

The effectiveness of human capital policies for disadvantaged groups in the Netherlands¹

Edwin Leuven
Hessel Oosterbeek

ABSTRACT

This paper presents results on the effectiveness of different human capital policies targeted to disadvantaged groups in the Netherlands. The policies considered are: class size reduction, extra resources for personnel, extra resources for computers, lowering the compulsory school attendance age and increasing the compulsory school leaving age. All results are based on some quasi-experimental research design. Lowering the compulsory school attendance age is the only intervention producing significantly positive effects. For the other interventions substantial positive effects can be ruled out.

I. Introduction

Governments around the world emphasize that investment in education is an important factor for the future wealth of their citizens. Yet, just increasing public (and private) expenditures on education is not a sensible policy. Crucial is to allocate resources as efficiently as possible. While no one will disagree with this general statement, an efficient allocation of the education budget requires knowledge about the effectiveness of separate interventions and policies. There is an inadequate supply of such knowledge because the number of convincing evaluations of education interventions is limited.

This paper summarizes some recent evaluation studies of education interventions in the Netherlands, which have been conducted by research group Scholar. The common element in all studies is that they are careful about the assumptions required to identify causal effects, and that these are relatively unrestrictive. Because of this, we will argue that these studies produce more convincing results. Another common element of these studies is that they pay special attention to the effects for disadvantaged pupils. For some of these interventions this is inevitable since they

¹ Paper prepared for the CESifo/PEPG-joint conference on Schooling and Human Capital Formation in the Global Economy: Revisiting the Equity-Efficiency Quandary.

are targeted towards disadvantaged groups. Other interventions are not targeted towards a specific group. For these policies effects for different groups are considered.

The evaluation studies reviewed here deal with the following policy measures/interventions:

- Changes in the age at which young children are allowed to attend school;
- Reducing class size in primary education;
- A scheme that gives extra funding for personnel to primary schools;
- A scheme that gives extra funding for computers to primary schools;
- Increasing the length of lower vocational programs from three to four years;

The next section summarizes the findings of these studies. For (technical) details we refer to the respective research papers. The aim of this summary is twofold. The first purpose is to present a number of interesting and policy relevant research outcomes. While the studies certainly do not cover the entire spectrum of possible interventions, the results indicate that some types of interventions may be more effective than others. As a by-product, the summary of recent research demonstrates different methods to attain credible identification of the effects of policies. We return to these issues in the final section.

II. Summary of evaluation studies

II.A. Effect of extra time in school on early test scores

The age at which pupils are allowed or required to starting school differs across countries and is within countries over time. In Scandinavian countries the typical school starting age is 7, it is 6 in most OECD countries, including Canada, Belgium, France, Germany, Italy, Spain and a majority of states in the United State, and 5 in for example the United Kingdom and New-Zealand. At the lower end of the spectrum is the Netherlands, where children are allowed to start attending school at the age of 4.

Whereas this brief description reveals that the school starting age is obviously a policy variable, little is known about the effects of starting at a younger age on achievement. Leuven et al (2004a) investigate this issue by exploiting two specific features of the Dutch regulations with regard to primary school enrollment. The first feature is that a child is allowed to enroll in school immediately after its fourth birthday. This is different from the situation in most other countries where typically all children of the same cohort start at the same day. The second feature is that a

school year cohort consists of the children born between October 1st of a year and September 30th of the next year. As a result children turning age 4 before, in and after the summer holiday are placed into the same class. Together these two features generate – conditional on age - variation in the maximum number of schooldays a child can have attended at any given date during its school career.

Leuven et al (2004a) report significantly positive effects of extra time in school on language and arithmetic scores in grade 2 for disadvantaged pupils. Minority pupils as well as Dutch pupils with low educated parents, benefit from the opportunity to spend more time in school. The effect is quite substantial; one more month of potential school enrollment increases early test scores by 6 percent of a standard deviation. To illustrate, the difference in average test scores of a school without any disadvantaged student and a school with only minority pupils, amounts to one standard deviation. Non-disadvantaged pupils do not benefit from the opportunity of extra time in school at a young age.

The paper argues that the reported effect of extra potential time in school is a lower bound of the effect of extra actual time in school. This implies that we may also expect beneficial effects from lowering the compulsory school attendance age for children with disadvantaged backgrounds. This last result is interesting in light of recent policy discussions in the Netherlands. Some years ago the vice-minister of education proposed to lower the compulsory school attendance age from 5 years to 4 years. But when the cabinet to which this vice-minister belonged was replaced by a new cabinet one of the first actions of the new minister of education was to withdraw this proposal. According to the new minister the proposed change would interfere too much with parents' freedom to choose.

II.B. Effects of class size reduction on achievement in primary schools

The effect of class size on achievement attracted substantial attention from researchers. This is a prime example of a case where simply comparing achievement of pupils placed in small and large classes is likely to give biased estimates of the causal effect. The bias can go either way. If parents who are more interested in the offspring's achievement opt for schools with smaller classes, a naïve comparison is likely to overestimate the true effect of class size on achievement. If instead schools place their more problematic pupils in smaller classes, the same comparison will probably produce an underestimate of the effect of interest.

Well-known are the results of the STAR field experiment conducted in the mid eighties in the state of Tennessee. Pupils and their teachers in kindergarten through third grade were

randomly assigned to classes of different sizes. A careful analysis of the results is reported in Krueger (1999); see also Krueger and Withmore (2002). The main finding is that pupils who have been assigned to smaller classes perform better than the pupils placed in larger classes, both in terms of short-term outcomes and in terms of longer term outcomes. Especially pupils from disadvantaged backgrounds seem to benefit from being placed in smaller classes.

Two reasons limit the external validity of the results of the STAR experiment. First, as is argued by Hoxby (2000), teachers are aware of their assignment to the treatment or control group and this may have a separate impact on their behavior. Second, the experiment was conducted in the mid eighties in Tennessee. Results are not necessarily valid for other populations and/or periods. The average treatment effect of smaller classes is positive, but the fact that disadvantaged students benefit more, already shows that there are heterogeneous treatment effects. This is also illustrated by Figure II in Krueger (1999, p. 526) that shows the distribution of average treatment effects within schools. For a large share of the schools the point estimate of the treatment effect has the wrong sign. Together these two facts illustrate that the overall positive impact of smaller classes depend on the composition of the pupil population and of the schools. In another context (state/country/period), the composition may be different such that the average treatment effect is larger or smaller than the effects reported in the STAR experiment.

A study that does not suffer from the concern of participants being aware of the experimental setting is Angrist and Lavy (1999). That paper uses a regression discontinuity design resulting from specific features of the funding rules applying to primary schools in Israel. According to the so-called Maimonides' rule, an extra teacher is added to a grade level as soon as the number of pupils at the grade level exceeds a multiple of 40. Therefore average class size is expected to equal 40 when the school has 40 pupils at the grade level, while it is expected to equal $20\frac{1}{2}$ when the school has 41 pupils at the grade level. By comparing the achievement of pupils in schools just above and just below the cutoffs, a credible estimate of the causal effect of class size on achievement is obtained. The identifying assumptions are that parents cannot choose schools based on their position around the cutoff, and that no other special events happen precisely at the grade level size of 40. Like Krueger (1999), Angrist and Lavy find positive effects of smaller classes, and effects are larger at schools with higher proportions of disadvantaged pupils.

Given these positive results of class size reduction, it is interesting and important to investigate whether similar effects are realized in other countries. Inspired by Angrist and Lavy's study, Dobbelsteen et al (2002) conducted a similar type of analysis using features of the Dutch funding scheme for primary schools (and Dutch data). The Dutch funding scheme too, has

discontinuities in the relation between number of pupils (at the school level) and teacher formation. The discontinuities are less pronounced than those in Israel. The disadvantage of this is that the size of the treatment is much smaller making it more difficult to identify effects with sufficient precision. The advantage is that the size of the discontinuities in the Dutch funding scheme is fairly close to the reductions in class size that have been implemented in the Netherlands in recent years.

Dobbelsteen et al. report separate effects of class size on language and arithmetic in grades 4, 6 and 8 (grade 8 is the last year of primary school, when pupils are age 11 or 12). Most of the point estimates based on instrumental variable techniques are positive implying that pupils in smaller classes perform worse than pupils in larger classes. The factors underlying this counterintuitive conclusion are further explored by including an extra variable in the achievement regressions. This extra variable measures for each pupil the (absolute) number of pupils in the class that has almost the same level of ability (measured by IQ). This variable intends to measure the peer effects that are predicted by social cognitive learning theories (cf. Bandura 1986; Schunk 1987). According to these theories, pupils benefit most from peers that have about the same level of cognitive ability. The variable “number of similar classmates” has the positive sign predicted by the learning theories, while at the same time inclusion of the variable reduces the positive class size effect. Hence, the non-negative and sometimes even positive relation between class size and achievement in Dutch primary schools can – at least partially – be attributed to the fact that reduction of class size also reduces the (expected) number of pupils in the class with a similar level of competence. This reduction apparently limits a pupil’s scope to learn from her classmates.

While the results reported above are interesting it should be stressed that the analysis is only exploratory. It attempts to find an explanation for the positive class size effects. The results with respect to the effects of the “number of similar classmates” should be interpreted with caution because no attempt has been made to purge these effects for endogeneity.

In a follow-up study, Levin (2001) uses the same specifications but now uses quantile regression techniques. This produces effect estimates at different percentiles of the conditional achievement distribution. Effects are reported for the 10th, 25th, 50th, 75th and 90th percentiles, again separately for language and arithmetic and for grades 4, 6 and 8. The results obtained without inclusion of the number of similar classmates reveals an erratic pattern. Sometimes the estimated effect of class size on achievement is larger at lower percentiles, sometimes the opposite holds. When the number of similar classmates is included the pure class size effect remains irregular. More clear-cut are the results with regard to the effect of the number of similar

classmates. For all grade levels, Levin reports a significant and monotonic decrease in the estimated peer effect when estimated at higher levels of the conditional achievement distribution. This implies that pupils who are in the lower tail of the conditional achievement distribution benefit more from being placed in classes with pupils of similar ability.

II.C. Effect of extra funding for personnel for schools with minority students

In an attempt to improve the performance of schools with a large share of minority students, the Dutch ministry of Education decided in 2000 to give extra funding to primary schools with at least 70 percent minority students. All schools with at least 70 percent minority students received the same extra funding of about 13,000 guilders per teacher. This amount was spread over two years and is slightly less than 10 percent of the annual personnel costs of a school. Schools could spend the additional resources the way they wanted, as long as it contributed to improving teachers' working conditions. Schools with less than 70 percent minority students (including schools with 69.9 percent minority students) were not eligible for the subsidy. To prevent strategic behavior from the schools, the share of minority students was based on the composition of the student population some years prior to the announcement of the policy in 2000.

This feature of the policy creates two different groups: schools with at least 70 percent minority students and schools with less than 70 percent minority students. Leuven et al (2004b) use this feature to evaluate the effects of this policy. By restricting the analysis to schools close to the cutoff, it is likely that there are no systematic differences between the two groups. Leuven et al restrict the analysis to schools with at least 60 percent and at most 80 percent minority students. This choice of bandwidth balances the trade-off between having more comparable schools and having more observations. Assuming that there are no systematic differences just around the cutoff point, this subsidy scheme is like an experiment that assigns schools randomly to the two groups. The empirical analysis also controls for the direct effect of the share of minority students on achievement. This is an example of a policy that treats almost identical cases very differently.

As outcome measure the study takes the scores of students in 8th grade for language, arithmetic and information processing at a nationwide test. Eighty percent of primary schools in the Netherlands participate in this test, and this test is considered to be quite important. Results are used to allot students to different levels of secondary schools. Secondary schools of higher levels require specific minimum levels of performance on this test. Moreover, the average score on this test of the students of a school is used in procedures to assess the performance of schools. This all implies that doing well on the test is important for both students and schools. Outcome

measures are available for 1999, 2000, 2002 and 2003. The first two are pre-intervention measures; the last two are post-intervention measures. As a result the analysis focuses on changes in achievement rather than levels of achievement.

Between the payments of the first and last payment of the subsidy Beerends and Van der Ploeg (2001) interviewed schools' headmasters about the subsidy. The results of these interviews show that around 90 percent of the extra funding schools received is spend in accordance with the ministry's intentions. Hiring and recruitment of extra personnel, extra payments of personnel and extra facilities appear to be the main components. Ten percent of the extra funding was not spend immediately but was added to schools' reserves. It is noteworthy that it was also attempted to evaluate the effectiveness of the scheme. This was done, by asking to headmasters of the treated schools whether the scheme had an effect. Over 80 percent of the respondents gave an affirmative answer. Obviously this method does not meet the minimum standards of a proper evaluation study.

Leuven et al's choice of achievement as a relevant outcome measure reflects the view that ultimately extra personnel, extra payments and extra facilities should translate into higher achievement of students. This seems reasonable given that the extra resources were directed to schools with large shares of disadvantaged students; these students are regarded as disadvantaged because they perform worse (especially on the nationwide test).

None of the effect estimates differ significantly from zero. However, the reported estimates are quite precise so that even modest positive effects can be excluded. For instance, for language the study can rule out effects in excess of 3 percent of a standard deviation with 95% probability.

For the interpretation of this result it is important to realize that the main funding scheme for primary schools in the Netherlands already channels a substantial amount of compensatory resources to schools with large shares of disadvantaged students. In this main funding scheme minority students enter with a weight of 1.9 relative to a unit weight for a non-disadvantaged student. This implies that schools with at least 70 percent minority students already receive over 50 percent more resources than a school with no disadvantaged students. The results from the evaluation study suggest that a level of resource adequacy has been reached (or surpassed).

II.D. Effect of extra funding for ICT for schools with disadvantaged students

A twin of the intervention just described is a subsidy scheme that provided a fixed amount of 209 Dutch guilders per student to all schools with a share of at least 70 percent disadvantaged

students. Disadvantaged students cover minority students and students with lower educated parents as the two main categories. Schools with less than 70 percent disadvantaged students (again including schools with 69.9 percent disadvantaged students) were not eligible for the subsidy. To prevent strategic behavior of the schools, the share of minority students was based on the composition of the student population some years prior to the announcement of the policy in 2000. This intervention is also evaluated in Leuven et al (2004b).

This design of the policy creates two different groups: schools with 70 percent or more disadvantaged students and schools with less than 70 percent disadvantaged students. The study again restricts the analysis to schools with at least 60 percent and at most 80 percent of the targeted type of students. Again assuming that there are no systematic differences just at the cutoff point, this subsidy scheme is like an experiment that assigns schools almost randomly to the two groups. The empirical analysis also controls for the direct effect of the share of disadvantaged students on achievement. Again, very similar cases are treated very dissimilar.

Outcome measures are once more the scores of students in 8th grade for language, arithmetic and information processing at the CITO final test. Test results are available for 1999, 2000, 2002 and 2003. The first two are pre-intervention measures; the last two are post-intervention measures. As a result the study focuses on changes in achievement rather than levels of achievement.

All estimation results reported in Leuven et al (2004b) are negative and in some cases significantly different from zero. This holds especially for language and arithmetic. The extra funding for ICT seems to have adverse effects on students' achievement. Leuven et al also report results from a questionnaire sent to schools in the 65 to 75 percent interval. The questionnaire included items concerning the computer/student-ratio, the "age" of the computers and the intensity of computer use in general and for language and arithmetic in particular. The results reveal no significant differences in computer/student-ratio and age of computers. In both the treatment and control groups this ratio is on average 1:5, which is high compared to the standard of 1:10 in primary schools. There is, however, a significant difference in the amount of school-time students use a computer. Students in the treated schools spend on average 50 minutes per week more using a computer than students in the control group. Part of this extra time is used for language and arithmetic instruction. Hence, extra resources for computers and software increase school-time using a computer and reduce test scores.

The negative findings on computer usage on test scores found in this Dutch study are consistent with findings from two other recent studies. Angrist and Lavy (2002) evaluate the effects of a program in which the Israeli State Lottery funded new computers in elementary and

middle schools in Israel. They use several estimation strategies and find “a consistently and marginally significant relationship between the program-induced use of computers and the 4th grade Math scores”. For 8th graders and for scores on Hebrew, the estimated effects are mostly negative although not significantly different from zero. Rouse et al. (2004) study the effects of the instructional computer program Fast ForWord (FFW). They find no evidence that the use of FFW results in gains in language acquisition or actual reading skills. Interestingly the time spent using FFW was in addition to the amount of time they spent in regular reading instruction. Although Rouse et al. do not find negative effects broader use of computers in instruction is likely to substitute regular instruction. If computer based learning is less effective than more traditional forms of classroom teaching, negative effects cannot be ruled out.

II.E. Effects of extending lower vocational programs from three to four years

It is often argued that low skilled workers should receive more general education or training because it equips them better to participate in the so-called “knowledge economy”. The effectiveness of such an intervention is, however, unknown. In the mid seventies the Dutch government implemented a reform that did exactly what the current proposals aim at. Until then lower vocational education programs had a length of either three or four years. The reform extended the length of all three-year programs to four years, and left the programs that already took four years unchanged. The focus of the extra year had to be on general skills rather than on vocational skills. This change in the program length was accompanied by an increase of compulsory education in the Netherlands from nine to ten years thereby raising the minimum school leaving age from 15 to 16.

Oosterbeek and Webbink (2004) evaluate the effect of the increased program length on the wages of graduates of the extended courses. They use a difference-in-differences (DD) approach where the graduates of the lower vocational courses that did not change in length form the control group. The analysis is related to previous studies that have exploited changes in compulsory school laws to obtain credible estimates of the wage effect of an extra year of schooling.

Aakvik, Salvanes and Vaage (2003) use an increase in the amount of compulsory schooling from seven to nine years in Norway to identify the wage effect of an extra year of schooling. For an extra year of the lowest level of vocational education a return of 0.7 percent is reported. Meghir and Palme (2003) evaluate a social experiment in Sweden. One ingredient of the experiment was an increase in the number of years of compulsory schooling from seven or eight

to nine years. The results suggest that the extra education obtained by those with low ability did not significantly affect their earnings. Oreopoulos (2003) analyses changes in school leaving laws for the US, Canada and the UK thereby concentrating on the effects for dropouts. In the UK, students with less than high school experienced an earnings increase of 5.2 percent as a consequence of the reform. Increasing the minimum school leaving age has substantial positive effects on the number of years of schooling of dropouts in the US and on the highest grade attended by dropouts in Canada. Also the wage effects for dropouts in both countries are substantial.²

Oosterbeek and Webbink fail to find a significantly positive effect of this reform. Their best estimate is -0.018 with a standard error of 0.019 , thereby excluding positive effects of 0.02 or more with 95% probability. This result cannot be explained by the fact that it took some period to fully implement the change. The result may be biased due to changes in the composition in the control group. In that case, the results are likely to provide upper bounds of the true effect. The different results for the simple difference specifications and the DD specifications suggest that previous results based on simple difference specifications are biased.

The findings seem to be at odds with the many studies that report highly significant and substantial returns to a year of schooling. Explanations for this may be that the extra year of schooling did not change the highest degree obtained, and that it is conceivable that the old three-year program was spread out more thinly over the new four-year program. Pischke (2003) offers comparable explanations for his finding of no adverse earnings effect from less time in school. Oosterbeek and Webbink's finding is consistent with that of Pischke and also with the results reported by Meghir and Palme (2003) and Aakvik, Salvanes and Vaage (2003) who also report negligible effects for groups comparable with the group affected by the reform we study.

The findings of Oosterbeek and Webbink suggest that individuals attending lower vocational programs do not benefit (in terms of later wages) of additional general education. This finding contrasts sharply with current policy initiatives that aim to provide young people with minimum levels of general skills. Of course the results relate to a different period of time and a different situation, which limits their external validity. Yet, the results at least cast some doubt on the effectiveness of the current initiatives.

III. Discussion and conclusion

² Related are also the studies by Harmon and Walker (1995), Vieira (1999) and Pischke (2003).

The previous section summarized the results from evaluation studies of five different education interventions. All evaluations are based on some quasi-experimental design. These designs produce results, which have a high degree of internal validity. That is, one can validly conclude from these studies that the differences in outcomes are caused by the differences in treatment (cf. Meyer 1995, p. 152). The possibility that the findings are corrupted by one of the usual threats to internal validity like omitted variables, trends in outcomes, simultaneity, selection or attrition seems negligible. Table 1 lists the key features of each of the studies.

Table 1: Summary of effects

<i>Intervention</i>	<i>Outcome variable(s)</i>	<i>Average effect(s)</i>	<i>Disadvantaged group(s)</i>	<i>Effect(s) for disadvantaged group</i>
Lowering school starting age by one month	Early language and math scores	0.024 (0.016) s.d. (language); 0.022 (0.016) s.d. (math)	Dutch pupils with low educated parents; minority pupils	0.062 (0.034) s.d. (language); 0.061 (0.034) s.d. (math) 0.060 (0.028) s.d. (language); 0.071 (0.029) s.d. (math)
Increasing class size by one pupil	Language and math scores in grades 4, 6, 8	-0.261 (0.270) percentile points (language 6 th grade) to 0.857 (0.367) percentile points (math 8 th grade)	Low achieving pupils (25 th percentile)	-0.245 (0.424) percentile points (language 6 th grade) to 1.050 (0.414) percentile points (math 6 th grade)
Extra resources for personnel	Language, math and information processing scores in grade 8		Minority pupils	-0.055 (0.043) s.d. (language); -0.023 (0.047) s.d. (arithmetic); -0.035 (0.043) s.d. information processing
Extra resources for computers	Language, math and information processing scores in grade 8		Dutch pupils with low educated parents; minority pupils	-0.079 (0.030) s.d. (language); -0.061 (0.033) s.d. (arithmetic); -0.032 (0.030) s.d. information processing
Extending lower vocational program	Earnings after 20 years		Students in lower secondary vocational programs	-0.018 (0.019) percentage earnings change

An important limitation of the results is, however, the extent to which they generalize to other contexts. The fact that Dutch primary schools with around 70 percent of disadvantaged students fail to transform extra resources for computers into higher test scores, does not prove that extra

resources for computers would have no effects at schools that have no disadvantaged students. The Dutch findings about the negative effects of extra resources for computers get, however, more weight when complemented with the other recent findings from other countries and other contexts. At the same time the Dutch findings give more weight to these other studies as well.

When we limit attention to the Dutch context, the evaluations of the five interventions paint a clear picture. Extra resources, for personnel or for computers, for schools with a high share of disadvantaged students have no impact on students' achievement. Class size reduction on the scale recently implemented in Dutch primary education has no beneficial impact on pupils' achievement. And extending the length of lower vocational programs with an extra year of general training has no positive impact on later earnings of graduates. The only intervention that has a clear beneficial effect is to allow young children from disadvantaged families to attend school at a younger age. This policy recommendation concurs with Heckman's (1999) advice to start investing in human capital at a young age.

The summary of findings has also an important methodological edge. Two of the five interventions could be evaluated with a convincing approach due to specific features of the implemented policy. The personnel subsidy for primary schools treats schools with at least 70 percent minority students very differently from schools with less than 70 percent minority students. The computer subsidy for primary schools treats schools with at least 70 percent disadvantaged students very differently from schools with less than 70 percent disadvantaged students. In both cases the specific design of the policy offers a neat research design, but in both cases this was not intentional. The policy makers were not aware of the evaluation opportunities they created. This gives rise to two recommendations. First, it seems likely that more policies with discontinuous treatments have been implemented without the intention to use this for evaluation. Researchers should utilize these features whenever possible. Second, the examples show that it is possible to implement policies that treat very similar cases rather differently. Policymakers should take advantage of this by including such features in the designs of new programs. These features can then be exploited to evaluate the policy.

References

- Aakvik, A, K. Salvanes and K. Vaage, 2003, Measuring heterogeneity in the returns to education in Norway using educational reforms, IZA Discussion Paper No. 815.
- Angrist, J.D. and V. Lavy, 1999, Using Maimonides' rule to estimate the effect of class size on scholastic achievement, *Quarterly Journal of Economics* 114, 533-575.

- Angrist, J.D. and V. Lavy, 2002, New evidence on classroom computers and pupil learning, *Economic Journal* 112, 735-765.
- Bandura, A., 1986, *Social Foundations of Thought and Action: A Social Cognitive Theory*, Englewood Cliffs, NJ, Prentice Hall.
- Beerends, H. and S. van der Ploeg, 2001, Onderzoek vergoeding schoolspecifieke knelpunten, Report OA-230, Regioplan.
- Currie, J., 2001, Early childhood interventions, *Journal of Economic Perspectives* 15, 213-238.
- Dobbelsteen, S., J. Levin and H. Oosterbeek, 2002, The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition, *Oxford Bulletin of Economics and Statistics* 64, 17-38.
- Harmon, C. and I. Walker, 1995, Estimates of the economic return to schooling for the United Kingdom, *American Economic Review* 85, 1278-1286.
- Heckman, J.J., 1999, Policies to foster human capital, Working Paper 7288, NBER.
- Hoxby, C.M., 2000, The effects of class size on student achievement: new evidence from population variation, *Quarterly Journal of Economics* 115, 1239-1285.
- Krueger, A.B., 1999, Experimental estimates of education production functions, *Quarterly Journal of Economics* 115, 1239-1285.
- Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2004a, The effect of potential time in school on early test scores, Working Paper.
- Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2004b, The effect of extra funding for disadvantaged pupils on achievement, Working Paper.
- Leuven, E. and H. Oosterbeek, 2004, Evaluating the effects of a tax deduction on training, *Journal of Labor Economics* 22, 461-488.
- Leuven, E., H. Oosterbeek and B. van der Klaauw, 2004, The effect of financial rewards on students' achievement: evidence from a randomized experiment, Working Paper.
- Levin, J.D., 2001, For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement, *Empirical Economics* 26, 221-246.
- Meghir, C. and M. Palme, 2003, Ability, parental background and education policy: empirical evidence from a social experiment, Working Paper 5/03, IFS, London.
- Meyer, B.D., 1995, Natural and quasi-experiments in economics, *Journal of Business and Economics Statistics* 13, 151-161.
- Oosterbeek, H. and D. Webbink, 2004, Wage effects of an extra year of lower vocational education: Evidence from a simultaneous change of compulsory school leaving age and program length, Scholar Working Paper 44/04.

- Oreopoulos, P., 2003, Do dropouts drop out too soon? International evidence from changes in school-leaving laws, NBER Working Paper Series No. 10155.
- Pischke, J.-S., 2003, The impact of length of the school year on student performance and earnings: evidence from the German short school years, NBER Working Paper Series No. 9964.
- Rouse, C.E., A.B. Krueger and L. Markman, 2004, Putting computerized instruction to the test: A randomized evaluation of a “scientifically-based” reading program, *Economics of Education Review*, forthcoming
- Schunk, D.H., 1987, Peer models and children’s behavioural change, *Review of Educational Research* 57, 149-174.
- Shadish, W.R., T.D. Cook and D.T. Campbell, 2002, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin Company.
- Vieira, J.A.C., 1999, Returns to education in Portugal, *Labour Economics* 6, 535-542.