Review

The effect of peer socioeconomic status on student achievement: A meta-analysis

Reyn van Ewijk [a,*], Peter Sleegers [b]

[a] TIER, Amsterdam School of Economics, University of Amsterdam, P.O. Box 94208, 1090 GE Amsterdam, The Netherlands
[b] Department of Educational Organisation and Management, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

ABSTRACT

Previous studies on the effects on students' test scores of their peers' socioeconomic status (SES) reported varying results. A meta-regression analysis including 30 studies on the topic shows that the compositional effect that researchers find is strongly related to how they measure SES and to their model choice. If they measure SES dichotomously (e.g. free lunch eligibility) or include several average SES-variables in one model, they find smaller effects than when using a composite that captures several SES-dimensions. Composition measured at cohort/school level is associated with smaller effects than composition measured at class level. Researchers estimating compositional effects without controlling for prior achievement or not taking into account the potential for omitted variables bias, risk overestimating the effect. Correcting for a large set of not well-thought-over covariates may lead to an underestimation of the compositional effect, by artificially explaining away the effect. Little evidence was found that effect sizes differ with sample characteristics such as test type (language vs. math) and country. Estimates for a hypothetical study, making a number of "ideal" choices, suggest that peer SES may be an important determinant of academic achievement.

Contents

* Corresponding author. Tel.: +31 20 525 4302.
  E-mail address: R.vanewijk@uva.nl (R. van Ewijk).

1. Introduction

Since the influential Coleman report (Coleman et al., 1966) first brought the topic into the spotlights, the effect of peer average socioeconomic status (SES) on students' school performance has been discussed widely among researchers from different disciplines (Economics, Educational Sciences, Sociology). As known, students generally perform better in school if their own SES-background is higher. If the SES of their peers has a separate effect above this, then this has some important implications. In school choice debates, it is a well-known argument that choice increases sorting of students. This may lead students with a low-SES-background to miss positive effects from attending school with high-SES peers while high-SES students profit from getting a "better" peer group. In total, this would widen the achievement gap between both groups of students. In school accountability systems that judge schools based on their students' test scores, the presence of peer effects alters the level that a school can be expected to attain with its students. Schools with many low-SES students will perform poorer than would be expected based on the school's quality. This holds even if the individual backgrounds of its students are corrected for. The opposite happens for schools with a relatively high average-SES intake.

Despite the large amount of research findings that are available, researchers still have not reached consensus on the subject. This lack of agreement is related to a substantial variation in the reported research results, which ranges from no effect at all (e.g. Bondi, 1991; Evans, Oates, & Schwab, 1992) to strong peer group effects (e.g. Ho Sui Chu & Willms, 1996; Robertson & Symons, 2003). Although scholars acknowledge that there are differences in approaches (McEwan, 2003; Thrupp, 1995) and have tried to improve their studies and models, no attempts have been made to summarize or synthesize the findings from previous studies by conducting a meta-analysis. We will try to fill in this gap by conducting a meta-analysis which will help to increase our understanding of the nature and size of the effect of peer SES on student achievement. Meta-analysis is a set of techniques used to systematically review the literature in a domain and can be used to estimate how characteristics of studies such as choice of sample and research design affect the results reported in different studies (Stanley, 2001).

The terminology used to describe the effect of peer characteristics on achievement differs between disciplines. In Economics, usually the term "peer effect" or "peer group effect" is used (Evans et al., 1992; Hoxby, 2000b; Zimmer & Toma, 2000). In the Social Sciences, the effect is variously described as "compositional effect" (Strand, 1998; Van Damme, De Fraine, Van Landeghem, Opdenakker, & Onghena, 2002), "contextual effect" (Hauser, 1970; Willms, 1986), "school mix effect" (Lauder & Hughes, 1999; Thrupp, 1995) or "aggregated group-level effect" (Hutchison, 2003). Although fine differences between these terms exist (Harker & Tymms, 2004), they all refer to the same underlying principle, namely the effect on a student's achievement associated with the background of the children she attends school with. Hence, we will make no fundamental distinction between studies according to the term used and will use the terms interchangeably.

Despite the large number of studies conducted on this issue, few studies have tried to investigate the channels through which peer group composition would affect achievement. Most studies treat the effect as a "black box". Nevertheless, several causal paths have regularly been proposed through which the effect could take place: average SES may affect the disciplinary climate or atmosphere in a class (Hoxby, 2000b); the teacher may adjust her style of teaching to the type of students in the class (Harker & Tymms, 2004); high-SES schools may benefit from greater support from parents (Opdenakker, Van Damme, De Fraine, Van Landeghem, & Onghena, 2002), while peer pressure and peer competition may stimulate students to work harder (OECD, 2001). Lastly, peer effects may be statistical artifacts that only show up in analyses because of poor controls for endogeneity (Evans et al., 1992).

The aim of the present meta-analysis is to systematically review the findings from previous studies and try to come to an understanding of why researchers have alternately found small effects, large effects, or no effects at all. The analysis will focus on the effects of student population composition associated with SES on primary and secondary school children's academic achievement. We argue that the large differences between results reported in previous studies are related to the types of samples used, the operationalization of peer social background, and the estimation models employed. We will therefore analyze if and to what extent differences in approaches used and choices made by researchers affected the size of the peer effect they reported. The model specification we use also enables us to estimate the effects a hypothetical "ideal" study – fulfilling a certain set of limiting conditions on study and model characteristics – would find. In order to check the robustness of our results, we will conduct study fixed effects meta-regression analyses: an addition to current meta-analytic techniques. This will help us to increase our understanding of the characteristics that studies aiming to give good estimates of the peer effect should have. This paper proceeds as follows: Section 2 describes how the studies included in this meta-analysis were identified and selected. Section 3 discusses the sources of variation between the studies that might affect the sizes of the peer effects they found and in Section 4, we discuss our estimation strategy. In Section 5 the results are presented. Section 6 concludes.

2. Selection of studies

This meta-analysis synthesizes studies that assess the effect of the average SES of peers on individual students' academic results. To be included in this review, a study has to meet the following criteria:

1. It has to report estimates of the effect of an increase in mean SES of the peer group, being the children a student attends school with, by one individual-level standard deviation, or the estimates can be converted into such a measure. Studies

that define composition only by means of categories (e.g. schools with more vs. less than a certain percentage of students coming from poor families) are not included, since effects of this type of variable cannot be reliably transformed into estimates of the required type.

2. The dependent variable has to be individual students' educational achievement as measured by scores on tests of mathematics, language, science or general academic achievement (being combinations of the three other types of tests). Studies measuring educational achievement only by rough categories such as dropping out of school or the passing of exams were not included, because their focus is different. These studies focus on only one specific point of the distribution, namely the lower end (i.e. the borderline between passing and failing), whereas we focus on a shift in the total distribution.

3. The estimation model as used in a study has to include as a covariate the individual-level variable corresponding to the average SES-variable. Not doing this would lead the aggregated variable to serve as a proxy for individual students' own SES, because of the strong correlation between both variables. This would cause a considerable overestimation of the peer effect.

4. The students in the sample have to be in primary or secondary (high) school (6–18 years old).

5. The study has to be published or presented no earlier than January 1986 and before January 2006.

6. The study has to be written in English.

7. The study has to use level/present test scores as the dependent variable in its model. A great majority of the studies that met the aforementioned criteria used level/present test scores as their dependent variable; a few used gain scores (being the present test score minus the test score at an earlier point in time). Estimates coming from gain equations refer to a different type of effect than estimates coming from level equations (even if the latter include a covariate for test score at an earlier point in time). Therefore, both types of estimates cannot be compared or taken together in the same meta-analysis. The low number of studies using gain score models precluded performing a separate meta-analysis on them. Hence, our analyses remain constrained to estimates from level equations.

Studies, both published and unpublished, that met these criteria for inclusion were identified by systematic searches of electronic databases related to different disciplines including EconLit, Sociological Abstracts and ERIC. Search terms included combinations of the terms peer, peer effect, peer influence, composition, socioeconomic influences, socioeconomic status, socioeconomic background, classroom environment and achievement. Each of the studies identified by the electronic searches was thoroughly examined for references to other studies on the subject of peer effects. This yielded a substantial number of additional studies.

All studies eligible for inclusion were coded by one of the researchers by means of a formal scheme. To obtain the required high degree of reliability in the codings (cf. Cooper & Hedges, 1994; Lipsey & Wilson, 2001) the accuracy of the codings was independently verified by the other researcher who checked all the data. Differences in codings were discussed until consensus was reached among the two researchers.

Using the coding form, information was recorded on the relevant characteristics and identifiers of the study, as well as on the factors that were hypothesized to influence the sizes of the peer effects that the study found. That is, the following aspects were systematically coded: (a) the way in which the compositional (average SES-) variables were operationalized and measured, (b) characteristics of the samples that were used and (c) how the models used for estimation of the peer effects were specified. These aspects will be discussed in detail in the next section. Whenever information necessary for coding was not reported in the study, we contacted the author(s). A few studies could not be included in the meta-analysis, because the authors could not be contacted, or because the authors were unable to retrieve information that was essential for the study to be included.

Most studies gave several estimates of the peer effect of SES. These estimates differed in the subject of the achievement test, the sub-sample, or in the model specification. In some cases, different models were shown in order to arrive at one or more best models. Whenever this was the case, the other ("non-optimal") estimates were excluded from this meta-analysis. In other cases, however, no clear "best" model was identified. Instead, a set of different models was reported in which no estimate of the peer effect was valued over the others. When this was the case, all alternative models were included in this meta-analysis. This had as an added advantage that it increased the variance in our set of predictors and hence the exactness of our identification.

The final database included 188 estimates from 30 studies. In Table 1, the included studies are summarized. The average estimated effect of composition varied considerably between the studies: the lowest estimates suggest that increasing the average socioeconomic status of a student's peer group by one student-level standard deviation leads to 0.03 standard deviation higher test scores; the highest estimates suggest an effect of 0.59. Six studies came from the field of Economics; the other 24 studies came from the field of Social Sciences. Most studies focused on western countries, but two studies focused on South America, and the OECD-studies included estimates on countries from all over the world. The age of the students in the various samples ranged from 6.5 to 18. The other columns of the table show characteristics of studies that may moderate the size of the compositional effect they found. These characteristics will be discussed below. The table thus indicates that there is considerable variation on these, which will also be necessary for our analyses. Finally, it should also be noticed that a considerable number of estimates were derived from a few OECD-studies. As will be described, because of the weighting procedure we use, this does not lead these studies to have a very high weight in our regressions. We will show in a sensitivity analysis that our results are robust to exclusion of these studies.

Table 1
Summary of the 30 studies used in the basic meta-regressions.

Author(s) (publication year)	Contributed estimates	Discipline	Countries in sample	Average SES measured at level of	Average student age	Type of test used	Attempts to overcome omitted vars bias	Prior attainment included as a covariate	>1 average SES-var. in one model	Type of average SES-variable(s)	Average weighted ES
Bankston and Caldas (1996)	3	Soc. Sc.	USA	Cohort/school	16	GAA	No	No	Some models	Dich.; composite	0.03
Bankston and Caldas (1998)	2	Soc. Sc.	USA	Cohort/school	16	GAA	No	No	Yes	Dich.; composite	0.18
	1	Soc. Sc.	UK: Scotland	Cohort/school	11.5	Lang.	No	Yes	No	Occup.	0.03
Caldas and Bankston (1997)	2	Soc. Sc.	USA	Cohort/school	16	GAA	No	No	Yes	Composite	0.03
Caldas and Bankston (1998)	1	Soc. Sc.	USA	Cohort/school	16	GAA	No	No	No	Composite	0.20
De Fraine, Van Damme, Van Landeghem, Opdenakker, and Onghena (2003)	1	Soc. Sc.	Belgium: Flanders	Class	14	Lang.	No	Yes	No	Composite	0.29
Harker and Nash (1996)	3	Soc. Sc.	New Zealand	Cohort/school	16	Lang.; Math; Science	No	Yes	No	Occup.	0.13
Ho Sui Chu and Willms (1996)	2	Soc. Sc.	USA	Cohort/school	14	Lang.; Math	No	No	No	Composite	0.26
Hutchison (2003)	3	Soc. Sc.	UK	Cohort/school	8; 10	Lang.	No	Yes	No	Dich.	0.06
Lee and Bryk (1989)	1	Soc. Sc.	USA	Cohort/school	18	Lang.	No	No	No	Composite	0.34
Ma and Klinger (2000)	8	Soc. Sc.	Canada	Cohort/school	12	Lang.; Math; Science	No	No	No	Home resources	0.16
McEwan (2003)	6	Ec.	Chile	Class	14	Lang.; Math	Yes	No	Yes	Educ.	0.43
McEwan (2004)	8	Ec.	Chile and Bolivia	Cohort/school	9; 10; 12; 14	Lang.; Math	No	No	No	Educ.; Dich.	0.43
OECD (2001)	6	Soc. Sc.	OECD average	Cohort/school	15	Lang.; Math; Science	No	No	No	Composite	0.59
OECD (2003)	36	Soc. Sc.	36 countries	Cohort/school	15	Lang.	No	No	No	Occup.	0.42
OECD (2004)	35	Soc. Sc.	34 countries and OECD average	Cohort/school	15	Math	No	No	No	Composite	0.45
OECD (2005)	35	Soc. Sc.	35 countries	Cohort/school	15	Lang.	No	No	No	Occup.	0.36
Opdenakker et al. (2002)	2	Soc. Sc.	Belgium: Flanders	Class	14	Math	No	Yes	No	Composite	0.13
Paterson (1991)	1	Soc. Sc.	UK: Scotland	Cohort/school	16	GAA	No	Yes	No	Composite	0.27
Peetsma et al. (2005)	2	Soc. Sc.	Netherlands	Class	10	Lang.; Math	No	Yes	No	Dich.	0.05
Rivkin (2001)	1	Ec.	USA	Cohort/school	18	GAA	Yes	Yes	No	Dich.	0.04
Rumberger and Willms (1992)	12	Soc. Sc.	USA	Cohort/school	17	Lang.; Math	No	No	No	Educ.	0.18
Schindler-Rangvid (2003)	1	Ec.	Denmark	Cohort/school	15	Lang.	Yes	No	No	Educ.	0.14
Schneeweis and Winter Ebmer (2005)	4	Ec.	Austria	Cohort/school	15	Lang.; GAA	Yes	No	No	Occup; home resources	0.16
Strand (1997)	2	Soc. Sc.	UK	Class	6.5	GAA	No	Some models	No	Dich.	0.25
Strand (1998)	3	Soc. Sc.	UK	Class	11	Lang.; Math; Science	No	Yes	No	Dich.	0.14
Willms (1986)	2	Soc. Sc.	UK: Scotland	Cohort/school	16	Lang.; Math	No	Yes	No	Occup.	0.23
Young and Fraser (1992)	1	Soc. Sc.	Australia	Cohort/school	14	Science	No	Yes	No	Composite	0.12
Young and Fraser (1993)	1	Soc. Sc.	Australia	Cohort/school	14	Science	No	Yes	No	Composite	0.05
Zimmer and Toma (2000)[a]	3	Ec.	Belgium, USA, Canada, New Zealand and France pooled	Class	13.5	Math	No	Yes	Yes	Dich.	0.06

Soc. Sc.: Social Sciences; Ec.: Economics; Lang.: language; GAA: general academic achievement test; Dich.: dichotomously based; Educ.: parental education; Occup.: parental occupation. The total number of included effect estimates was 188 from 30 studies.

[a] Zimmer and Toma's estimate using average mother's occupational status as the average SES-variable was excluded from this meta-analysis, since it referred to whether the mother was working outside the home; we find it doubtful whether, and if so, how, this measure of occupational status indicates socioeconomic status.

3. Sources of variation between studies

As mentioned earlier, the large differences between studies may be related to approaches researchers used and choices they made to analyze the presence and size of the effect of peer average SES. Before describing the estimation strategy for our meta-regression analysis, we first discuss important characteristics on which the studies included in the meta-analysis differ. These can be divided into three sources of variation: (1) measurement of the compositional variable, (2) sample characteristics, and (3) model specification.

3.1. Measurement of the compositional variable

Although a high degree of consensus among authors from different disciplines exists about what SES should measure, the way in which authors operationalize the compositional variable, average SES, still differs considerably. Researchers generally agree that SES refers to the extent to which individuals, families or groups have access (either realized or potential) to, or control over valued resources, including wealth, power and status (Mueller & Parcel, 1981; Oakes & Rossi, 2003). There also seems to be agreement on a three-componental view of SES which states that SES can be indicated by either parental education, parental occupation, or parental income (Duncan, Featherman, & Duncan, 1972; Hauser, 1994; Mueller & Parcel, 1981). Sirin (2005) adds as a fourth indicator home resources, which refers to the extent to which a student's home situation provides an environment that is conducive to learning. In a meta-analysis on the effect of individuals' own SES on their own primary school academic achievement, Sirin (2005) shows that the use of different SES-measures is associated with significant differences in the reported strength of the relation between individual SES and academic achievement.

In the studies included in this meta-analysis, the average SES-variable used to measure composition, was often a composite that included two or more of the above mentioned components. In practice, such a composite always included both parental education and parental occupation and usually also a measure of home resources. In a few cases, it also included family income. Other studies operationalized SES as parental education or parental occupation. Two studies included in the meta-analysis used only home resources to operationalize SES. Family income was never used as an SES-measure except as part of a composite. In this study, we will investigate whether the type of average SES-variable used can influence the effect size of the peer effect a researcher finds.

Several studies used SES-measures based on dichotomies. Because of their low reliability, we treat dichotomously based measures as a separate category. In most cases, the dichotomy referred to the proportion of students' peers that were eligible for free or reduced price lunches. In other cases, parental education or occupation was measured dichotomously.[1] Dichotomously based measures of SES are unreliable approximations of SES, since the true value of the underlying concept is continuous. Poverty status as measured by the dichotomy of being eligible for free/reduced price lunch status has the additional disadvantage of being a very unstable measure. Hauser (1994) strongly advises researchers to refrain from using free lunch status in studying the effects from economic deprivation. Hill and Jenkins (2001) show that between 1991 and 1996, in Britain around 25% of children aged 6–11 experienced at least 1–2 years of poverty. However, only 1.5% were in poverty for the full 6 years. This instability adds to the unreliability of free/reduced price lunch measures. Because of attenuation bias, we expect that where rough, dichotomous measures of average SES are used, the peer effect will be underestimated. This is not to say that non-dichotomous measures of SES will yield unattenuated estimates of the peer effect, but since reliabilities of used SES-measures are generally not known, we assume that this bias is considerably smaller.

Besides by type and reliability of average SES-variables, estimates of peer effects can also be influenced by the number of average SES-variables that are included as covariates in a single regression equation. If more than one average SES-variable has been included, the effect of each is estimated "cleaned" from the effect of the others. If this is the explicit goal of the researcher, this is of course a good strategy. If not, it leads to ambiguity in the interpretation of the parameters. For example, Caldas and Bankston in their studies (Bankston & Caldas, 1996, 1998; Caldas & Bankston, 1997), estimate effects of peers' parental education and occupation (taken together into one composite), while keeping peers' poverty level constant. This leads both parameters to be hard to interpret: neither one completely measures peer SES on its own now anymore. And (like the other studies that included more than one peer SES-variable as a covariate), they did not give a strong theoretical reason to include both SES-measures in the same model. Since SES is defined as people's position on a general social hierarchy (Mueller & Parcel, 1981; Oakes & Rossi, 2003; Sirin, 2005), a peer SES-variable should capture as much information on peers' average position in this social hierarchy as possible. Including a second peer SES-variable as a covariate, takes valuable information away from the first and hence leads to bias in its parameter.

A last aspect of the average SES-variable that can have a profound impact on the size of the peer effect is the level at which it is measured. Some studies measured the peer SES-variable at the level of the class. This is consistent with the view that the relevant peer group for a student is formed by the children she daily attends class with and not by the entire cohort or school,

---

[1] Six studies used free/reduced price lunch status. Peetsma et al. (2005), Zimmer and Toma (2000) and McEwan (2004) used the proportion of parents having an education above a certain level. Zimmer and Toma (2000) also used skilled versus unskilled level of father's occupation. Note that Sirin (2005) treats the dichotomous variable free/reduced-price lunch status as conceptually different from the other four types. We do not agree with this, because we consider a child's lunch status as an indicator of her family's income. Instead, we use a separate category that includes all the dichotomously based measures.

which also includes many children that the student rarely or never interacts with. In many studies, however, average peer SES was measured at the level of the cohort or school.[2] To the extent that the composition of cohort or school differs from the composition of the relevant peer group unit (the class), composition measured at the cohort or school level can be viewed as a less reliable approximation of the true peer group variable. Estimates of the peer effect then suffer from (attenuation) bias towards zero. This problem is relatively small if the average cohorts for which SES-composition was established consists of little more than one class. Peetsma, Van der Veen, Koopman, & Van Schooten (2005), for example, have cohorts averaging 1.2 class. In such a case, the composition of the cohort will hardly differ from that of the class and the resulting attenuation bias will be much lower than when the cohort consists of around 100 students or more as is the case in Caldas and Bankston's study (Caldas & Bankston, 1998). Studies measuring composition at cohort level in which the average cohort consisted of 40 students or less, were classified as effectively measuring peer effects at class level (Angrist & Lavy, 1999); if cohorts averaged more than 40 students, we classified the study as measuring composition at the cohort/school level.[3]

3.2. Sample characteristics

In addition to being affected by the way in which average SES is measured, the size of the peer effect can also depend on some characteristics of the sample that is used. These sample characteristics include the type of achievement test used, students' age and country in which the study was carried out. Achievement tests could generally be classified into language, mathematics and science tests. Some studies used general academic achievement tests, which always consisted of different ratios of language, mathematics and science tests. These 'constructed' general achievement tests were coded for our analyses accordingly, as partially language, partially mathematics, and/or partially science tests.

The extent to which students are amenable to peer effects may change with their age. As children get older, the influence of adults such as parents and teachers on their behavior may decrease, while the influence of peers of their own age increases. Based on this assumption, the peer effect would be expected to increase in size as students get older.

The main reason why differences in the size of a peer effect between countries may arise is that countries differ in their strength of social hierarchy or social inequality. We standardized SES for each study separately when calculating effect sizes (see below). This meant that moving up one standard deviation on the SES-distribution was related to a larger increase in access over resources if the country studied experienced greater social inequalities. In our models, we will therefore include as a covariate the country's standardized GINI-coefficient as an indicator of wealth inequality in a country.[4]

3.3. Model specification

A last great source of differences between studies examining the compositional effect of SES lies in their model specifications and more specifically in the way they deal with endogeneity and omitted variables bias. When estimating peer effects, the potential issue of endogeneity bias is a well-known problem: statistically established peer effects may be artifacts if students do not score lower *because* they are in a class with a certain composition, but are in this class because of other factors that make them more likely to attend this class *and* that simultaneously negatively affect their scores (Evans et al., 1992; Harker & Tymms, 2004; Hauser, 1970; Nash, 2003). For example, when a child from highly educated parents goes to a school with many low-SES children and performs poorly, this effect cannot be automatically attributed to the lower SES of his classmates. The reason that he performs poorly may be the same as the reason why he goes to this particular school instead of one with more high-SES children. Perhaps his parents are somewhat a-typical for higher educated people, e.g. they have poorer paying jobs (which makes them end up living in a poorer neighborhood amidst lower SES families and close to the low-SES school) and provide a poorer home-environment to their child (leading to his lower performance). This child would then also have performed poorly if he would have gone to a higher SES school. Hoxby (2000b) adds that even within schools, there may be selective sorting as motivated parents (who stimulate their children to perform well in school) try to get their child in the class with the best teacher or with the best (or highest-SES) fellow students.

Despite the fact that these endogeneity problems have been described often, very few studies formally take them into account in their models when estimating the SES-peer effect. Those that did, were in all cases studies from the field of Economics. Here, we come to a fundamental difference in approaches between the studies on this topic from the field of

---

[2] Note that in both instances, average peer SES was generally described as "school average SES". Sometimes, this referred to a cohort average and sometimes it was not clear whether it referred to the average of the cohort or of the entire school. Because of this, we are not able to distinguish between average SES measured at cohort and at school level.

[3] To put this into perspective: Hoxby (2000a) found that 1% of Connecticut primary school classes had 34 pupils or more. Class sizes in developing countries, are often higher: although average secondary school class size in the OECD countries is around 25, in some countries, it is up to 39 (OECD, 2003). Angrist and Lavy (1999) use Maimonides' rule, stating that classes should be split up as soon as their size exceeds 40, to study the effects of class size reductions in Israel. We apply a similar "rule of 40", to distinguish between studying students' relevant peer group and studying a broader group that also includes many children irrelevant for the student in question.

[4] The GINI-coefficient measures the area between a cumulative distribution line and a straight (45 degree) line in a graph plotting the cumulative share of income earned against the cumulative share of people earning less than a certain income. Data on countries' GINI-coefficients were obtained from CIA's World Factbook 2007 (CIA, 2007). Estimates from OECD (2003, 2004, 2005) on Albania, Iceland, Luxembourg and Serbia were removed because of unknown GINI-values.

Economics and those from the (other) fields of Social Sciences.[5] Economists generally confined themselves to a relatively small topic and had as their purpose the exact estimations of one, or a few clearly related parameters. If covariates were added to their estimation model, this was usually done to improve the estimation of the parameter of interest. The Social Scientific studies, on the other hand, often aimed at studying several phenomena at the same time: not only peer effects. As a consequence, in several of these studies large numbers of predictors were included in one model without thorough concern about whether and how inclusion of one would influence the coefficients on the others. This difference in approaches is most easily illustrated by the goals set forth in a few studies. Social Scientists Young and Fraser (1993) and Bondi (1991) aim at broad goals: "to investigate science achievement (. . .) and how this achievement can vary from school to school" (Young & Fraser, 1993, p. 265) and "to investigate factors influencing the attainment of students" (Bondi, 1991, p. 204). Economists McEwan (2003) and Schindler-Rangvid (2003) have a much narrower goal: they aim at "estimates of peer effects on student achievement" (McEwan, 2003, p. 131) and "to estimate educational peer effects" (Schindler-Rangvid, 2003, p. 107). Although the broader scope of Social Scientific studies leads to a gain in content and may yield a wide array of important results, this may come at the cost of a higher risk of bias in individual parameters. This can be especially troublesome for effects that are as difficult to estimate free from bias as peer effects. Particularly at risk may be four studies from the OECD on the PISA-databases (OECD, 2001, 2003, 2004, 2005). These voluminous studies report on everything from competitive vs. cooperative learning to the effects of school climate on achievement and compare such effects between the few dozen countries in their database. When estimating peer effects, the same model is used for each of the countries, without adjustments for any country-specific situations. Such a one size fits all approach may overlook country-specific issues that authors focusing on only one country would have found necessary to solve by fine-tuning their model to the requirements for that specific country. Schneeweis and Winter Ebmer (2005), for example, use PISA-data on Austria, and argue that including a set of school type dummies is necessary for Austria to take into account the substantial sorting into the different Austrian school types. The OECD-studies ignore this country-specific issue, which may bias their results. Because of this general difference between Economic and Social Sciences in approach and model specification, we expect studies from the field of Economics, especially those that used a formal strategy to overcome endogeneity problems, to give less biased (smaller) estimates than studies from the field of Social Sciences.

One specific characteristic of the models used in the included studies is the use of a covariate for individual students' prior attainment or ability. It has often been pointed out (Goldhaber & Brewer, 1997; Hanushek, Kain, & Rivkin, 2002; Ho Sui Chu & Willms, 1996; Rumberger & Palardy, 2005) that not correcting for prior scores leads to an overestimation of effects. The reason behind this is twofold: first, prior attainment may have influenced the school or track a student currently attends. Students going to a low track because of poor prior attainment, will often have lower SES peers. Estimating the effect of composition without correcting for prior attainment, may lead to mistaking the effect of a student's poor performance in the past for an effect of having low-SES peers. Second, the student's prior attainment is affected by his peer group composition in the past. If past and present composition are correlated, not correcting for prior attainment leads the coefficient on current peer group composition to pick up the effects of composition in the past. Through both channels, leaving out prior attainment/ability will lead to an overestimation of the peer effect.

4. Estimation strategy

4.1. Estimation of basic meta-regression models

To make effect estimates comparable across studies, we standardize each effect estimate that was reported in a study. The original effect estimates were the regression coefficients of average SES on test scores. We linearly transform those, so that they now refer to the effect on standardized test scores of increasing the average peer group-SES by one *individual*-level standard deviation. (We could also have let our estimates refer to effects of going up one standard deviation in the *school* (or class/cohort) *average* SES-distribution. Since the standard deviation of school average SES, however, depends on the degree of school segregation in a population, this would make a comparison across studies that focus on different populations (with hence different degrees of segregation) problematic. Going up one standard deviation in the individual-level SES-distribution is much more comparable across populations.) The standard errors to the estimates, which are used to determine the weights in the meta-regression as will be described below, underwent the same linear transformation.[6]

---

[5] Some researchers see Economics as one of the Social Sciences; some see it as a separate discipline. There is a clear division in methods here between studies from the field of Economics and those from (other) Social Sciences. The latter were conducted by Sociologists, Educational Scientists and scholars from a few related fields. For ease of terminology, we will henceforth refer to the latter disciplines as "Social Sciences" and to the former as "Economics".

[6] If no standard errors, but only significance levels were reported, we computed standard errors' assuming a p of 0.05 if significance at 5% was reported, etc. (cf. Cooper & Hedges, 1994). For effects reported as "not significant", we took p as halfway between the significance level used for testing (usually 0.05) and the p going with no effect at all (0.50 for two-sided testing). If no parameter was reported, but an effect was only referred to as "not significant", we interpreted this conservatively as an effect of 0 and imputed the corresponding standard error from other estimates presented in the same study. Some studies reported OLS regressions without appropriately taking into account the clustered nature of the data. Their reported standard errors were adjusted based on the distribution of variances over class/cohort and school (which if not available was estimated from studies using similar datasets) and group sizes. Lee and Bryk (1989) used a group-mean centered multilevel model. The standard error to the estimate we were interested in (a difference between two parameters) should be adjusted using the covariances between the two parameter estimates (Bryk & Raudenbush, 1992). This covariance was unknown; instead, we used the unadjusted standard error, which is probably a slight underestimate of the true value.

Each (standardized) estimate T_{ij} reported by a study j is an estimate of the "true" size of the peer effect, θ_{ij}:

$$T_{ij} = \theta_{ij} + e_{ij} \tag{1}$$

The estimation or sampling error, e_{ij}, is the standard error to the estimate as reported in the study and standardized as described above. The squared of this is the estimation variance, denoted by v_{ij}. The true effect is not constant across all estimates, but differs according to a number of characteristics of study and model, X_k, that were discussed in Section 3:

$$\theta_{ij} = \beta_0 + \sum_{k=1}^{l} \beta_k X_{kij} + u_{ij} \tag{2}$$

The term u_{ij} captures systematic variance between the estimates that arises because of (often unobserved) differences between those estimates that are not included among the X_k. Its associated variance is σ_θ^2. Combining (1) and (2), we come to a meta-regression equation of the form:

$$T_{ij} = \beta_0 + \sum_{k=1}^{l} \beta_k X_{kij} + e_{ij} + u_{ij} \tag{3}$$

Adding a systematic variance component means that the estimates included in the meta-regression are treated as a random sample from all estimates that could have potentially been made on the effect. It hence enables the researcher to generalize beyond the particular set of estimates that are included to the entire universe of all potential studies. Therefore, the use of such models that include both a sampling error and a systematic variance term, is propagated by many meta-analysts (Hedges & Vevea, 1998; Lipsey & Wilson, 2001; Overton, 1998; Raudenbush, 1994).[7] The study and model covariates, X_k, are generally dummy variables. Because of the way we assign the 0- and 1-values to these, the constant, β_0 refers to the peer effect that a hypothetical "ideal" study is expected to find. This "ideal" study would possess all the characteristics that we argued are best (i.e. an attempt would be made to overcome endogeneity/omitted variables bias; the compositional variable would be a composite, etc.). Only age and the standardized GINI-coefficient are no dummy variables. Age is coded as deviations from a student age of 18: the maximum age at which a study would be eligible for inclusion into this meta-analysis. Thus, to the characteristics of the "ideal" study, we added that the age of the students in the sample would be 18 and that the standardized GINI-coefficient would be 0 (about the value for the USA). Also, the dummies are set so that the "ideal" study studies the effect on language and is from the field of Economics. The latter we chose because these studies generally focused on estimating as exactly as possible only the one, specific parameter that we also focus on, while several of the Social Scientific studies had a much broader goal, which may increase the risk for bias.

Commonly, in meta-analytic models, each estimate is weighted by the inverse of its total variance (Lipsey & Wilson, 2001; Raudenbush, 1994):

$$w_{ij} = \frac{1}{v_{ij} + \sigma_\theta^2} \tag{4}$$

Note that taking into account systematic variance by σ_θ^2 lowers the weights of all estimates in the meta-analysis, leading to larger standard errors to the parameters in the model. Overton (1998) argues that under certain conditions, it can be assumed that σ_θ^2 equals zero, in which case estimation would become more efficient by omitting the u_{ij} term. Also, if the aim of the meta-analysis would not be to generalize to all studies that could potentially be performed on the topic, but only to make statements on the particular set of studies in the meta-analysis, u_{ij} should be omitted (Hedges & Vevea, 1998). Neither of these is the case here: we do want to make this generalization, and, as estimates of σ_θ^2 will show, there are substantial differences between the estimates in our sample that cannot be explained away by the available set of covariates. We therefore include the term in our model. The resulting lower accuracy of the estimates is, as Raudenbush (1994) notes, the "price we pay" for approaching the studies in our analyses as a random sample from the universe of potential studies, instead of as forming the complete population themselves. We follow the general weighting strategy from Eq. (4), but first have to take one more thing into account. As discussed before, most studies provided more than one effect estimate T_{ij}, for which some of the study and model characteristics X_{kij} are usually different. Including each estimate separately and independently in our meta-regression and weighting it by w_{ij}, gives studies contributing several effect estimates a disproportionately large weight in determining our overall outcomes and would lead to overconfidence in the overall accurateness of the estimated coefficients, since the multiple estimates from a single estimate are often not independent observations. Many authors, including Lipsey and Wilson (2001) argue for a conservative approach by either selecting only one of the multiple estimates supplied by a study selected – at random or based on certain criteria – or to take an average over the estimates. Adopting one of these approaches here, would lead to an important loss of valuable information, because the effect estimates within one

---

[7] Note that in meta-analysis literature, such a model is often referred to as a "random effects model" (Cooper & Hedges, 1994). This use of terms can be somewhat confusing because of the fixed effects models we describe below, in which "fixed effects" refers to something entirely different from the "random effects" in the present model. To avoid confusion, we will avoid the use of the term "random effects" here.

study differ on some of the predictors. Therefore, we decide to include all estimates from each study in our analysis. Simply correcting for clustering of estimates within studies using multilevel meta-analytic models as proposed by Hox (2002) would not suffice in this situation: estimates are often not just coming from the same study (as is for example the case when a study reports a set of estimates on several sub-samples), but even come from exactly the same data. Hence, the estimates are not just correlated, but are in essence codetermined. If multiple estimates are made on the same data, then these data determine all estimates that can be made on it at the same time. Often, the only difference between two reported estimates is that in the second one, some covariates are added to the model; the codeterminedness is caused by the fact that the values of all respondents on both the predictor of interest and the dependent variable will not change between those two estimates. A more restrictive approach is needed that takes this codeterminedness into account. We therefore propose the following strategy.

We make the assumption that we can get no more accurate information from a set of simultaneously determined estimates than the most accurate of these estimates, being the one with the smallest standard error. The accurateness of this estimate equals its inverse estimation variance, v_{ij}^{smallest}. The sum of the inverse estimation variances of all estimates in the set should not be lower (or higher) than exactly this. We therefore divide the inverse estimation variance of the most accurate estimate proportionally over all codetermined estimates, to arrive at an adjusted sampling variance for each estimate of:

$$v_{ij}^* = v_{ij} \times v_{ij}^{\text{smallest}} \times \sum_{\text{all codetermined estimates}} \frac{1}{v_{ij}} \qquad (5)$$

In several cases, the same authors used one database with test scores in multiple studies that were taken up in this meta-analysis, or the same database was used by different authors in their studies. In those cases, we used the same strict procedure in treating estimates as simultaneously determined if they came from the same database and used the same (sub-)sample of students taking the same test and if the same compositional variable was used. Whenever estimates on the same dataset do not fulfill these criteria (e.g. use different sub-samples of students), estimates were not treated as being dependent and were not combined using the procedure described above.[8] The weights for our meta-regression, combining (4) and (5), now become:

$$w_{ij}^* = \frac{1}{v_{ij}^* + \sigma_\theta^2} \qquad (6)$$

Our model is now similar to a weighted least squares regression, with as weights $1/(\sigma_\theta^2 + v_{ij}^*)$. As Lipsey and Wilson (2001) note, however, applying a regular WLS analysis when estimating a meta-regression, leads to incorrect standard errors, since the weights do not represent different numbers of subjects, as is usually the case, but variance in estimates. The standard errors therefore have to be divided by the square root of the mean squared regression error from the WLS (Hedges, 1994; Lipsey & Wilson, 2001). The meta-regression will be estimated using restricted maximum likelihood (Hox, 2002; Thompson & Higgins, 2002).

4.2. Study fixed effects models

To check the robustness of our findings, we conduct study fixed effects meta-regression analyses, in which we combine meta-analytic with fixed effects regression analysis. In meta-analyses, it is especially important to check whether results are robust against omitted variables bias. This is because studies included in a meta-analysis will usually vary on a large number of characteristics, not all of which will be included as covariates in the model. Some of these characteristics are unobserved, while others are observed, but specific to only one study included in a meta-analysis. Due to this idiosyncratic nature, these observed characteristics will generally not be included as covariates in the meta-analysis (e.g. the use of certain very original covariates in one study or a somewhat different choice of sample in another study). The systematic variance component treats such differences between studies as randomly distributed error variance. While this commonly used strategy leads to reasonably more conservatism (and lower weights) in the meta-analytic estimates, it does not take into account that certain (un)observed study characteristics may covary with included covariates. If this happens, then some X_k in Eq. (3) will be correlated with u_{ij}. Such a correlation is particularly a problem in meta-analyses since the number of data points is generally relatively small in comparison to many other (non-meta-analytic) studies, while at the same time each data point receives a high weight. That means that the simultaneous occurrence of some study characteristic captured in a covariate X_k, with a characteristic that is either unobserved or not included as a covariate, in only a few studies can already cause serious problems. The risk of such omitted variables bias is especially large in a so-called meta-ANOVA in which one covariate at

---

[8] Four OECD-studies using PISA-data (OECD, 2001, 2003, 2004, 2005) gave a large number of estimates on different countries. We treated the estimate on the pooled set of countries (i.e. for "OECD combined"/"All countries in the PISA-data") with the smallest standard error as the most accurate one. Separate estimates for individual countries that were also included in this pooled set were treated as simultaneously determined with the pooled estimate. An alternative weighting procedure would treat the estimates for individual countries as completely independent. This would lead to extremely, and unrealistically, high cumulative weights for the total set of four studies: the combined weight (before adding the systematic variance component u_{ij}) would have been around 50% of the total weight of all included studies.

a time is tested. In a meta-regression analysis, multiple covariates can be tested simultaneously. This decreases systematic variance (see e.g. Jarrell & Stanley, 2004 for an application of this). Although this approach does diminish the problem, it does not solve it completely.

An analysis strategy that is often used to solve such a problem in non-meta-analytic studies when panel-data are available, is fixed effects analysis. We believe that a variation of this approach is very promising for meta-analyses if several effect estimates per study are available, as is the case in the rich dataset we constructed. We propose a combination of meta-regression estimation with fixed effects analysis which enables us to filter out all systematic between-studies variation and in this way to obtain estimates that are free from bias due to omitted variables. This analysis serves as an excellent robustness-check on the results from our regular models. We estimate a meta-regression of the form:

$$T_{ij} = \alpha_j + \sum_{k=1}^{l} \beta_k X_{kij} + e_{ij} \tag{7}$$

In this, α_j stands for a fixed effect per study.[9] Since all systematic differences between studies are captured in the fixed effects term, u_{ij} becomes trivial and can be omitted. Note that, since no assumption on the distribution of the α_j is made, using this model, no information on a constant can be obtained. Also, no estimates on characteristics that are constant within each of the studies can be made.

Since some studies contributed only one effect estimate to our dataset, or contributed several that were constant on the X_k, the number of estimates included in the study fixed effects meta-regression was lower than in the basic model: 172 effect estimates from 18 studies were included. Three of the seven parameters that were estimated, were only identified by variation coming from a single study. Although this does not invalidate these parameter estimates, robustness-checks would be stronger if results can be shown to hold across several studies. The dataset we constructed allows us to study whether this is the case. For this, we estimate a second model in which we add a few estimates that were previously excluded, because they were only shown in a study to arrive at a final/preferred model. If such an estimate only differed from the study's final model on one or more of the characteristics we study, such an estimate can be included in order to strengthen the bias-free estimation of specific parameters. In the alternative analysis, 13 more effect estimates from three studies are included. This includes a few estimates from a study by Harker and Tymms (2004), that was previously excluded, since its purpose was not to estimate "true" peer effects, but to show under what conditions peer effects may appear as statistical artifacts. Table 2 shows the studies included in the fixed effects meta-regression models and the information they contributed to them.

5. Results

5.1. Estimates from the basic meta-regression model

Table 3 presents the results from the meta-regression estimates that were derived using Eq. (3). The left column shows a model in which no regressors are included. The resulting constant is the average weighted effect size over all our studies. An increase of the average socioeconomic status of a student's peer group with one student-level standard deviation leads to an increase of her test score with 0.320 *SD*. The effect for a (hypothetical) "ideal" study, given by the constant in the right column, has almost the same size: 0.315, although the standard error to this is considerably larger. This is an effect of considerable size. Sirin (2005) finds an effect of about the same magnitude from increasing a student's own SES by one standard deviation. In the empty model, there is substantial systematic variance between the studies, as can be seen from the highly significant estimate of the random effects variance component, σ_θ^2. By adding a number of predictors, this variance is appreciably reduced, but remains significant. The results show that the large differences in effect estimates reported in the different studies can to a considerable extent be explained by differences between those studies in their operationalization of peer SES and in their estimation strategies.

The size of the compositional effect a researcher finds varies greatly with the type of SES-measure (s)he uses. Composite measures of SES are built up of several of the components of SES and are therefore the measures that best capture the entire concept of SES as the extent of access to valued resources. Hence, there is a clear advantage to using such a type of measure. The results show that measures that only used information on parental education lead to much smaller (−0.16) effect sizes. In contrast, composite measures (which include parental occupation) and measures solely consisting of parental occupation are associated with about the same effect size. When home resources are used as the average SES-variable (as was done by only two of the studies), lower effect sizes were found. Furthermore the results show that if the hypothetical ideal study we defined would use a dichotomously based average SES-measure such as free lunch eligibility, instead of a composite, the peer effect would presumably all but disappear. This result emphasizes the problematic nature of dichotomously based measures, which tend to be very unreliable and, in the case of free lunch, also instable (Hauser, 1994; Hill & Jenkins, 2001).

---

[9] Note that "fixed effect" here refers to something that is entirely different from what is usually meant by fixed effects meta-regression (cf. Cooper & Hedges, 1994; Lipsey & Wilson, 2001): these models are similar to our Eq. (3), but omit error term u_{ij}. Whenever we mention fixed effects, we do not refer to this type of model.

Table 2
Studies included in the fixed effects meta-regressions.

	Estimates in basic model	Estimates in extended model	Age	Math vs. language	Science vs. language	Prior attainment	SES dichotomous vs. composite	>1 average SES in one model	GINI
Bankston and Caldas (1996)	3	3	N	N	N	N	Y	Y	N
Bankston and Caldas (1998)	2	2	N	N	N	N	Y	N	N
Caldas and Bankston (1997)	0	5	N	N	N	N	Y*	Y*	N
Harker and Nash (1996)	3	3	N	Y	Y	N	N	N	N
Harker and Tymms (2004)	0	6	N	N	N	Y*	N	N	N
Ho Sui Chu and Willms (1996)	2	2	N	Y	N	N	N	N	N
Hutchison (2003)	3	3	Y	N	N	N	N	N	N
Ma and Klinger (2000)	8	8	N	Y	Y	N	N	N	N
McEwan (2003)	6	6	N	Y	N	N	N	N	N
McEwan (2004)	8	8	Y	Y	N	N	N	N	Y
OECD (2001)	6	6	N	Y	Y	N	N	N	N
OECD (2003)	36	36	N	N	N	N	N	N	Y
OECD (2004)	35	35	N	N	N	N	N	N	Y
OECD (2005)	35	35	N	N	N	N	N	N	Y
Peetsma et al. (2005)	2	2	N	Y	N	N	N	N	N
Rumberger and Willms (1992)	12	12	N	Y	N	N	N	N	N
Schneeweis and Winter Ebmer (2005)	4	4	N	Y	Y	N	N	N	N
Strand (1997)	2	2	N	N	N	Y	N	N	N
Strand (1998)	3	3	N	Y	Y	N	N	N	N
Willms (1986)	2	4	N	Y	N	Y*	N	N	N

Note: Y/N indicate that the study did/did not contribute information to the fixed effects meta-regression models; Y* indicates that the study only contributed this information to the extended fixed effects meta-regression model.

Table 3
Parameter estimates and (standard errors) for the meta-regression models.

	Empty model	Basic meta-regression
Constant	0.320 (0.016)**	0.315 (0.105)**
Compositional variable is (omitted category is "composite")		
Parental education		−0.156 (0.055)**
Parental occupation		−0.020 (0.043)
Home resources		−0.258 (0.067)**
Dichotomously based		−0.246 (0.060)**
>1 average SES-variable in one model		−0.193 (0.058)**
SES-variable is measured at cohort-/school level (omitted category is "at class level")		−0.168 (0.066)*
Test (omitted category is language)		
Math		0.001 (0.040)
Science		−0.016 (0.067)
18 minus age		0.008 (0.010)
zGINI		0.005 (0.014)
Prior attainment NOT included as a covariate		0.258 (0.053)**
Does NOT attempt to overcome omitted vars bias		0.118 (0.082)
Social Sciences (omitted category is Economics)		−0.130 (0.064)*
N	188	188
R^2	0.00	0.39
Systematic variance component (σ_θ^2)	0.0322 (0.0047)**	0.0181 (0.0030)**

* Significant at 0.05 level.
** Significant at 0.01 level.

Another way in which low reliability in the compositional variable seems to affect the effect estimates in many studies is through the level at which measurement takes place. If the average SES-variable is measured at cohort/school level instead of at class level, the magnitude of the effect is reduced by about half. Since a student's relevant group of peers is formed by her classmates, with whom she interacts daily, and not by the cohort or school in its entirety, any measure of composition that does measure characteristics at a higher level than that of the class, is a noisy measure and using it, because of attenuation bias, leads to underestimation of the true peer effect.

Interestingly, three main characteristics of samples used did not seem to be related to the differences between studies in their reported effect sizes. Peer SES has about an equally sized effect on students' language, mathematics and science test scores. Peer effects did also not differ between children of different ages. This means that the results do not confirm our hypothesis on age (as children get older, the influence of peers on behavior increases at the expense of the influence of adults). The small, insignificant coefficient for standardized GINI-coefficient shows that between countries that differ in their extent of social inequality, the peer effect does not vary. Peer effects can be found in every country, and in each country, they are about equally large.

As was expected, not including a prior attainment covariate leads to considerably higher effect estimates: in fact, it almost doubles the effect sizes found by researchers. As was pointed out earlier, this should be seen as an overestimation of the true effect size.

Furthermore, the results show that not making an explicit attempt to overcome omitted variables bias does not lead to significantly different effect estimates. Nevertheless, the coefficient was in the expected direction and quite large. Not making such an attempt seems to be associated with finding effect sizes of about a third higher. That the coefficient is insignificant, may be related to the low number of studies (four) that made such an attempt and hence to the large standard error to the coefficient. It may, however, also be related to the arguably imperfect ways in which these four studies tried to deal with this bias. Schindler-Rangvid adds a set of covariates that was carefully selected, but which may not capture all omitted variables bias. Rivkin (2001) adds region/community type fixed effects, Schneeweis and Winter Ebmer (2005) add school type fixed effects and McEwan (2003) adds school fixed effects. These fixed effects may not completely account for the fact that students may be non-randomly allocated to schools within regions or to classes within schools. To account for the remaining omitted variables bias, McEwan (2003) therefore adds family fixed effects and looks at differences between twins attending different classes. This is a promising approach, but greatly reduces his sample size, so that his estimates become very imprecise. The upshot of all this is that our results cannot give a definitive answer to the question whether formal approaches for dealing with endogeneity, such as instrumental variables or fixed effects models, are needed when studying peer effects, or that applying a carefully chosen set of covariates is sufficient, or that endogeneity is not a threat at all. Given the size and direction of the parameter estimate and given the strong arguments for the possible dangers of endogeneity in estimating peer effects (Evans et al., 1992; Harker & Tymms, 2004; Hauser, 1970; Nash, 2003), it seems reasonable to assume that endogeneity is a potential problem that researchers should carefully take into account when modeling peer effects.

The coefficient on the difference between studies from the fields of Social Sciences and Economics suggests that the differences between the two research traditions translates itself in a difference in reported effect estimates. Ceteris paribus,

Table 4
Parameter estimates and (standard errors) for the fixed effects meta-regression models.

	Without added effect estimates	With added effect estimates
Compositional variable is dichotomously based (vs. is a composite)	−0.259 (0.064)**	−0.254 (0.061)**
>1 average SES-variable in one model	−0.225 (0.104)*	−0.244 (0.083)**
Test (omitted category is language)		
Math	0.004 (0.012)	0.004 (0.012)
Science	−0.056 (0.048)	−0.058 (0.048)
18 minus age	−0.011 (0.005)*	−0.011 (0.005)*
zGINI	−0.006 (0.009)	−0.006 (0.009)
Prior attainment NOT included as a covariate	0.000 (0.166)	0.217 (0.079)**
N	172	185

* Significant at 0.05 level.
** Significant at 0.01 level.

Social Scientific studies report effect estimates that are about 0.13 smaller. This is contrary to our expectations: the studies from the field of Economics in our sample generally confined themselves to an attempt to obtain unbiased estimates of only the peer effect, while the Social Scientific studies had a much broader goal. Studying peer effects was often only one of their aims. We therefore expected Economics studies to give less biased and lower effect estimates. One possible explanation for this result could be that the Social Scientific studies, using models that for various reasons contained many covariates, coincidentally reached the same results as the Economics studies reached using models that were specifically designed to obtain unbiased estimates. (If we add up the coefficients on discipline and on attempting to overcome omitted variables bias/endogeneity, we find that Social Scientific studies, ceteris paribus, find about the same results as Economics studies that did attempt to overcome this bias.) The large sets of covariates included by some Social Scientific studies often included variables that could actually be seen as part of the peer effect or as a channel through which it works. Examples of this are learning climate or average motivation in the class or school and teacher characteristics. Climate and motivation may be affected by the average socioeconomic status and in turn themselves affect learning outcomes. Teacher characteristics may be affected by average socioeconomic status in that schools with a low-SES intake have difficulties in finding good teachers (Clotfelter, Ladd, & Vigdor, 2006; Hanushek, Kain, & Rivkin, 2004) and thus end up with lower quality teachers. This teacher quality in turn affects students' outcomes. Taking up such variables as covariates may artificially explain away the peer effect. Such covariates are not valid substitutes for a well-thought-over strategy to deal with the problems of estimating unbiased parameters, but might coincidentally lead to the same results. So, two sources of bias may compensate each other by chance.

A potential concern in the present analysis is that the results may be determined to a disproportionally large extent by a few studies that contribute a high number of effect estimates. This may be a concern, even though in our weighting procedure these studies do not receive an extraordinarily high weight. In the Appendix, we present the results from a meta-regression which does not include these studies by the OECD on PISA-data and show that our results are robust to the exclusion of these studies.

Another concern is that the results may be influenced by publication bias: even though we include both published and non-published studies, perhaps studies have a higher chance of appearing if they do find substantial effects. A test for this is in the correlation between the standard errors and the effect sizes reported in studies. The effect size researchers are expected to find, should be independent of their sample size or, equivalently, the precision of, or standard errors to their effect size. However, the smaller the sample size, the more variation there will be in the effect sizes researchers will actually find. In the classical publication bias pattern, some studies with a small sample size and small effects will not be published, whereas studies with a small sample size and large effects have a higher chance of appearing. This creates a negative correlation between effect size and sample size or, equivalently, in this case a positive correlation between effect sizes and their accompanying standard errors (Begg, 1994). We find a correlation of 0.29 ($p < 0.001$), which suggests that there may be some publication bias. However, this correlation is entirely due to variation between the various effects reported by McEwan (2003): in his twin fixed effects estimates, his sample size is reduced from 163,075 to 443. His estimates consequently become very imprecise, which is reflected in a strongly increased standard error, while his point estimates also go up. When we remove this one study, the correlation becomes −0.05 ($p = 0.48$), which indicates no publication bias.

5.2. Estimates from study fixed effects meta-regressions

The results from the study fixed effects meta-regressions are presented in Table 4. The left column shows the results from the estimation that only includes effect estimates that were also included in the basic meta-regressions discussed before. The characteristic that has been most often compared within studies is test subject. Again, the peer effect turns out to have the same size for language, mathematics, and science tests. Most other coefficients fall within the 95% confidence interval of the estimates from the basic meta-regression as well. The most notable exception is the coefficient on inclusion of a prior attainment covariate, which has an estimated value of 0.00. This coefficient was only identified by the study of Strand (1997)

that looks at 6.5-year-old students. For such young children, prior attainment, its measurement and inclusion as a covariate may be of a somewhat different nature and the effect of inclusion may arguably differ from that in samples with children later in their school career. We should therefore be careful in interpreting this coefficient. The coefficients on dichotomously based vs. composite compositional variables and on including more than one compositional variable in one model are also identified through one study only. The second model, which includes some previously excluded estimates, does not have this limitation. The results from this estimation almost all lie easily within the 95% confidence interval of the parameters from the basic meta-regression. This is an important finding. Although not for all of the characteristics from our original model, parameters could be estimated, the fixed effects analysis confirms the robustness of the estimates from our basic model for most of the important sample and model characteristics which we theorized might have an impact on the size of the peer effect. Only the coefficient on age has changed sign and is now significant, which suggests that the peer effect is stronger for older children.

6. Conclusion

The aim of this meta-analysis was to systematically review the findings from previous studies into peer effects on student achievement and to try to come to an understanding of why researchers have alternately found small effects, large effects, or no effects at all.

The results show that the approach a researcher takes for estimating the peer effect of socioeconomic status strongly affects the effect size found. The average weighted effect size over all our studies was 0.32. The exact size a researcher will find, however, may deviate considerably from this, depending on the operationalization of the average SES-variable and the model specification chosen. Choosing a dichotomously based compositional variable, such as free lunch eligibility, or including several average SES-covariates in the same model, leads to a very low and attenuated estimate of the peer effect. The use of a thoroughly constructed composite that includes several of the dimensions of SES is associated with much higher effects than the use of SES-measures based only on parental education or home resources. Our results also suggest that a researcher examining peer effects would generally be strongly advised to include a control for prior attainment in some form. Not doing so would lead to a strong upward bias in effect estimates.

In contrast to the strong relations between the operationalization of the SES-variable, the model specification chosen and the measured size of the peer effect, there was little evidence for an effect of sample choice on the peer effect. The effect did not differ between language, mathematics, and science tests, nor did it differ between countries. There was some evidence suggesting that the peer effect is stronger for older children. Robustness-checks we performed using a fixed effects meta-regression, a promising advancement on current meta-analytic techniques, supported our conclusions.

Although many scholars have described problems due to endogeneity and omitted variables in estimating peer effects, very few studies formally take them into account in their models when estimating the SES-peer effect. Only studies from the field of Economics sometimes explicitly tried to overcome omitted variables bias, often by including many covariates. Studies in the field of the Social Sciences never used this strategy explicitly. The results of our meta-analysis however do not give strong indications for the biasing role omitted variables bias and endogeneity play in the estimation of the peer effect. The studies that used an explicit strategy to deal with such bias, found somewhat lower effects, although the difference was not significant. The number of studies using such a strategy was relatively limited, however, and the strategies they used were arguably not capable of completely getting rid of all omitted variables bias. In all cases, there might have been some bias left and estimates from perfectly unbiased strategies might deviate to some extent. We found that studies from the field of Social Sciences, ceteris paribus, found smaller effects than studies from the field of Economics. These results seemed surprising because several Social Scientific studies in our sample lacked a focus on unbiased estimation of only the peer effect. We argued that some Social Scientific studies that artificially explained away the peer effect by including covariates such as learning climate and teacher characteristics might be accountable for the reported lower effect sizes. Alternative explanations would be that either the strategies used by the Economics studies in our sample that explicitly tried to overcome omitted variables bias/endogeneity were somewhat flawed, or that endogeneity and omitted variables do not play a seriously biasing role here. Without solid proof for the latter, we suggest that it is best to consider omitted variables/endogeneity as a possibly serious problem and that it is advisable to use solid strategies aimed at overcoming it. Studies in education that focus on many issues at the same time and that do not fine-tune their estimation models to the bias-free estimation of the compositional effect may otherwise run a serious risk of obtaining only biased estimates of the compositional effect.

We argued for a number of best choices a researcher could make when examining the SES-peer effect: measuring SES by a thoroughly constructed composite that includes several of the dimensions of SES, not including more than one average SES-covariate in one regression model, controlling for prior attainment, and dealing with the risk of bias due to omitted variables/endogeneity in a correct way. A counterfactual estimate for the effect that such a hypothetical "ideal" study would find, shows that increasing peer SES with one student-level standard deviation is associated with an increase in test scores of about 0.31 of a standard deviation. Because of the large standard error to this estimate and since such an "ideal" study has not been carried out yet, it would be hard to argue that this is "the" exact size of the peer effect. Especially, studies that in a better way deal with the risk of overestimation due to omitted variables/endogeneity, may find that the true effect size is lower. There is clearly still a need for such studies. Findings from such studies can help to increase the quality of future research, both quantitative and qualitative, on peer effects. Our results, however, do suggest that the SES of a student's

Table A1
Parameter estimates and (standard errors) for the meta-regression model excluding estimates derived from the OECD-studies.

	Basic meta-regression excluding estimates from OECD-studies
Constant	0.296 (0.061)**
Compositional variable is (omitted category is "composite")	
Parental education	0.004 (0.038)
Parental occupation	−0.123 (0.056)*
Home resources	−0.076 (0.048)
Dichotomously based	−0.179 (0.039)**
>1 average SES-variable in one model	−0.083 (0.035)*
SES-variable is measured at cohort-/school level (omitted category is "at class level")	−0.093 (0.042)*
Test (omitted category is language)	
Math	−0.021 (0.029)
Science	−0.076 (0.052)
18 minus age	0.009 (0.006)
Prior attainment NOT included as a covariate	0.079 (0.039)*
Does NOT attempt to overcome omitted vars bias	0.109 (0.051)*
Social Sciences (omitted category is Economics)	−0.188 (0.040)**
N	76
R^2	0.67
Systematic variance component (σ_θ^2)	0.0038 (0.0013)**

  * Significant at 0.05 level.
  ** Significant at 0.01 level.

classmates potentially has a substantial effect on her test scores and that obtaining unbiased estimates of this effect, taking into account the pitfalls we discussed, is worth pursuing. If the effect is indeed as large as this meta-analysis suggests, this would have some important implications for school choice and school accountability debates. School choice usually increases sorting of students with similar SES into similar schools. It might therefore lead to a widening of the achievement gap, as high-SES students would profit from having high-SES peers, whereas low-SES students would miss these benefits from attending school with high-SES peers. School accountability systems that judge schools based on their students' test scores would put low-SES schools at a disadvantage, since the SES-peer effect would make it more difficult for them to induce their students to high performance. Correcting for individual students' backgrounds would not be sufficient to deal with this. High-SES schools, on the contrary, would reach good scores with relatively little effort, because their students' performance is boosted by the SES-peer effect. Whether this effect is indeed as large as this meta-analysis indicates, should follow from future research, that takes into account the quality criteria identified here.

Acknowledgements

Appendix A. Estimate excluding the OECD-studies on the PISA-data

Table A1 shows estimates for the basic meta-regression in which the estimates contributed by the four OECD-studies on the PISA-data are excluded. We conduct this analysis in order to check whether our results are sensitive to the large number of estimates contributed by these studies. In this regression, GINI, which indicates social inequality within countries, has been excluded as a covariate, because removing the OECD-studies substantially reduced the variation on this variable. Our results are robust to the removal of these studies: the constant, which indicates the effect for a study making all the "best" choices, remains virtually unchanged and for most predictors, sign and significance stay the same. The parameters indicating the type of SES-variables used by a study change somewhat, but the main finding here, that composite measures are related to stronger effects than measures which only capture a single aspect of SES, is confirmed.

References[10]

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, *114*(2), 533–576.
*Bankston, C., III, & Caldas, S. J. (1996). Majority African American schools and social injustice: The influence of de facto segregation on academic achievement. *Social Forces*, *75*(2), 535–555.

---

[10] Articles marked with an asterisk were included in the meta-analysis.

*Bankston, C., III, & Caldas, S. J. (1998). Family structure, schoolmates, and racial inequalities in school achievement. *Journal of Marriage and the Family*, *60*(3), 715–723.

Begg, C. B. (1994). Publication bias. In H. E. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York, NY, US: Russell Sage Foundation.

*Bondi, L. (1991). Attainment at primary schools: An analysis of variations between schools. *British Educational Research Journal*, *17*(3), 203–217.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models applications and data analysis methods*. Newbury Park, CA, US: Sage Publications.

*Caldas, S. J., & Bankston, C., III. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research*, *90*(5), 269–277.

*Caldas, S. J., & Bankston, C., III. (1998). The inequality of separation: Racial composition of schools and academic achievement. *Educational Administration Quarterly*, *34*(4), 533–557.

CIA. (2007). The World Factbook 2007: CIA.

Clotfelter, C., Ladd, H., & Vigdor, J. (2006). Teacher–student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778–820.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York, NY, US: Russell Sage Foundation.

*De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal*, *29*(6), 841–860.

Duncan, O. D., Featherman, D. L., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York, NY, US: Seminar Press.

Evans, W. N., Oates, W. E., & Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, *100*(5).

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, *32*(3), 505–523.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). *New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement*. NBER working papers, no. 8741. National Bureau of Economic Research.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*(2), 326–354.

*Harker, R., & Nash, R. (1996). Academic outcomes and school effectiveness: Type "A" and Type "B" effects. *New Zealand Journal of Educational Studies*, *31*(2), 143–170.

*Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, *15*(2), 177–200.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology*, *75*(4, Part 2), 645–664.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, *65*(6), 1541–1545.

Hedges, L. V. (1994). Fixed effects models. In H. E. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York, NY, US: Russell Sage Foundation.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological methods*, *3*(4), 486–504.

Hill, M. S., & Jenkins, S. P. (2001). Poverty among British children: Chronic or transitory? In B. Bradbury, S. P. Jenkins, & J. Micklewright (Eds.), *The Dynamics of Child Poverty in Industrialised Countries* (pp. 174–195). Cambridge, UK: Cambridge University Press.

*Ho Sui Chu, E., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education*, *69*(2), 126–141.

Hox, J. (2002). *Multilevel analysis techniques and applications*. Mahwah, NJ, US: Lawrence Erlbaum Associates.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, *115*(4), 1239–1286.

Hoxby, C. M. (2000b). *Peer effects in the classroom learning from gender and race variation*. NBER working papers, no. 7867. National Bureau of Economic Research.

*Hutchison, D. (2003). The effect of group-level influences on pupils' progress in reading. *British Educational Research Journal*, *29*(1), 25–40.

Jarrell, S. B., & Stanley, T. D. (2004). Declining bias and gender wage discrimination? A meta-regression analysis. *Journal of Human Resources*, *38*, 828–838.

Lauder, H., & Hughes, D. (1999). *Trading in futures: Why markets in education don't work*. Philadelphia, PA, US: Open University Press.

*Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*(3), 172–192.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA, US: Sage.

Ma, X., & Klinger, D. A. (2000). Hierarchical Linear Modeling of Student and School Effects on Academic Achievement. *Canadian Journal of Education*, *25*(1), 41–55.

*McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, *22*(2), 131–141.

*McEwan, P. J. (2004). The indigenous test score gap in Bolivia and Chile. *Economic Development and Cultural Change*, *53*(1), 157–190.

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, *52*(1), 13.

Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School Effectiveness and School Improvement*, *14*(4), 441–457.

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science and Medicine*, *56*(4), 769.

*OECD. (2001). *Knowledge and skills for life: First results from the OECD programme for International Student Assessment (PISA), 2000* (no. 92-64-19671-4). Paris, France: OECD Publications.

*OECD. (2003). *Literacy skills for the world tomorrow: Further results from PISA 2000* (no. 9264102868). Paris, France: OECD.

*OECD. (2004). *Learning for tomorrow's world first results from PISA 2003* (no. 9264007245). Paris, France: OECD.

*OECD. (2005). *School factors related to quality and equity: Results from PISA 2000*. Paris: OECD.

*Opdenakker, M. C., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement*, *13*(4), 399–427.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3*(3), 354–379.

*Paterson, L. (1991). Socio-economic status and educational attainment: A multi-dimensional and multi-level study. *Evaluation and Research in Education*, *5*(3), 97–121.

*Peetsma, T., Van der Veen, I., Koopman, P., & Van Schooten, E. (2005). *Class composition influences on pupils' cognitive development* (working paper). Amsterdam, The Netherlands: University of Amsterdam: SCO-Kohnstamm Institute.

Raudenbush, S.-W. (1994). Random effects models. In H. E. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York, NY, US: Russell Sage Foundation.

*Rivkin, S. G. (2001). Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review*, *20*(3), 201–209.

Robertson, D., & Symons, J. (2003). Do peer groups matter? Peer group versus schooling effects on academic achievement. *Economica*, *70*, 31–53.

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, *107*(9), 1999–2045.

*Rumberger, R. W., & Willms, J. D. (1992). The impact of racial and ethnic segregation on the achievement gap in California high schools. *Educational Evaluation and Policy Analysis*, *14*(4), 377–396.

*Schindler-Rangvid, B. (2003). *Educational peer effects. Quantile regression evidence from Denmark with PISA 2000 data. Chapter 3 in Do schools matter?* PhD Thesis. Aarhus, Denmark: Aarhus School of Business.

*Schneeweis, N., & Winter Ebmer, R. (2005). *Peer effects in Austrian schools*. London, UK: Centre for Economic Policy Research.

Sirin, S.-R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, *15*(3), 131–150.

*Strand, S. (1997). Pupil progress during key stage 1: A value added analysis of school effects. *British Educational Research Journal*, *23*(4), 471–488.

*Strand, S. (1998). A 'value added' analysis of the 1996 primary school performance tables. *Educational Research*, *40*(2), 123–137.

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1574.

Thrupp, M. (1995). The school mix effect: The history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education*, *16*(2), 183–204.

Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement*, *13*(4), 383–398.

*Willms, J. D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review*, *51*(2), 224–241.

*Young, D. J., & Fraser, B. J. (1992, March). School effectiveness and science achievement: Are there any sex differences? Paper presented at the *Annual meeting of the National Association for Research in Science Teaching*.

*Young, D. J., & Fraser, B. J. (1993). Socioeconomic and gender effects on science achievement: An Australian perspective. *School Effectiveness and School Improvement*, *4*(4), 265–289.

*Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, *19*(1), 75–92.