

**THE IMPACT OF 'EQUAL
EDUCATIONAL OPPORTUNITY' FUNDS.
A REGRESSION DISCONTINUITY DESIGN**

ERWIN OOGHE

April 2011



**FLEMOSI DISCUSSION PAPER DP8:
IMPACT 'EQUAL EDUCATIONAL OPPORTUNITY' FUNDS**

This paper was written as part of the SBO-project “FLEMOSI: A tool for ex ante evaluation of socio-economic policies in Flanders”, funded by IWT Flanders. The project intends to build ‘FLEmish MOdels of Simulation’ and is joint work of the Centre for Economic Studies (CES) of the Katholieke Universiteit Leuven – the Centre for Social Policy (CSB) of the Universiteit Antwerpen– the Interface Demografie of the Vrije Universiteit Brussel – the Centre de Recherche en Économie Publique et de la population (CREPP) of the Université de Liege and the Institute for Social and Economic Research (Microsimulation Unit) of the University of Essex.

For more information on the project, see www.flemosi.be.

THE IMPACT OF 'EQUAL EDUCATIONAL OPPORTUNITY' FUNDS.
A REGRESSION DISCONTINUITY DESIGN

ERWIN OOGHE (*)

April 2011

Abstract: Many countries provide extra resources to schools serving disadvantaged pupils. We exploit a discontinuity in the assignment of such personnel subsidies in Flanders to estimate the impact on cognitive outcomes via a regression discontinuity (RD) design. Because bias can be substantial in RD designs, we include a bias correction in the specification of the control function. Overall, we find positive effects for mathematics, reading and spelling, but the impact is significant for spelling only. The effects are larger for disadvantaged pupils defined on the basis of family background, smaller – or less reliable – for low initial performers, and again larger at schools that used the resources to foster socio-emotional development.

* Center for Economic Studies, KULeuven

The impact of ‘equal educational opportunity’ funds a regression discontinuity design

Erwin Ooghe*

March 31, 2011

Abstract

Many countries provide extra resources to schools serving disadvantaged pupils. We exploit a discontinuity in the assignment of such personnel subsidies in Flanders to estimate the impact on cognitive outcomes via a regression discontinuity (RD) design. Because bias can be substantial in RD designs, we include a bias correction in the specification of the control function. Overall, we find positive effects for mathematics, reading and spelling, but the impact is significant for spelling only. The effects are larger for disadvantaged pupils defined on the basis of family background, smaller—or less reliable—for low initial performers, and again larger at schools that used the resources to foster socio-emotional development.

Acknowledgments: I would like to thank Lex Borghans, Koen Decancq, André Decoster, Geert Dhaene, Jolien Gay, Steven Groenez, Yuemei Ji, Olivier Marie, Raymond Montizaan, Ides Nicaise, Frank Vandenbroucke, Jan Van Damme, Maarten Vendrik, Jean-Pierre Verhaeghe, Dirk Verwerft as well as seminar participants in Leuven and Maastricht for helpful comments, and the SiBO-team and the Flemish Department of Education for the data and kind help.

1 Motivation

Investing in the human capital of disadvantaged young children is a policy intervention that promotes both equity and efficiency; see, e.g., Heckman (2006) for a summary. Understanding *why* disadvantaged children lag behind is therefore crucial to

*Department of Economics, KULeuven, Belgium. E-mail to erwin.ooghe@econ.kuleuven.be.

improve the design of education policies. Jacob and Ludwig (2009) provide three possible explanations—the lack of sufficient resources, good practices and good incentives at school—, and they discuss the effectiveness of different policy interventions in each of these areas.

A widespread policy intervention in OECD countries is to provide extra resources to schools or school districts serving large numbers of disadvantaged pupils. Guryan (2001), Card and Payne (2002) and Papke (2005) report on equalization reforms to narrow the spending gap between the different school districts in the US. They find evidence that equalization improved test scores and pass rates, particularly for low-scoring students, and that it led to a reduction in test score gaps between students with a different family background. Ludwig and Miller (2007) analyze “Head Start”, a US federal program to reduce differences in education and health between young children with a different family background. They find a clear drop in mortality rates, but only suggestive evidence of an improvement in educational performance. Van der Klaauw (2008a) finds no evidence that “Title I” funding—a federal program aimed at low-achieving students in schools with high concentrations of disadvantaged students—, improved student outcomes in New York City public schools. Machin et al. (2004) report on the “Excellence in Cities” program that targets resources to schools in disadvantaged urban areas in England to alleviate underachievement. They find a positive but modest impact on test scores and a significant improvement in attendance for 14-year-old children. Leuven et al. (2007) evaluate the effect of two subsidy schemes aimed at schools with large proportions of disadvantaged students in the Netherlands. They find no impact, sometimes even a negative one, on a range of test scores. Also Bénabou et al. (2009) find no evidence that the “Zones d’Education Prioritaire”, a program providing additional resources to disadvantaged school districts in France, had an impact on a range of student outcomes.

We report here on a similar program in Flanders (the North of Belgium) where schools could receive extra personnel subsidies depending on the family background of their pupils. This policy measure was one of the three pillars of a broader ‘Equal Educational Opportunity’ decree, introduced in 2002 following the alarming signals of relatively high inequalities in educational outcomes in Flanders at that time; see, e.g., UNICEF (2002). Some of the key features of the educational system in Flanders—free school choice on the demand side, free entry and autonomy on the supply side, and no central exams—can help to explain the combination of high average test scores and high inequalities.

In this paper, we exploit a discontinuity in the assignment of the extra resources to estimate the impact on cognitive pupil outcomes via a regression discontinuity design. Section 2 provides details about the program and the data. Section 3 presents the empirical set-up, specifies the model via cross-validation, and tests the validity of the identification assumption. Section 4 shows the results and a final section 5 concludes.

2 Program and data

We focus on ‘basic’ education in Flanders: 2349 schools with 33905 full-time equivalent teachers serving 643769 pupils in pre-primary (3-6 years old) and primary education (6-12 years old) at the start of the ‘Equal Educational Opportunity (EEO)’-programme. The EEO-decree of June 2002 stipulates that schools could receive extra personnel subsidies depending on the family background of their pupils. These extra resources are fixed for a period of three years and schools can autonomously decide how to use them, but within *at least* one of the following themes: (1) to remedy lags in cognitive development and to realize value-added, (2) to foster language proficiency, (3) to stimulate a positive self-image and to improve social skills. To sketch the size of the programme, 4671 extra full-time equivalent teachers were hired during the first EEO-cycle (2002-2005) on top of the 101939 regular teachers for the same period, or a 4.58% increase.

The funding formula was based on a disadvantage index, calculated for each pupil as a weighted sum—with a maximum of 1.2—of the following 5 binary pupil indicators (weights between brackets): the pupil is not living with one of the biological parents (0.8), the pupil’s family belongs to a travelling population (0.8), the income of the pupil’s household consists only of replacement incomes (0.4), the mother of the pupil does not have a degree of secondary education (0.6), and—only in combination with one of the former indicators—the language spoken at home is different from Dutch (0.2).¹ Pupils who meet at least one of the first four mentioned pupil indicators are called disadvantaged pupils in the sequel. The disadvantage index of a school is the sum of the indices of their pupils, multiplied with 1.1 if the percentage of disadvantaged pupils is equal to or higher than 80%, and multiplied with 1.5 if the school lies in the regional capital of Brussels. The total budget is allocated in proportion to the disadvantage

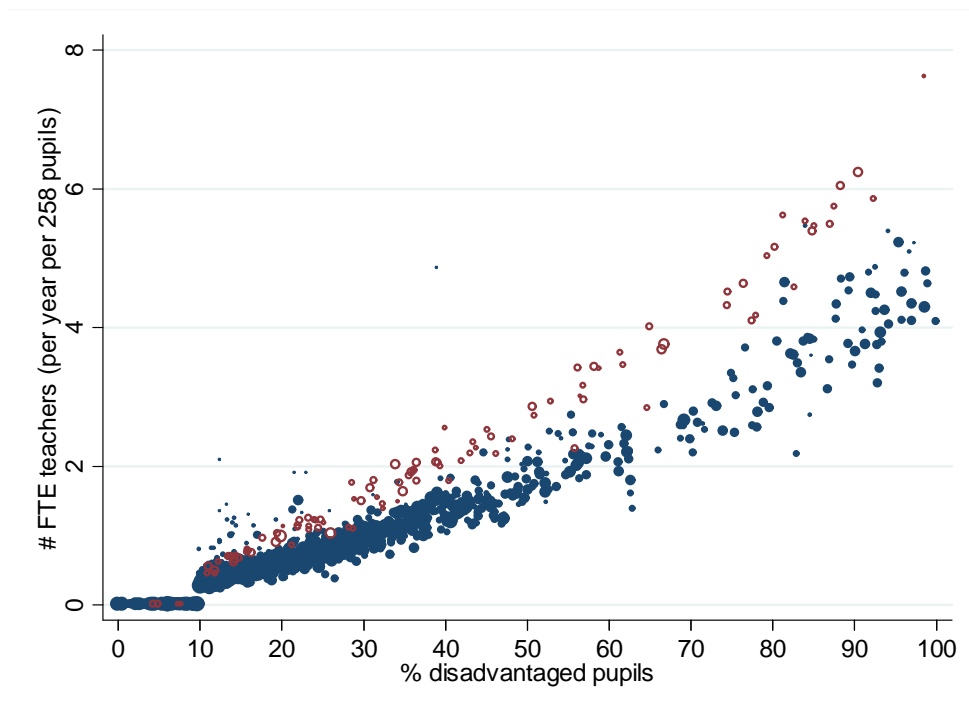
¹Schools have to collect the data and to prove their authenticity via certificates (for the first two mentioned indicators) and via written declarations by one of the parents or foster parents (for the last three indicators).

index of each school, with a minimum of 0.25 full-time equivalents. One interesting exception applies: schools do not receive anything if their percentage of disadvantaged pupils is lower than 10%.

To show the sharp discontinuity, we use administrative data from the Flemish Department of Education. This data set contains all schools in Flanders with the percentage of disadvantaged students (the assignment or running variable), the extra personnel received (treatment variable), a location dummy (Brussels or not) and school size. Figure 1 presents the extra personnel subsidies as a function of the percentage of disadvantaged pupils for all schools in Flanders during the first EEO-cycle (2002-2005); the extra personnel is expressed in full-time equivalents per year per 258 pupils (the median school size in February 2004).

Figure 1: Extra resources during the first EEO-cycle (2002-2005)

○ = only Brussels ● = all except Brussels



Note: bubble size is proportional to school size.

Figure 1 shows that schools do not receive extra resources if the percentage of disadvantaged pupils is less than 10%. The difference at the cut-off is equal to one third of a full-time per year. Above the cut-off the resources increase approximately linearly, steeper for schools in Brussels and slightly steeper for schools above 80%. Due to the

minimum amount of 0.25 full-time equivalents some extremely small schools (less than 40 pupils) do receive a relatively large amount of extra resources.

As there are no central exams in Flanders, we use output data from the SiBO-project aimed at describing and explaining differences in the school curriculum of a representative sample of about 4000 Flemish pupils in 120 schools. The data collection started in September 2002, which is—not coincidentally—also the start of the first EEO-cycle in which schools could receive extra personnel subsidies. We have standardized test scores in mathematics and language proficiency at the start of the first cycle (in September-October 2002, at the age of 5, if not retarded) and in mathematics, reading and spelling for the same pupils at the end of the first cycle (in May-June 2005, at the age of 8). We also have an index of socio-economic status, which is based on the education level of the parents, the profession of the parents and the household income; see Reynders et al. (2005) for details. For each school we know the percentage of disadvantaged students and the extra personnel received. The number of pupils to the left and right of the 10% cut-off is equal to 408 and 3400, respectively. In Appendix A we provide some summary statistics.

3 Empirical set-up

At first sight, the idea of a regression discontinuity is simple. If schools do not have perfect control over the percentage of disadvantaged pupils, then the resulting treatment variation near the cut-off is as good as randomized; see Imbens and Lemieux (2007), Van der Klaauw (2008b), and Lee and Lemieux (2010) for overviews. As a consequence, a regression discontinuity (RD) estimate of the (local) average treatment effect can be obtained by estimating β via a regression

$$s_i = \alpha + \beta d_i + \epsilon_i, \tag{1}$$

on the basis of observations close to the cut-off (i.e., satisfying $c - bw < a_i < c + bw$), with s_i a test score result for pupil i , α a constant, d_i the treatment dummy indicating whether pupil i is in a school to the right of the cut-off, ϵ_i an idiosyncratic error term, c the cut-off (10% in our case), a_i the assignment variable (the % of disadvantaged pupils at i 's school), and bw the bandwidth, with bw approaching zero.²

²Because the treatment variable—the number of full-time equivalents—is continuous, we can also estimate the effect of the treatment variable, instrumented by the treatment dummy. Because the

If the variation around the cut-off is as good as randomized, there is no theoretical reason to include baseline covariates and/or fixed effects. Still, it might help to improve the precision of the estimates and it can serve as a robustness check; see Lee and Lemieux (2010). Let $s_{i,0}$ and $s_{i,1}$ denote initial and final test score results respectively, let ses_i be the socio-economic status and let u_i be a pupil-specific effect. If

$$s_{i,0} = \alpha_0 + \delta_0 ses_i + u_i + \epsilon_{i,0}, \quad (2)$$

$$s_{i,1} = \alpha_1 + \beta d_i + \gamma s_{i,0} + \delta_1 ses_i + u_i + \epsilon_{i,1}, \quad (3)$$

then we can difference out the pupil-specific effect to estimate β via

$$\Delta s_i = s_{i,1} - s_{i,0} = \Delta\alpha + \beta d_i + \gamma s_{i,0} + \Delta\delta ses_i + \Delta\epsilon_i, \quad (4)$$

again on the basis of observations satisfying $c - bw < a_i < c + bw$. We call the estimate of β via (4) a difference-in-difference (DID) estimate, here corrected for initial test score and socio-economic status.

Before we present the RD and DID estimates of the local treatment effect, two questions have to be answered. Is the identification assumption of imperfect control valid? And how can we estimate the treatment effect “near the cut-off”? We start with the last question because we need it to handle the first one.

3.1 Bandwidth and control function

How can we estimate the treatment effect “near the cut-off”? We face a classical bias-variance trade-off. Too narrow a bandwidth is not feasible in practice because there are either no observations, or too little observations to obtain a reliable estimate. But the wider the bandwidth, the more (negatively) biased our estimate will be, because test scores decrease on average with the percentage of disadvantaged pupils at school.

Imbens & Lemieux (2008) show that the bias in RD designs is likely to be substantial. Therefore we follow the literature and add a control function—a function of the assignment variable—to the right-hand side of equations (1) and (4). We use a polynomial of order $o = 0, 1, 2 \dots$ and we allow some of the parameters to be different on both sides of the cut-off; more precisely, we add

$$(1 - d_i) \sum_{k=0}^o \zeta_k^- (a_i - c)^k + d_i \sum_{k=0}^o \zeta_k^+ (a_i - c)^k. \quad (5)$$

resulting IV-estimate is equal to β in equation (1) divided by the number of full-time equivalents at the cut-off (approximately 0.4 in the SiBO sample), and because the significance is not affected substantially, we do not report the IV estimates.

The specification is now defined up to a bandwidth bw , an order o for the polynomial, and, new to the RD literature, two parameters ζ_0^- and ζ_0^+ . Note that the average treatment effect at the cut-off equals $\beta + (\zeta_0^+ - \zeta_0^-)$ when adding the above control function. Therefore, only the difference $\Delta\zeta_0 = \zeta_0^+ - \zeta_0^-$ is relevant for our purposes. Choosing an optimal specification for $\Delta\zeta_0$ is equivalent to specifying an optimal bias-correction as proposed by Simar and Wilson (2007) to measure efficiency in a production context.

Inspired by Imbens and Lemieux (2007) and Lee and Lemieux (2010), and to better mimic the estimation process in (1) or (4), we propose a ‘leave-two-out’ cross-validation to assess the predictive performance of different choices for $(bw, o, \Delta\zeta_0)$.³ First, choose a specification $(bw, o, \Delta\zeta_0)$.⁴ Second, select a pair of schools, one to the left and one to the right of the cut-off, and let a_ℓ and a_r denote their percentage of disadvantaged students. Third, use all observations with assignment levels in $[a_\ell - bw, a_\ell[\cup]a_r, a_r + bw]$ to estimate equation (1) or (4), while adding

$$(1 - d_i) \sum_{k=0}^o \zeta_k^- (a_i - a_\ell)^k + d_i \sum_{k=0}^o \zeta_k^+ (a_i - a_r)^k \quad (6)$$

as a control function.⁵ This provides us with a prediction $\widehat{\beta} + \Delta\zeta_0$ of the treatment effect that can be compared with the true effect β , being the observed difference in average test scores between the two selected schools. Fourth, repeat the previous three steps for all pairs of schools ‘close’ to the cut-off and calculate the mean squared error, i.e., the average of $(\beta - (\widehat{\beta} + \Delta\zeta_0))^2$ over the different school pairs.⁶

The cross-validation function can guide us in choosing reasonable specifications. In appendix B we plot the mean squared error as a function of the bandwidth for each order of the polynomial $o = 0, 1, 2$, once without bias correction (dotted lines) by fixing $\Delta\zeta_0 = 0$, and once with optimal bias correction (full lines) by setting $\Delta\zeta_0 = \Delta\zeta_0^*$, with

³Note that the usual leave-one-out procedure performed separately on both sides of the cut-off does not take into account the possible correlation structure in the bias. For example, a specification with exactly the same bias on both sides of the cut-off is still unbiased in estimating the difference, which is what we are ultimately interested in.

⁴In principle, one could allow for a different bandwidth and a different order for the control function on either side of the cut-off. Experiments with this more flexible specification do not change the cross-validation results in a qualitative way.

⁵Note the difference between (5) and (6): the cut-off c is replaced by the new cut-offs a_ℓ and a_r .

⁶We follow Imbens and Lemieux (2007) and use the median value of the assignment variable on either side of the cut-off as border cases to define ‘close to the cut-off’. Because the bias correction is sensitive to this choice, we will report sensitivity results in the next section.

$\Delta\zeta_0^*$ the bias correction that minimizes the mean-squared error for a given bandwidth and order. The difference between a dotted line and the corresponding full line is the squared bias.

We retain the following three guidelines. The bias correction can be substantial—up to half a standard deviation—, and therefore we will only report results with bias-correction. Given the bias-correction, the mean squared error remains more or less stable given a bandwidth of at least 10. We show estimates for bandwidths from 10 to 80 in steps of 10 in the sequel. Given the bias-correction, order 0 typically performs better than order 1 and order 1 in turn outperforms order 2. As the differences can be substantial, we will report results for order 0 in the main text and provide estimates based on a local linear regression (order 1) as a robustness check.

3.2 Validity

Is the identification assumption of imperfect control valid? Because the funding rules were announced in June 2002, but based on pupil data collected in February 2002, manipulation could only occur if schools anticipated the funding rules (in particular, the 10% cut-off). While manipulation is therefore less likely in the first funding cycle (2002-2005), this is definitely not the case for the second one (2005-2008) as the rules of the game were well-known.

Direct validity tests look for a discontinuity in the density of the assignment variable at the cut-off. Indeed, if schools just below the cut-off would try to get the extra resources by attracting more disadvantaged students or by manipulating the data, we should see an ‘abnormally’ low density to the left of the cut-off and the opposite to the right of it. Figure 2 presents the density of the assignment variable in the first two cycles for *all* Flemish schools. There is some indication of manipulation, especially in the second cycle. But in the first cycle it is less clear, as expected. To summarize, there is no strong reason to believe that the RD identification assumption for the first cycle would be invalid, while we have to be much more reluctant to make the same assumption in the second cycle.

The visual direct validity test is confirmed by a more formal indirect validity test that checks whether there exist discontinuities in the baseline covariates at the cut-off. The main drivers of test score results are initial test scores—the mathematics and language proficiency test at age 5—and socio-economic status.

Figure 2: Density of the assignment variable in the first two cycles

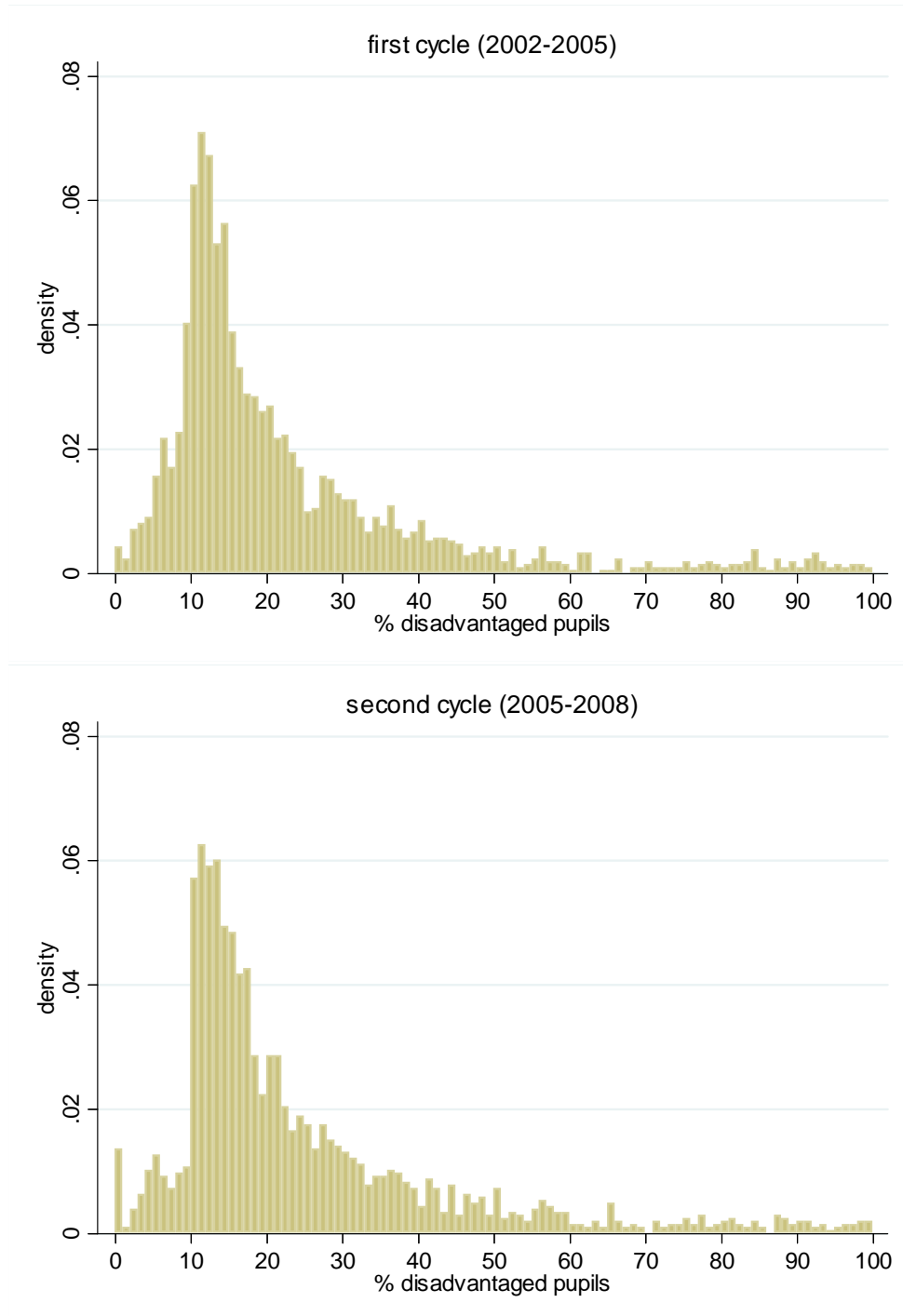


Table 1 presents the estimated difference in initial test score results and socio-economic status for different bandwidths.⁷ The joint null hypothesis—i.e., no differences in the

⁷We estimate a seemingly unrelated regression with initial test scores and socio-economic status as dependent variables and the treatment dummy as the covariate; see Lee and Lemieux (2010).

initial test scores and socio-economic status at the cut-off—, is never rejected for the first cycle (2002-2005).

Table 1: Testing continuity of the baseline covariates in the first cycle (2002-2005)

$bw =$	10	20	30	40	50	60	70	80
$math_0$	0.06	0.06	0.04	0.05	0.04	0.04	0.06	0.05
$lang_0$	0.05	0.05	0.02	0.03	0.03	0.03	0.04	0.04
ses	-0.07	-0.05	-0.05	-0.03	-0.05	-0.05	-0.03	-0.04
$\text{Prob} > \chi^2$	0.20	0.29	0.53	0.56	0.53	0.52	0.47	0.49

This can again be contrasted with Table 2, presenting estimates of the same differences in the second cycle (2005-2008): the null is always rejected.

Table 2: Testing continuity of the baseline covariates in the second cycle (2005-2008)

$bw =$	10	20	30	40	50	60	70	80
$math_0$	0.17	0.18	0.18	0.18	0.17	0.18	0.18	0.18
$lang_0$	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	0.00	0.00
ses	-0.27	-0.26	-0.26	-0.26	-0.26	-0.26	-0.25	-0.25
$\text{Prob} > \chi^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

In appendix C we report estimates based on a local linear regression (order 1) for both cycles. Although the figures can be different, the overall picture is the same: no rejection of the null hypothesis in cycle 1—except for $bw = 10$ —and full rejection in cycle 2. Based on these validity tests, we only report estimates for cycle 1 in the next section.

4 Results

Table 3 presents the RD and DID estimates for the different standardized test scores. Almost all effects are positive, but only the effects for spelling tend to be significant. This overall picture is quite robust. The differences between the RD and DID estimates and between the different bandwidths are small. In addition, the local linear regression estimates in appendix C are also similar: with the exception of reading—the effects for reading become negative, but never significantly different from zero—the overall picture remains the same. In appendix D we repeat the above RD estimates (in the middle

row denoted ‘50|50’) and report estimates when the cross-validation—including the bias correction—is based on larger (first two rows) and smaller subsamples (last two rows). Although the figures change due to differences in the bias correction, the effects remain typically positive, and only significant for spelling.

Table 3: RD and DID estimates of the treatment effect

math	$bw =$	10	20	30	40	50	60	70	80
RD		0.10	0.10	0.09	0.10	0.09	0.09	0.10	0.10
DID		0.08	0.07	0.06	0.07	0.07	0.07	0.07	0.07
read	$bw =$	10	20	30	40	50	60	70	80
RD		-0.01	-0.01	-0.01	0.00	-0.01	-0.01	0.00	0.00
DID		0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
spel	$bw =$	10	20	30	40	50	60	70	80
RD		0.25***	0.26***	0.25***	0.26***	0.25***	0.25***	0.27***	0.26***
DID		0.26***	0.26***	0.26***	0.26***	0.25***	0.26***	0.26***	0.26***

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.

Next, we want to check whether the treatment effect is different for different groups of pupils. Table 4 presents estimates of the treatment effect for advantaged (a) and disadvantaged (d) pupils separately; the DID estimates were again very similar and not reported here.

Table 4: RD estimates for advantaged and disadvantaged pupils

math	$bw =$	10	20	30	40	50	60	70	80
a		0.11	0.10	0.10	0.12	0.11	0.11	0.13	0.14
d		0.31	0.32	0.31	0.29	0.29	0.29	0.29	0.26
read	$bw =$	10	20	30	40	50	60	70	80
a		-0.04	-0.04	-0.04	-0.03	-0.04	-0.04	-0.02	-0.01
d		0.39*	0.38*	0.36*	0.35*	0.36*	0.38**	0.36*	0.33*
spel	$bw =$	10	20	30	40	50	60	70	80
a		0.24***	0.26***	0.26***	0.27***	0.27***	0.27***	0.29***	0.30***
d		0.38*	0.39*	0.36*	0.34	0.33	0.34	0.32	0.29

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.

Due to the small number of disadvantaged pupils in the control group (34 pupils only), one should be cautious when interpreting these effects. Still, the results suggest that disadvantaged pupils benefit more from the extra resources according to all tests.

We find a similar, but less pronounced picture if we look at the dependence of the treatment effect on socio-economic status. This stands to reason, because socio-economic status and disadvantage do correlate: being disadvantaged explains about 37% of the variation in socio-economic status. We split up all pupils to the left of the cut-off in three equally sized groups according to socio-economic status and compare them with groups of pupils to the right of the cut-off based on the same quantiles. From Table 5 we infer again that individuals with a low and middle socio-economic status tend to gain somewhat more for math and reading, while for spelling the evidence is less clear for pupils with a low socio-economic status.

Table 5: RD estimates according to socio-economic status

math	$bw =$	10	20	30	40	50	60	70	80
	low	0.23**	0.22**	0.20**	0.21**	0.20**	0.20**	0.20**	0.19**
	mid	0.24***	0.23***	0.23***	0.24***	0.24***	0.24***	0.26***	0.28***
	high	0.10	0.09	0.12	0.14	0.14	0.14	0.16	0.18
read	$bw =$	10	20	30	40	50	60	70	80
	low	0.08	0.09	0.07	0.07	0.07	0.08	0.08	0.07
	mid	0.02	0.03	0.04	0.04	0.04	0.04	0.05	0.06
	high	0.00	-0.02	0.01	0.02	0.02	0.02	0.03	0.04
spel	$bw =$	10	20	30	40	50	60	70	80
	low	0.25*	0.27*	0.25*	0.26*	0.25*	0.25*	0.26*	0.24
	mid	0.36***	0.37***	0.39***	0.39***	0.39***	0.39***	0.42***	0.43***
	high	0.28***	0.29***	0.31***	0.33***	0.32***	0.33***	0.35***	0.37***

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.

Table 6 looks at the effect for pupils with low, middle and high initial test scores. For mathematics we use the initial mathematics scores, while for reading and spelling we use the initial language proficiency scores to split up the sample in 3 subgroups as before. The picture is less clear now. With the exception of one estimate for spelling, low initial performers never significantly improve their test scores. Middle and also high initial performers benefit more of the extra funds.

Table 6: RD estimates according to initial performance

	$bw =$	10	20	30	40	50	60	70	80
math	low	0.18	0.18	0.17	0.18	0.16	0.16	0.17	0.16
	mid	0.07	0.09	0.12	0.13	0.12	0.13	0.14	0.16
	high	0.12	0.19***	0.21***	0.23***	0.22***	0.23***	0.24***	0.26***
read	low	-0.10	-0.12	-0.12	-0.11	-0.11	-0.11	-0.11	-0.11
	mid	0.05	0.07	0.08	0.08	0.09	0.09	0.10	0.10
	high	0.07	0.09	0.08	0.09	0.07	0.07	0.09	0.10
spel	low	0.20	0.20*	0.19	0.20	0.19	0.19	0.20	0.19
	mid	0.39***	0.43***	0.43***	0.43***	0.42***	0.43***	0.45***	0.46***
	high	0.21	0.24*	0.26*	0.25*	0.24*	0.25*	0.27**	0.28**

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.

This is not necessarily contradicting the previous tables: for example, being disadvantaged is only a weak signal of initial test score performance: it explains 9.46 % and 5.47% of the variance in initial maths and language proficiency, respectively.

Finally, recall that schools could autonomously decide how to use the extra personnel subsidies, but within *at least one* of the following three themes: (1) to remedy lags in cognitive development and to realize value-added, (2) to foster language proficiency, (3) to stimulate a positive self-image and to improve social skills. In our sample, the percentages of pupils within these themes are equal to 76%, 52% and 43%, respectively. Because the themes are chosen by the school, the DID design seems more appropriate to control for the potential endogeneity problem (but again, RD estimates point to the same qualitative result).

Table 7 presents the DID estimates of the treatment effect within the different themes. Schools that worked on remediation (theme 1) did slightly better for math and spelling, but worse for reading. Schools that focussed on language proficiency (theme 2) did worse, not only for mathematics but more surprisingly, also for reading and spelling. Finally, schools that worked on socio-emotional skills (theme 3) always performed better, on average at least. The figures in Table 7 are average treatment effects, so one could still wonder whether schools within themes 1 and 2 do better to remedy pupils who lag behind. In appendix E we presents estimates for the different themes for low

initial performers only. Roughly speaking, the same picture emerges: socio-emotional development is more effective to foster cognitive test scores.

Table 7: DID estimates of the treatment effect for the different themes

math	$bw =$	10	20	30	40	50	60	70	80
	1	0.13	0.10	0.10	0.10	0.09	0.09	0.09	0.08
	2	-0.22**	-0.12	-0.10	-0.06	-0.07	-0.06	-0.06	-0.06
	3	0.21*	0.19*	0.17	0.19*	0.19*	0.18	0.19*	0.19*
read	$bw =$	10	20	30	40	50	60	70	80
	1	-0.10	-0.07	-0.06	-0.04	-0.04	-0.04	-0.02	-0.02
	2	0.05	0.00	-0.02	-0.02	-0.03	-0.02	-0.03	-0.03
	3	0.13	0.09	0.05	0.05	0.05	0.04	0.02	0.02
spel	$bw =$	10	20	30	40	50	60	70	80
	1	0.26**	0.27**	0.26**	0.26**	0.25**	0.26**	0.28***	0.26**
	2	0.16	0.16	0.19*	0.20*	0.19*	0.20*	0.19*	0.19*
	3	0.49***	0.48***	0.46***	0.47***	0.46***	0.44***	0.42***	0.42***

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.

It is not clear to what extent the DID estimates sufficiently control for endogeneity, but the results suggest that fostering socio-emotional skills (e.g., a positive self-image) is more effective in improving cognitive test scores. This is in line with the evidence in Borghans et al. (2008) and Cunha et al. (2010) showing that non-cognitive skills influence cognitive test scores.

5 Conclusion

In this paper, we exploit a discontinuity in the assignment of extra personnel subsidies in basic (i.e., pre-primary and primary) education to estimate the impact on cognitive outcomes via a regression discontinuity (RD) design. As bias can be substantial in RD designs, we propose to include a bias-correction in the specification of the control function. Overall, we find robust positive effects for mathematics, reading and spelling, but the effects are only significant for spelling. The effects tend to be larger for disadvantaged pupils defined on the basis of family background, and smaller—or less reliable—for low initial performers. This suggests that social inequality, i.e., the dependence of outcomes on family background, has decreased, while output inequality, the dependence

of outcomes on initial test score results, has increased at the same time. We also find that the impact is larger for pupils at schools that used the resources to stimulate the socio-emotional development of their pupils.

References

- [1] Bénabou, R., Kramarz, F. and Prost, C., 2009, The French zones d'éducation prioritaire: Much ado about nothing?, *Economics of Education Review* 28, 345-356.
- [2] Borghans, L. Meijers, H., and Ter Weel, B., 2008, The role of noncognitive skills in explaining cognitive test scores, *Economic Inquiry* 46(1), 2-12.
- [3] Card, D., Payne, A., 2002, School finance reform, the distribution of school spending, and the distribution of student test scores, *Journal of Public Economics* 83, 49-82.
- [4] Field, S., Kuczera, M., and Pont, B., 2007, *No More Failures. Ten Steps to Equity in Education*, OECD education and training policy.
- [5] Guryan, J., Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts, NBER working paper 8269.
- [6] Heckman, J., 2006, Skill formation and the economics of investing in disadvantaged children, *Science* 312, 1900-1902.
- [7] Cunha, F., Heckman, J., and Schennach, S., 2010, Estimating the technology of cognitive and noncognitive skill formation, *Econometrica* 78(3), 883-931.
- [8] Imbens, G., Lemieux, T., 2008, Regression discontinuity designs: a guide to practice, *Journal of Econometrics* 142, 615-635.
- [9] Jacob, B. and Ludwig, J., 2009, Improving educational outcomes for poor children, *Focus* 26(2), 56-61.
- [10] Lee, D., Lemieux, T., 2010, Regression discontinuity designs in economics, *Journal of Economic Literature* 48, 281-355.
- [11] Leuven, E. and Lindahl, M., Oosterbeek, H., Webbink, D., 2007, The effect of extra funding for disadvantaged pupils on achievement, *The Review of Economics and Statistics* 89(4), 721-736.
- [12] Ludwig, J. and Miller, D., 2007, Does head start improve children's life chances? evidence from a regression discontinuity design, *The Quarterly Journal of Economics* 122(1), 159-208.

- [13] Machin, S., McNally, S. and Meghir, C, 2004, Improving pupil performance in English secondary schools: excellence in cities, *Journal of the European Economic Association*, 2(2-3), 396-405.
- [14] Papke, L., 2005, The effects of spending on test pass rates: evidence from Michigan, *Journal of Public Economics* 89, 821-839.
- [15] Reynders, T., Nicaise, I., Van Damme, J., 2005, De constructie van een SES-variabele voor het SiBO-onderzoek, LOA-rapport nr. 31.
- [16] Simar, L., Wilson, P., 2007, Estimation and inference in two-stage, semi-parametric models of production processes, *Journal of Econometrics* 136, 31-64.
- [17] UNICEF, 2002, A league table of educational disadvantage in rich nations, Innocenti Report Card 4, UNICEF Innocenti Research Centre, Florence.
- [18] Vanderklaauw, W., 2008a, Breaking the link between poverty and low student achievement: An evaluation of Title I, *Journal of Econometrics* 142, 731-756.
- [19] Vanderklaauw, W., 2008b, Regression-discontinuity analysis: a survey of recent developments in economics, *Labour* 22(2), 219-245.

A. Summary statistics

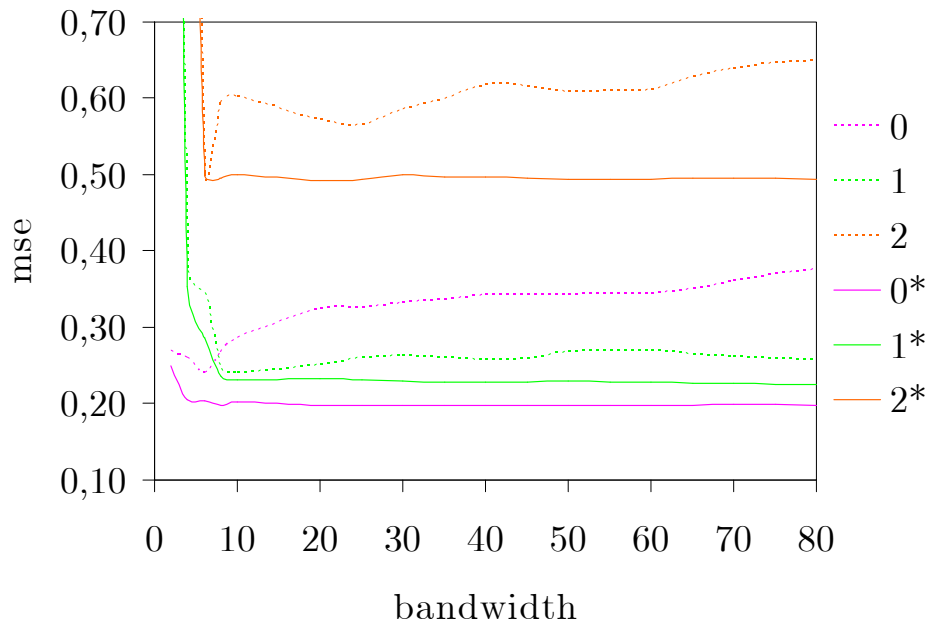
Summary statistics

	mean	stddev	min	p05	p25	p50	p75	p95	max
math ₀	4.26	0.87	0.71	2.72	3.70	4.31	4.92	5.59	5.91
lang ₀	4.55	0.88	1.43	2.83	4.08	4.65	5.15	5.81	6.75
ses	0.06	0.86	-2.36	-1.35	-0.55	0.03	0.68	1.49	2.07
math	9.87	1.00	6.33	8.20	9.18	9.87	10.57	11.50	12.28
reading	3.00	1.00	0.18	1.37	2.27	3.03	3.71	4.63	6.24
spelling	8.84	1.00	3.98	7.29	8.14	8.90	9.37	10.41	12.02
% disadvantaged	21.98	16.61	2.38	6.47	12.58	15.7	25.82	62.73	89.51
# fte teachers	0.82	0.93	0.00	0.00	0.33	0.54	0.96	2.50	5.21

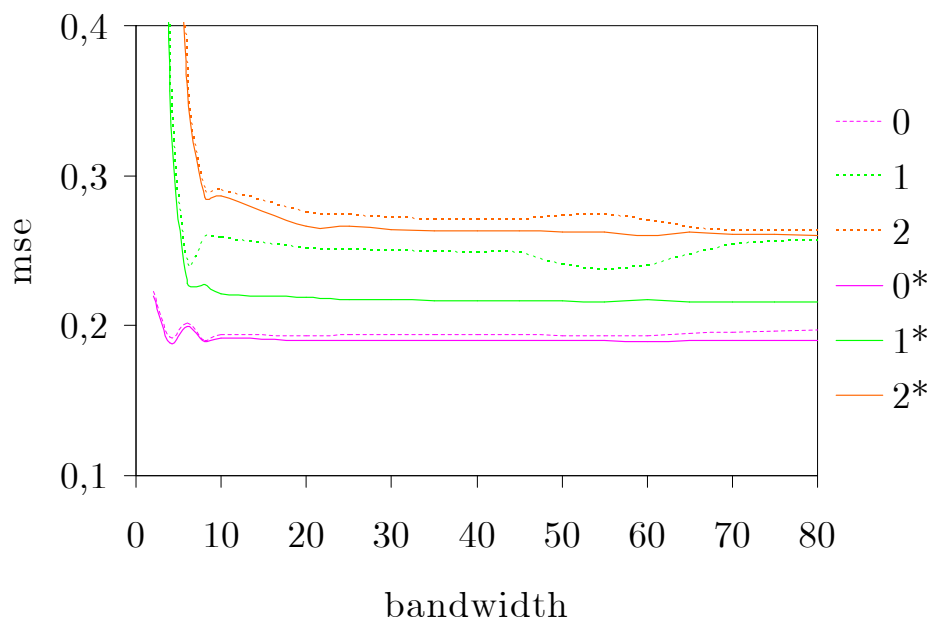
Note: Due to attrition, standard deviations differ from 1 for the initial test scores and socio-economic status. Statistics in the last two rows are calculated at the school level.

B. Cross-validation functions

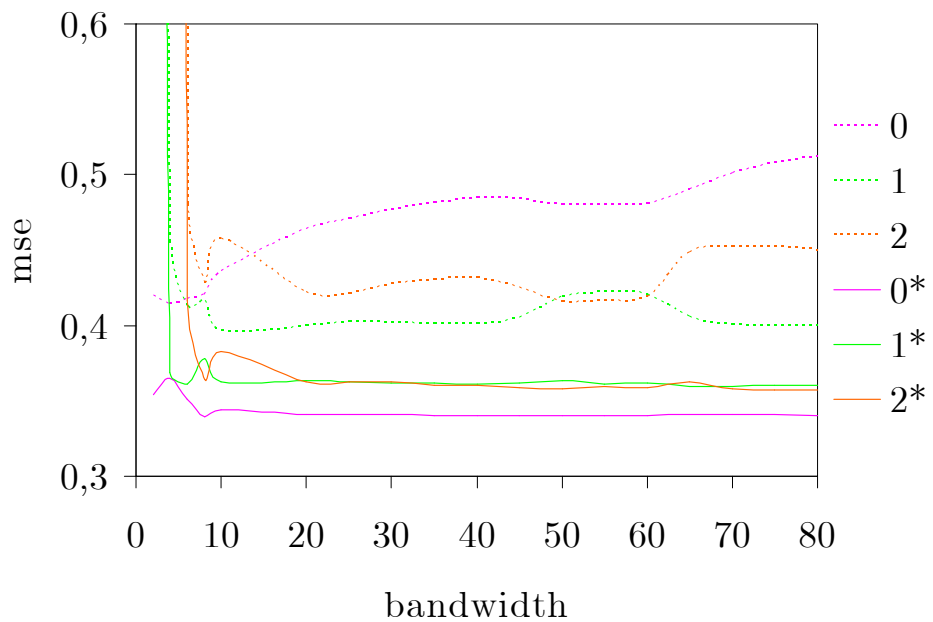
Cross-validation for mathematics



Cross-validation for reading



Cross-validation for spelling



C. Local linear regressions

Testing continuity of the baseline covariates for cycle 1 and 2

	$bw =$	10	20	30	40	50	60	70	80
	$math_0$	-0.03	0.04	0.08	0.06	0.07	0.06	0.04	0.05
1	$lang_0$	0.17	0.14	0.17	0.15	0.14	0.13	0.11	0.12
	ses	0.04	-0.03	-0.04	-0.07	-0.04	-0.05	-0.05	-0.05
	$Prob > \chi^2$	0.02	0.20	0.12	0.14	0.24	0.24	0.35	0.33
	$bw =$	10	20	30	40	50	60	70	80
	$math_0$	-0.25	-0.23	-0.19	-0.19	-0.19	-0.20	-0.20	-0.19
2	$lang_0$	0.20	0.25	0.24	0.25	0.24	0.24	0.23	0.23
	ses	-0.08	-0.10	-0.10	-0.12	-0.11	-0.12	-0.12	-0.11
	$Prob > \chi^2$	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01

RD and DID estimates for the different test scores

	$bw =$	10	20	30	40	50	60	70	80
math									
	RD	0.08	0.12	0.12	0.09	0.11	0.10	0.09	0.09
	DID	0.08	0.08	0.06	0.05	0.06	0.06	0.06	0.05
reading									
	RD	-0.19	-0.13	-0.11	-0.11	-0.10	-0.11	-0.12	-0.11
	DID	-0.21	-0.14	-0.12	-0.12	-0.11	-0.11	-0.12	-0.11
spelling									
	RD	0.24	0.26*	0.28*	0.27*	0.29**	0.27*	0.26*	0.27*
	DID	0.20	0.23*	0.24*	0.25**	0.27**	0.25*	0.24*	0.25**

Note: *, ** and *** mean significance at the 90%, 95% and 99% confidence level.

D. Sensitivity for cross-validation percentiles

RD estimates for different percentiles used in the cross-validation

	<i>bw</i> =	10	20	30	40	50	60	70	80
math									
40 60		0.03	0.02	0.01	0.01	0.01	0.01	0.02	0.02
45 55		0.08	0.08	0.07	0.08	0.07	0.07	0.08	0.08
50 50		0.10	0.10	0.09	0.10	0.09	0.09	0.10	0.10
55 45		0.11	0.11	0.10	0.11	0.10	0.10	0.11	0.11
60 40		0.03	0.04	0.03	0.04	0.03	0.03	0.04	0.04
reading									
40 60		0.10	0.10	0.09	0.10	0.09	0.10	0.11	0.10
45 55		0.02	0.02	0.01	0.02	0.01	0.02	0.02	0.02
50 50		-0.01	-0.01	-0.01	0.00	-0.01	-0.01	0.00	0.00
55 45		0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04
60 40		0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
spelling									
40 60		0.19**	0.20**	0.19**	0.19**	0.18**	0.19**	0.21**	0.20**
45 55		0.24***	0.25***	0.25***	0.25***	0.24***	0.25***	0.26***	0.26***
50 50		0.25***	0.26***	0.25***	0.26***	0.25***	0.25***	0.27***	0.26***
55 45		0.28***	0.29***	0.29***	0.29***	0.28***	0.28***	0.30***	0.29***
60 40		0.32***	0.34***	0.33***	0.34***	0.33***	0.33***	0.34***	0.34***

Note: *, ** and *** mean significance at the 90%, 95% and 99% confidence level. In the first column the symbol 'x' in the notation 'x|1-x' stands for the percentile used to the left of the cut-off.

E. Low performers within the different themes

DID estimates of the treatment effect for the different themes for low initial performers

math	$bw =$	10	20	30	40	50	60	70	80
	low	0.18	0.18	0.17	0.18	0.16	0.16	0.17	0.16
	low 1	0.15	0.07	0.06	0.05	0.03	0.03	0.03	0.01
	low 2	-0.18	-0.06	-0.07	-0.04	-0.05	-0.04	-0.05	-0.04
	low 3	0.35**	0.27*	0.23	0.24	0.24	0.24	0.23	0.23
read	$bw =$	10	20	30	40	50	60	70	80
	low	-0.10	-0.12	-0.12	-0.11	-0.11	-0.11	-0.11	-0.11
	low 1	-0.14	-0.12	-0.11	-0.09	-0.09	-0.09	-0.06	-0.07
	low 2	0.10	-0.06	-0.15	-0.14	-0.14	-0.14	-0.16	-0.16
	low 3	0.12	-0.03	-0.10	-0.09	-0.09	-0.11	-0.17	-0.17
spel	$bw =$	10	20	30	40	50	60	70	80
	low	0.20	0.20*	0.19	0.20	0.19	0.19	0.20	0.19
	low 1	0.13	0.15	0.13	0.13	0.13	0.13	0.18	0.13
	low 2	0.15	0.15	0.11	0.13	0.12	0.12	0.09	0.10
	low 3	0.46**	0.40*	0.32	0.34	0.33	0.32	0.25	0.26

Note: *, ** and *** mean significantly $\neq 0$ at the 90%, 95% and 99% confidence level.