557-577.
- Finn, J. D. and Achilles, C.M. (1999), Tennessee's class size study: Findings, implications and misconceptions, *Educational Evaluation and Policy Analysis* 21, 97-110.
- Finn, J.D., Fulton, D., Zaharias, J. and Nye, B.A. (1989), Carry-over effects of small classes, *Peabody Journal of Education* 67, 75-84.
- Glass, G. V. and Smith, M.L. (1979), Meta-analysis of the research on class size and achievement, *Educational Evaluation and Policy Analysis* 1, 2-16.
- Granström, K. (1998), Stora och små undervisningsgrupper. Forskning om klasstorlekens betydelse för elevers och lärares arbetsituation, FOG rapport no. 37, Institutionen för pedagogik och psykologi Linköping universitet.
- Greenwald, R., Hedges, L.V. and Laine, R.D. (1996a), The effect of school resources on student achievement, *Review of Educational Research* 66, 361-396.
- Greenwald, R., Hedges, L.V. and Laine, R.D. (1996b), Interpreting research on school resources and student achievement, *Review of Educational Research* 66, 411-416.
- Gustafsson, J.-E. and Myrberg, E. (2002), Ekonomiska Resursers Betydelse för Pedagogiska Resultat—en Kunskapsöversikt, Skolverket, Stockholm.
- Hanushek, E.A. (1979), Conceptual and empirical issues in the estimation of educational production functions, *Journal of Human Resources* 14, 351-388.
- Hanushek, E.A. (1981), Throwing money at schools, *Journal of Policy Analysis and Management* 1, 19-41.
- Hanushek, E.A. (1986), The economics of schooling: Production and efficiency in public schools, *Journal of Economic Literature* 24, 1141-1117.
- Hanushek, E.A. (1991), When school finance "reform" may not be a good policy, *Harvard Journal on Legislation* 28, 423-456.
- Hanushek, E.A. (1992), The trade-off between child quantity and quality, *Journal of Political Economy* 100, 84-117.
- Hanushek, E.A. (1997), Assessing the effects of school resources on student performance: An update, *Educational Evaluation and Policy Analysis* 19, 141-164.
- Hanushek, E.A. (1999), Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects, *Educational Evaluation and Policy Analysis* 21, 143-164.
- Hanushek, E.A. (2003), The failure of input-based schooling policies, *Economic Journal* 113, F64-F98.

WHAT DO WE KNOW ABOUT EFFECTS OF SCHOOL RESOURCES ON
EDUCATIONAL RESULTS?, Jan-Eric Gustafsson

- Hanushek, E.A., Kain, J.F. and Rivkin, S.G. (1998), Teachers, schools and academic achievement, NBER Working Paper 6691, National Bureau of Economic Research.
- Hedges, L.V. and Greenwald, R. (1996), Have times changed? The relation between school resources and student performance, in G. Burtless (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, D.C.
- Hedges, L.V., Laine, R.D. and Greenwald, R. (1994), Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes, *Educational Researcher* 23, 5-14.
- Heckman, J., Layne-Farrar, A. and Todd, P. (1996), Does measured school quality really matter? An examination of the earnings-quality relationship, in G. Burtless (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, D.C.
- Hoxby, C.M. (2000), The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics* 115, 1239-1285.
- Klitgaard, R.E. and Hall, G.R. (1974), Are there unusually effective schools?, *Journal of Human Resources* 10, 90-106.
- Krueger, A.B. (1999), Experimental estimates of educational production functions, *Quarterly Journal of Economics* 114, 497-532.
- Krueger, A.B. (2003), Economic considerations and class size, *Economic Journal* 113, F34-F63.
- Krueger, A.B. and Whitmore, D. (2001), The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR, *Economic Journal* 111, 1-28.
- Lazear, E.P. (2001), Educational production, *Quarterly Journal of Economics* 116, 777-803.
- Levin, H. (1988), Cost-effectiveness and educational policy, *Educational Evaluation and Policy Analysis* 10, 51-69.
- Lindahl, M. (2000), Studies of causal effects in empirical labor economics, Ph.D. Thesis, Swedish Institute for Social Research, Stockholm University.
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A. and Ehrle, K. (1999), Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin, *Educational Evaluation and Policy Analysis* 21, 165-178.
- Monk, D.H. (1992), Education productivity research: An update and assessment of its role in education finance reform, *Educational Evaluation and Policy Analysis* 14, 307-332.

WHAT DO WE KNOW ABOUT EFFECTS OF SCHOOL RESOURCES ON
EDUCATIONAL RESULTS?, Jan-Eric Gustafsson

- Monk, D.H. (1994), Subject matter preparation of secondary mathematics and science teachers and student achievement, *Economics of Education Review* 13, 125-145.
- Mosteller, F. (1995), The Tennessee study of class size in the early school grades, *The Future of Children: Critical Issues for Children and Youths* 5, 113-127.
- Mosteller, F. and Moynihan, D.P. (eds.) (1970), *On Equality of Educational Opportunity*, Random House, New York.
- Murnane, R.J. (1981), Interpreting the evidence on school effectiveness, *Teachers College Record* 83, 19-35.
- Murnane, R.J. (1995), Supply of teachers, in M. Carnoy (ed.), *International Encyclopedia of Economics of Education*, 2nd edition Cambridge, Cambridge University Press, UK.
- Murnane, R.J. and Phillips, B.R. (1981), Learning by doing, vintage and selection: Three pieces of the puzzle relating teacher experience and teaching performance, *Economics of Education Review* 1, 453-465.
- Murnane, R.J., Singer, J.D., Willett, J.B., Kemple J.J. and Olsen, R.J. (1991), *Who Will Teach: Policies that Matter*, Harvard University Press, Cambridge, MA.
- Nye, B., Hedges, L.V. and Konstantopoulos, S. (1999), The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment, *Educational Evaluation and Policy Analysis* 21, 127-142.
- Nye, B., Hedges, L.V. and Konstantopoulos, S. (2000), Do the disadvantaged benefit more from small classes? Evidence from the Tennessee class size experiment, *American Journal of Education* 109, 1-26.
- Ogawa, R.T. and Huston, D. (1999), California's class-size reduction initiative: Differences in teacher experience and qualifications across schools, *Educational Policy* 13, 659-694.
- Pate-Bain, H., Boyd-Zaharias, J., Cain, V.A., Word, E. and Binkley, M. E. (1997), *STAR Follow-up studies 1996-1997*, Heros Inc. (<http://www.heros-inc.org/newstar.pdf>).
- Robinson, G.E. (1990), Synthesis of research on the effects of class size, *Educational Leadership* 47, 80-90.
- Smith, M.L. and Glass, G.V. (1980), Meta-analysis of research on class size and its relationship to attitudes and instruction, *American Educational Research Journal* 17, 419-433.
- Stecher, B.M. and Bohrnstedt, G.W. (2000), *Class Size Reductions in California: The 1998-1999 Evaluation Findings*, California Department of Education Sacramento, California.
- Strauss, R.P. and Sawyer, E. (1986), Some evidence on teacher and student competencies, *Economics of Education Review* 5, 41-48.

WHAT DO WE KNOW ABOUT EFFECTS OF SCHOOL RESOURCES ON
EDUCATIONAL RESULTS?, Jan-Eric Gustafsson

Wansbeek, T. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics*, Elsevier Science B S, Amsterdam.

Wenglinsky, H. (1998), Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement, *Educational Evaluation and Policy Analysis* 20, 269-283.