Does targeted school budget support reduce achievement inequality within schools?

Preliminary version

Joppe de Ree and Karen van der Wiel*†

June 29, 2017

Abstract

This paper studies the rationale for and effectiveness of unconditional budget support for primary schools with many disadvantaged students. First, we present a simple economic model that could explain why governments choose to hand out these grants unconditionally and why schools then spend the unconditional funds on remedial education of the under-performers. Second, we study an additional Dutch budget scheme for disadvantaged students at schools in relatively adverse neighborhoods. We show that school boards forward the additional compensatory budget directly to the eligible schools. This results contrasts basic economic theory and indicates the importance of flypaper effects at this level. Moreover we show that schools spend the money on extra teaching personnel, without reducing class sizes. However, by exploiting the geographic cut-off rule in the budget formula, RD-estimates show that extra teachers are not able to improve the performance of disadvantaged students, nor of the other students.

Keywords: Personnel economics, public sector labor markets, economics of education, principal-agent problem, compensatory funding programs, regression discontinuity design

JEL:

*Centraal Planbureau. joppederee@gmail.com

†The authors thank Bert ter Bogt, Dinand Webbink, Bas ter Weel, Statistics Netherlands, the Dutch Ministry of Education, Culture and Sciences.

1 Introduction

Many local and national governments world-wide engage in compensatory school funding programs. Karsten (2006) gives a historical overview of classic programs of positive discrimination such as the American Title I programme, and the English, Dutch, Flemish and French priority areas policies. Such programs typically increase funding levels for schools serving low SES communities. The extra funding is meant to compensate for the lack of support parents and communities are able to provide. The final goal is to reduce gaps in learning outcomes and school success between high SES and low SES students, by improving school success of the low SES students. The Netherlands has implemented such programs ever since the 1970s. Today, with inequalities rising, supporting the lower end of the achievement distribution seems ever more relevant.[1]

This paper looks at the effectiveness of compensatory funding programs in Dutch primary education. Schools receive extra funding based on the number of students with poorly educated parents. Besides biological differences that are transferred from parent to child, poorly educated parents are thought to have more difficulties on average to support their children with homework, etc. Schools can compensate for lower levels of support at home by reducing class size, organizing extra tutoring, or targeted after-school classes. By now, it is well documented that some of these interventions have lasting impacts on the educational performance of disadvantaged students (e.g. Fredriksson, Öckert, and Oosterbeek (2013) and Cortes, Goodman, and Nomi (2015)).

The fact that targeted school funding programs exist widely reveal that governments have a specific interest in reducing inequalities in educational inputs.[2] However, governments often have limited control over how schools – or other legal entities, such as lower level governments

[1]See for example OECD (2011) for evidence of increasing (economic) inequality within countries.

[2]School success is a due to the interplay of school inputs and environmental inputs. Inequalities (of opportunity) arise when environmental inputs (at home for example) differ between groups in the population.

or school boards – use this money. Some level of school autonomy is present in virtually all developed countries, but the type and extent differs. The Netherlands provides an interesting case study in this respect as *Freedom of Education* is embedded in the constitution.[3] Hanushek, Link, and Woessmann (2013) estimate that Dutch schools have among the highest level of autonomy in the world. But the Netherlands is not alone, as many of the (developed) countries listed have comparable levels of autonomy, including Belgium, the UK or Australia etc.

Governments thus provide funds for reducing inequalities in educational inputs, while freedom of education forces them to take a back seat when this money is spent. Whether these compensatory funding programs "work" therefore depends in part on the extent to which schools play along with the Government's objective. Schools might prefer to spend the money on other things, as in a typical principal-agent problem. The Government's objective of the subsidy is reducing inequalities by supporting low SES students. The objective of the school might be to maximize overall school success, not just that of disadvantaged students. Both objectives are not necessarily aligned.

One way to think about this is based on a model of an achievement maximizing school, with a money constraint. Simply relaxing the budget constraint would maximize average learning outcomes as schools would engage in higher return activities (or do more of them). Governments interested in improving average student achievement, would simply provide budget support to schools. Any budget component that is explicitly earmarked by the national government, then, could yield suboptimal returns. When the Government's objective is not maximizing average learning gains, but to reduce inequality, it is not clear why schools would follow.

Whether compensatory funding programs are successful in reducing inequalities in student achievement, therefore, depends on the following chain of events. First, do schools follow the government's objective and use (a substantial part of the) compensatory funding on supporting disadvantaged students? Second, when most of the extra money is actually spent in support of

[3] See article 23 of the Dutch constitution.

the target population, are schools able to reduce learning gaps, by transforming *more money* into *improved achievement* of low-SES students?

As many things can go wrong in this causal chain it is clear one would not necessarily expect to same result in each context. The previous literature on targeted budget support finds mixed results. Early literature focusing on compensatory programs in the US for example find no learning effects (Hanushek). Much more recent research on the other hand does find learning effects of school budget support (Lafortune, Rothstein, and Schanzenbach (forthcoming)). In The Netherlands Leuven, Lindahl, Oosterbeek, and Webbink (2007a) found little impact of additional IT-resources to schools with disadvantaged children.

Our empirical findings are summarized as follows. Based on a regression discontinuity design, we estimate the effects of a targeted 10 percent increase of the school budget. The funding differential maintained for 4-5 years within our sample period. We find that: (a) school boards transfer the money to schools on a 1-to-1 basis, indicating the importance of flypaper effects at this level (Hines and Thaler 1995), (b) schools spend the extra money directly on additional personnel without a meaningful delay. This suggest that money constraints are binding and that there is no substantial uncertainty regarding future money flows, which would justify some smoothing, (c) schools tend not to reduce class sizes, suggesting that the additional personnel is used in other settings (tutoring, remedial education, two teachers per classroom etc.). These results suggests that schools at the margin expect more from hiring remedial teaching staff than from reducing class size. The latter result is important as class size reductions benefit all students, not just the disadvantaged.

We do not find any evidence for average learning effects as measured by test scores from a high-stakes test administered at the end of primary school, nor on a teacher assessment of the student. Not for the average student nor for the group of disadvantaged students.[4] The results

[4]The test is administered by the national testing agency CITO for most (85 percent) of the grade 6 students. The test is administered three-day period and the stakes are high (`http://educatie-en-school.infonu.nl/diversen/128127-de-cito-toets-2015-uitslag-data-en-informatie.html`). The scores

combined suggest that at the margin, schools are not able to successfully translate more money into improved learning outcomes of low-SES students.

We conclude that government objectives are well translated to schools, even in a system with very large school autonomy. However, schools do not seem particularly capable of effectively obtaining the set goals. At the margin the budget support does not measurably reduce achievement inequality on average.

The rest of the paper is organized as follows. In section 2 we discuss the institutional context of the Dutch primary education system and the financing formulas. In section 3 we introduce the regression discontinuity design, the different administrative data records we use, and present the results of our analysis. Section 4 concludes. In the appendix we present our model teacher time allocation.

2 The institutional context and economic theory

Rising inequalities are an important concern for many countries. The existence of compensatory school funding programs provide evidence for this. In the Netherlands for example, the compensatory programs in primary school amount to roughly 5% of the overall primary school budget, roughly 300 million euro each year. Students with poorly educated parents are considered educationally disadvantaged in the current system. Poorly educated parents might not, on average, provide a home-environment as conducive to learning as highly educated parents. The compensatory funding programs in Dutch primary education target this group, on a per student basis. The explicit goal of this program is to help breaking the intergenerational cycle of poor education. The potential lack of support at home can be compensated for at school with smaller classes, additional tutoring, etc. These school-based interventions require extra money which the compensatory program provides.

on the test partly (complementing a teacher recommendation) determines eligibility for education tracks in secondary education.

The traditional freedom that schools have in the way they organize their teaching recognizes the right of parents to choose the kind of education that they want for their children. This liberty however is not complete. The state sets conditions on the quality of education for example and the type of teacher training. The law however does not present a set of well-defined rules and regulations: it does not completely describe where the powers of the state end, and where freedom of education begins. Inevitably, however, freedom of education prescribes that the state takes a back seat in this (social) equilibrium.

As the Dutch government is forced to take a backseat in any school-level decision-making, they face challenges in reducing the inequality of opportunity between socioeconomic groups. Rather than forcing schools to do what it wants, the Dutch government only provides weak incentives:

- m1 They provide extra funding for schools on the basis of the number of educationally disadvantaged students. All else equal, more money goes to schools with more disadvantaged students.

- m2 They compartmentalize funding components. This way the state signals to schools what the money is meant for. Schools typically know that they get extra funding for a specific purpose, in this case, the support of disadvantaged students.

Whether these "soft" approaches are effective in reducing social inequalities depend on the institutional characteristics of primary schools financing, the distribution of educationally disadvantaged students across schools, and the relevance of behavioral mechanisms like the flypaper effect at the level of schools and school boards. The rest of this section introduces the Dutch system of financing alongside a (neoclassical) economic model, in which schools and school boards maximize *average* achievement levels of students under a budget constraint. As this model yields predictions that do not likely reduce inequalities between socioeconomic groups, we show that the government does not believe in the standard economic model, but more in behavioral concepts such as flypaper effects. In the empirical section (3) we assess whether they are right in doing so, by comparing the model predictions to with what happens in reality.

Primary schools in the Netherlands are fully funded out of the state's budget, based on clear costing formulas. The transfer of funds however does not go directly from the state to school, but rather with an important intermediate step. The sequence goes as follows. The funds are disbursed periodically from the national government to school boards. School boards are the legal entities bearing most of the responsibility for management and deployment of resources at the school level. In the Netherlands there are about 1,000 school boards for 7,000 primary schools. School boards have considerable liberty in allocating and deploying funds in the way they see fit. This liberty reduces the influence of the state in funding schools. School boards for example may deviate from the state's costing rules on a "needs basis", and (re)allocate funds from one school to another.

For the purpose at hand we can simplify the costing formulas as follows. Schools receive a fixed component per student of about 3,000 euro plus some funding from additional sources of 700 euro per student. In addition to this, schools receive compensatory funding related to the number of type of disadvantaged student at the school level. The first compensatory program is the *weights program*, where two types of educationally disadvantaged students receive different weights in the costing formula.

$$W_s = (0.3D_s^1 + 1.2D_s^2 - 0.06N_s) \times 3,000 \text{ euro} \geq 0 \tag{1}$$

where N_s is the total number of students at school s, D_s^1 (D_s^2) is the number of disadvantaged students of type-1 (type-2). Students are type-1 disadvantaged when both parents have completed not more than the lowest echelons of vocational training. Students are type-2 disadvantaged when one parent has only primary education while the other has completed not more that the lowest echelons of (post-primary) vocational training. D_s^1 is the number of disadvantaged students of type-1 in school s. Type-2 disadvantaged students count four times as heavily as type-1 disadvantaged students. The weights program also has a built-in lower bound, below which schools are not eligible for funding out of the weights program. Schools are only eligible for this program when more than 20 percent of students are type-1 disadvantaged or when more than 5 percent of students are type-2 disadvantaged.

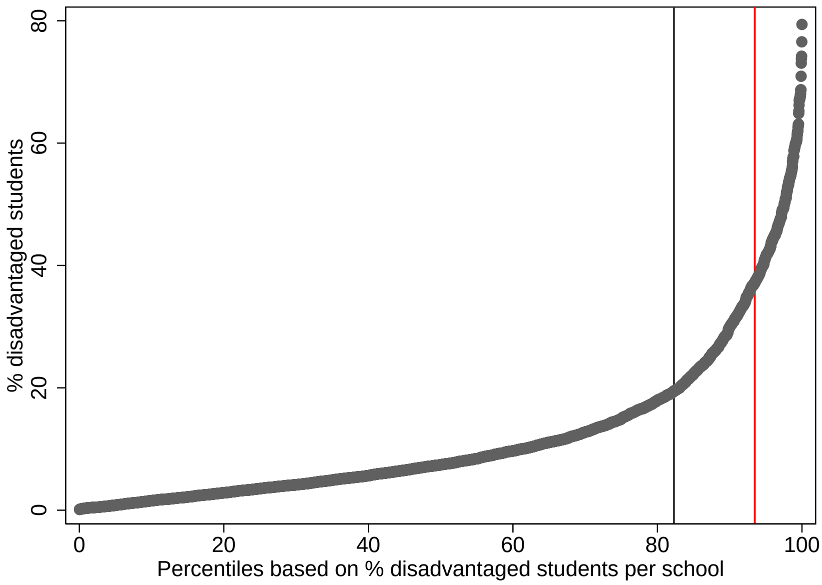
The second compensatory program is the *impulse area subsidy*. This subsidy is most relevant to our analysis as it provides a sharp discontinuity in the eligibility for this program based on neighborhood characteristics within which the school is located. For this program, the two types of disadvantaged students count equally in the costing formula:

$$I_s = (D_s^1 + D_s^2) \times 1,700 \times 1\,(\text{school } s \text{ located in impulse area}) \tag{2}$$

where $1\,(\text{school } s \text{ located in impulse area})$ is an indicator function that is 1 if school s is located in an impulse area, and 0 otherwise. Impulse areas are characterized by relatively high unemployment rates or high percentages of low-income households. In particular, Statistics Netherlands has ranked all (roughly 4,000) 4-digit area codes based on the percentage welfare recipients, and based on the percentage of low-income households. The bottom quintiles of both categories were subsequently classified as impulse areas. The classification if impulse areas has not changed since the impulse area subsidy was introduced for the first time in the 2009/10 school year. The (persistent) discontinuity in the eligibility for the impulse area subsidy provides an opportunity to use a regression discontinuity design, which is discussed further in section (3).

Costing rules (1) and (2) show that the compensatory funding is not precisely proportional to the number of disadvantaged students. The correlation however between the percentage disadvantaged students ($100\% \times D_s/N_s$) and total compensatory funding per student ($(W_s + I_s)/N_s$) is strong ($corr = 0.91$). One other important element here is that there is relatively modest clustering of disadvantaged students within set neighborhoods and schools. Figure (1) below shows how disadvantaged students are distributed between schools.

Figure 1



Notes: 2013/14 data. Black vertical line represents the school above and below which there are half of all disadvantaged students in the Netherlands. Red vertical line represents the school above and below which half of the compensatory budget is spent.

The figure shows that half the disadvantaged students are in schools of which the student population has only modest amounts of these students ($< 20\%$ of the total). The distribution of compensatory funding for disadvantaged students is more skewed than the distribution of disadvantaged students itself. Half of all the compensatory funding is spent at schools with student populations with less than 40% disadvantaged students and about a third is spent on schools with less than 25% disadvantaged students. There exist therefore substantial heterogeneity between Dutch primary schools and school-level responses to the compensatory funding may be very different between these schools.

The fact that these compensatory funding programs exist suggest that the government believes that schools and school boards use the money (at least for a big part) for extra care and support of disadvantaged students. The fact that school boards form a buffer between the state and the school, however makes that there are a lot of ways in which the money can be reallocated. So why does the governments maintain these compensatory funding schemes? Apparently they believe in the idea that schools and school boards actually spend most of this money on the care and support of disadvantaged students. But do they have any theoretical and empirical ground for holding these beliefs? In the remainder of this section we investigate some of these theoretical aspects, after which we evaluate this empirically in section (3).

The allocation of funds across different learning producing activities, can be seen as a two-step approach. In the first step, money is transferred from boards to schools. In the second step the school (under supervision of the board) allocates money across the learning generating activities. In appendix A we work out a theoretical model about the of allocating money across learning producing activities, subject to a budget constraint (exogenously determined by the government's funding rules). In particular we investigate whether there are reasons to suspect that schools and board would spend the compensatory funding on disadvantaged students in the way as foreseen by the government.

The model also has two elements, matching the two steps in the allocation of funds. In the first step, the model is pretty clear in its predictions. If the marginal returns to school level expenditures are decreasing, the board maximizes learning by allocating funds across schools in such a way that the marginal euro per student spent, equals between schools. It may be that some schools get more per student, as at a given level of per student spending, marginal returns differ between schools. It seems plausible therefore that schools with more disadvantaged students have higher levels of funding per student than schools with few disadvantaged students.

However, one clear prediction of this standard model is that when one school under the board is exogenously allocated extra money, that this money is distributed evenly (on a per student basis) across all schools under the board. As the median school operates under a board that also provides management to 13 other schools[5], basic economic theory predicts that most of the

[5]This is based on 2013/14 data from the *dienst uitvoer onderwijs*.

money “leaks” towards other schools. Governments, by virtue of having institutionalized these compensatory funding programs, reveals that it does not believe that this would actually happen. Governments appear to believe in more behavioral concepts, like the flypaper effect (Hines and Thaler 1995). Labeling funding components is enough to motivate boards to transfer the “extra” money directly to schools. There is however a lot of empirical evidence for the existence of flypaper effects in this context.

In the empirical section (3) we find strong evidence for flypaper effects at the level of the school board. The data suggests that school boards forward the extra compensatory funding to schools almost on a one-to-one basis. This empirical fact gives rise to the evaluation of the second step in the budgeting process. Is there any theoretical ground to believe that schools spend the extra money on disadvantaged students, at least in some disproportionate way. Schools for example, may just use the money for class size reduction. But as this would benefit all students in more or less equal ways, most disadvantaged students would not benefit significantly.

The model we have developed to describe this second step in the budget allocation identifies two types of students, students who are *on track* and students who are *behind*. Students who are behind would benefit from remedial education. There is however also a cost to providing the remedial education. As (classroom) teaching has public goods aspects, reallocating teacher time from general learning producing activities that benefit those who are on track, to a smaller group that is behind, would generally hurt more people (those who are on track) than that it helps (those who are behind). A naive conclusion on the basis of this argument would be that theory suggests that schools would not engage much in remedial education of students who are behind.

The model however builds in some indirect benefits to remedial education that seem plausible to us. Remedial education of students who are behind could prepare them better for general classroom activities, which increases the benefits to remedial education. We show that under some conditions schools engage in remedial education. Also, we show that more money could potentially increase the total amount of time spent on remedial education of those who are behind. The link between extra spending and more remedial education of students who are behind (such as the disadvantaged students) is not abundantly clear however. The model therefore (as it currently stands, but it is work in progress) does not yield clear predictions that align with the ideas that the government appear to have. That schools would use the extra compensatory funding predominantly for supporting disadvantaged students.

3 Data, empirical design, and results

3.1 Data

For the empirical analysis we rely on two different administrative data sources. We use school level administrative records from the Dutch ministry of education. These data sets include school level information on numbers of students, by type (disadvangated, or not), the number of teachers, by type, and their pay levels and workload (as measured full time equivalent, or fte). We use data from the past 5 years, from 2010/11 until 2015/16.

The second data source are individual level records of students provided by Statistics Netherlands through remote access. For this analysis we use the individual-level records of the entire population of students enrolled in primary education, from 2008/09 until 2015/16. For this group, we observe household disposable income (through the *Regionaal Inkomens Onderzoek*), place of birth of parents (through the *Gemeentelijke Basisadministratie*). In addition to this we observe school enrollment, class size, and end-of-primary-school test scores. These test scores are high stakes, and partly determine eligibility for track placement in secondary education. These test are developed by the Dutch center of test development (*CITO*). These tests are administered over a three-day period, and have very high rates of internal reliability.

In our analysis we exclude schools who receive small schools subsidies.[6] In 2013/14 bout 1/3 of the schools in the Netherlands receive small school subsidies. However, only 13 percent of the

[6]Schools with less than 145 students are eligible for small school subsidies.

Dutch primary students was enrolled in a small school. Not excluding the small schools yielded similar results, but noisier due to the high variability in per capita funding, due to the small school subsidies. We also exclude about 200 schools without disadvantaged students. The above mentioned selections on the data exclude approximately 17% of the student-level observations from the analysis. The analysis is done on about 4,000-4,500 school-year observations.

3.2 Design and balance tests

The costing rule for the impulse area subsidy gives rise to using a regression discontinuity design (see e.g. Imbens and Lemieux (2008) and Gelman and Imbens (2014) for background). The regression discontinuity design can be used when (eligibility to) treatment is a discontinuous function of some observed continuous assignment variable. In our setup, schools are eligible for impulse area subsidies of roughly 1,700 per disadvantaged student, as long as the school is located in an impulse area.

Impulse areas are characterized by relatively high unemployment rates or high percentages of low-income households. In particular, Statistics Netherlands has ranked all approximately 4,000 4-digit area codes based on the percentage welfare recipients, and based on the percentage of low-income households in the area.[7] The bottom quintile of both categories were classified as impulse areas. This way, roughly 25% of all 4-digit postal code areas were classified as impulse areas. Schools that are located in these areas receive 1,700 per year for each disadvantaged student.

There are two variables based on which areas were classified as impulse areas. Figure (2) presents the fraction on schools in impulse areas, as a function of the two potential assignment variables respectively. Figure (2) shows that the threshold for households on welfare is more informative than the threshold for low-income households, as only a minority of schools to the left of the threshold is located in an impulse area.

[7]The Netherlands has 17 million inhabitants, or 4000-5000 individuals per 4-digit postal code area.

Figure 2: Pairwise relationships between percentage on welfare, percentage low income household, and the fraction of schools located in impulse areas

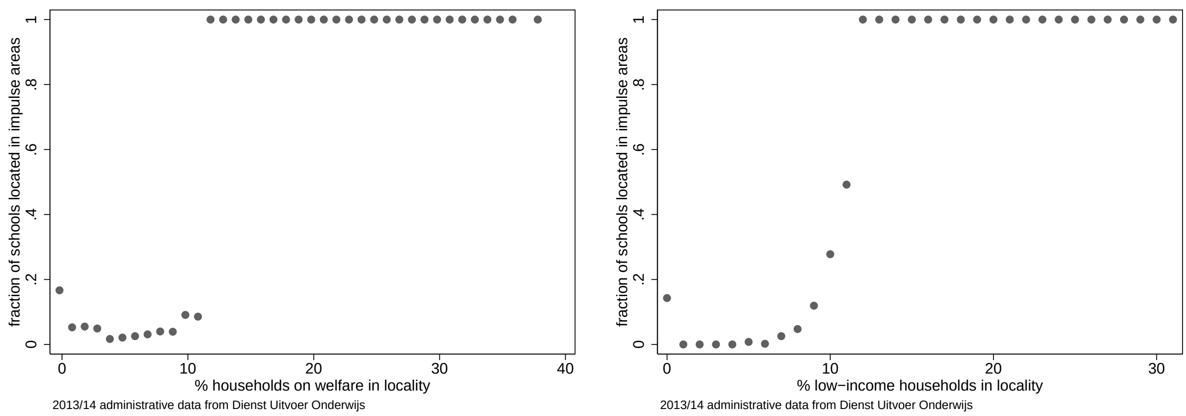


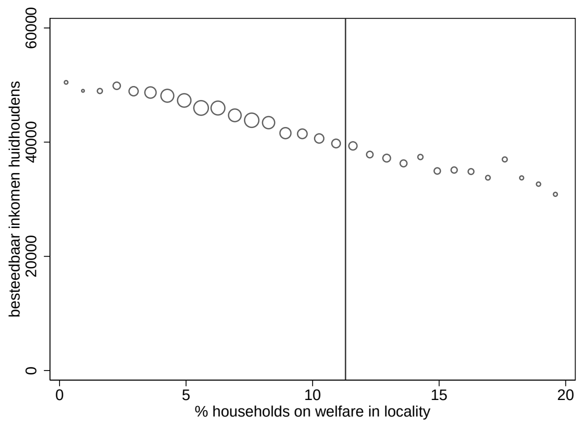
Figure (2 - right) shows that more than 60 percent of schools to the left of the low-income threshold are already located in impulse areas. The cutoff level for more households on welfare seems better suited for a regression discontinuity design. Just left of the threshold, only 10% of schools are located in impulse areas. For reasons of transparency we focus on this discontinuity in the analysis, rather than combining the two assignment variables. Combining two assignment variables is technically possible and potentially more efficient (e.g., Reardon and Robinson (2012)).

The key assumption behind any regression discontinuity design is that (prior to treatment) schools just left of the threshold are comparable on average to schools just right of the threshold. As schools just left of the threshold do not receive treatment, they can be used as a control group for schools just right of the threshold. Schools just right and just left of the threshold therefore need to be reasonably comparable. This assumption can be tested. We do not have access to data from before the 2009/10 school year, the year in which the impulse area subsidy was introduced. [In future versions of this paper we do have this data. Also in this version we have some, see table 4.] We cannot therefore compare treatment and control schools prior to treatment [in the new draft this is updated, based on results from CBS data that dates back to 2008/09]. We do however have some school- and individual-level variables that are not normally affected by the treatment.

In Table 1A we compare treatment and control schools on the total number of students per school, percentage disadvantaged students, household income of students, household income of disadvantaged students, ethnic background of students and the ethnic background of disadvantaged students, for the period from 2010/11 to 2014/15. On the whole we do not find any sizable differences between schools and students just right and just left of the threshold. The only variable for which we find some imbalances (occasionally statistically significant) is the percentage of disadvantaged students. We do not think that this poses a threat to internal validity. In our view the most plausible explanation for this phenomenon is some level of reporting bias. Schools themselves assess education levels of parents, prior to classifying students as disadvantaged or not. If there were actual differences in the number of disadvantaged students, we would expect to see difference in household income and ethnic background as well because these variables correlate strongly with being disadvantaged.

In figure (3) we show some of the results of Table 1 - panel A graphically. Schools just right and just left of the threshold for impulse areas are very comparable on average. Their parents have a similar level of disposable income [left] and a similar ethnic background [right].

Figure 3: Balance on student-level characteristics



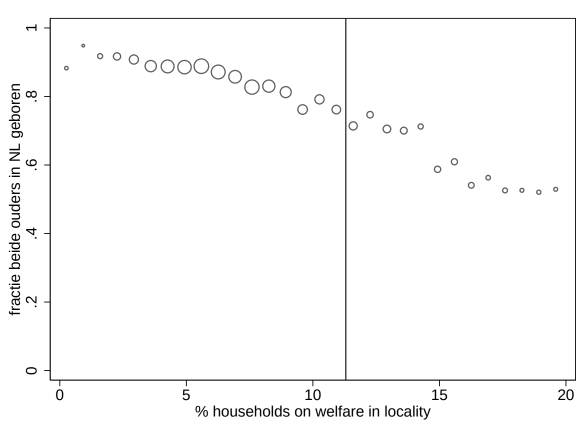


Table 1A shows that the percentage of disadvantaged students just to the left of the threshold for impulse areas is about 15% [this number is not reported in the table in this version of the paper]. From the costing formulas we know that the intensity of the treatment at the school-level, depends directly on the percentage of disadvantaged students. For schools of which 15% of the students are disadvantaged, the impulse area subsidies generate roughly 1,700 euro per disadvantaged student extra, or on a per student basis, $0.15 \times 1,700 \approx 255$ extra.

In the analysis we also look at a selection of schools for which the change in funding at the threshold is stronger – using terminology used in instrumental variables analysis, it has a *bigger first stage*. We have restricted the school-level data on those observations for which the percentage of disadvantaged students is above the median, conditional on the value of the assignment variable.[8] In Table 1B we report on the results of the balance check for this selected sample. Also for this subset of schools we do not find any particular differences between schools and students along the most important dimensions, except for the percentages of disadvantaged students.

3.3 Intermediate results

Schools just to the right of the threshold are eligible for the impulse area subsidy. They, therefore, should receive more money than schools just left of the threshold. Not surprisingly, Table 2 - panel A shows that this is indeed the case. On average, schools just right of the threshold receive roughly 5 percent extra funding per student. This differential is not surprising as it is a direct result of the costing rules. In Table 2 - panel B we have again looked at the selected sample. For the selected sample, schools just right of the threshold receive close to 10 percent more money per student than schools just to the left of the threshold. Table 2 shows furthermore that this difference persists for the entire sample period. The 10 percent extra funding per student amount to about 450 euro per student per year. Per disadvantaged student this amount is significantly

[8]This selection is less likely to be exogenous to treatment as would be a selection purely on the percentage of disadvantaged students as this variable is potentially affected by endogenous reporting bias.

higher, around 1,500 euro.

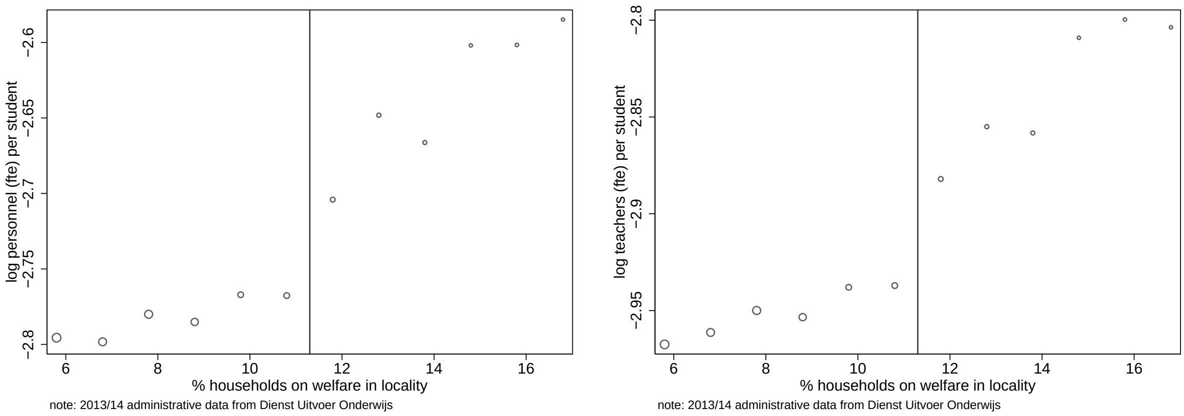
These numbers however only reflect the amount that the ministry of education assigns to the school. In section (2) we mentioned that in the Dutch system the money is not directly transferred to schools. The national government transfers the money to school boards, who, then, allocate the money to schools. As mentioned before, school boards may deviate from these funding rules and reallocate the funds on a needs-basis. Still, as schools seem generally aware of costing rules, school boards need to justify a reallocation of funds to the school leadership.

In Appendix A.1 we derive a model of an average test-score maximizing school board. The model predicts effectively that the extra money from the impulse area subsidies of a particular school under the board is divided equally across all schools under the board, on a per student basis. Our regression discontinuity design provides an ideal setup to test this prediction. Table 3 presents some of our results. Schools just to the right of the threshold also tend to have higher salary payments per student. The effect sizes at the margin are similar to the results for the transfers of funds reported in Table, suggesting that, at the margin, schools boards transfer extra funds to schools on a one-to-one basis, and that schools use the additional funds for hiring extra personnel. These effects are highly statistically significant and persist from 2011/12 to the end of the sample period in 2014/15. For the 2010/11 school year, we do not yet see these effects, indicating that the hiring process took about a year to take effect. These results reject the neoclassical model presented in Appendix A.1.

A next result is that while schools hire additional personnel they do not appear to reduce class size. As reducing class sizes would reduce the burden on the teaching personnel and (might) improve the learning environment for students (Krueger 1999), our suggest that schools and boards expect a higher return from other ways of using trained personnel. Apparently, money is not a binding constraint for class size reduction. This is interesting and suggest that schools have reached some optimal class size level already without the impulse area subsidies (see Lazear (2001) for economic theory on (optimal) class sizes).

The increase in the number of teachers per student at the threshold is also quite clearly visible graphically, even though teacher level data is noisier (not surprisingly) than the budget data. Figure (4) presents the results for the full sample, separately for all personnel [left] and teaching personnel [right]. Figure (5 presents the same variable for the selected sample. Although the discontinuity is visible for all years separately the noisiness of the teacher data sometimes produces odd outliers. For these figures we therefore averaged across the school years 2011/12 to 2014/15, reducing the impact of outliers (that are driven by measurement or coding errors).

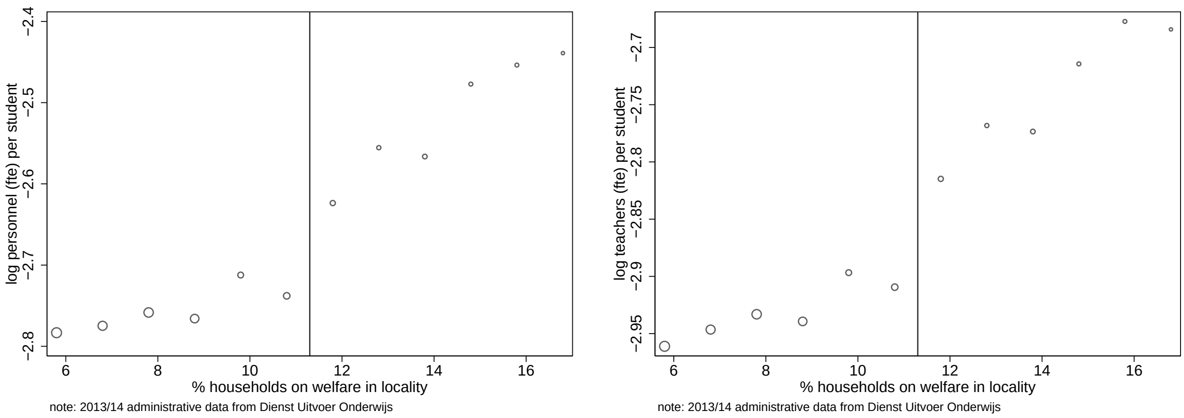
Figure 4: Effects on personnel for full sample. All personnel [left], teaching personnel [right]



Notes: data averaged across 2011/12 - 2014/15 data.

The results so far are not inconsistent with the Government's broad objective. The money is transferred by boards to schools that are entitled to them according the costing rules. And schools hire additional personnel, without reducing class size. The fact that schools do not reduce class size makes sense also aligns with the implicit objective of the government. If schools were in fact using this money to reduce class size, the extra money would reduce classes roughly by 5 (full sample) to 10% (selected sample), a reduction from, say 25 to 22 at best. Such group size reductions are unlikely to generate large learning returns, and would not, in general reduce inequalities as all students would tend to benefit equally.

Figure 5: Effects on personnel for selected sample. All personnel [left], teaching personnel [right]



Notes: data averaged across 2011/12 - 2014/15 data.

3.4 Effects of the subsidies on learning outcomes

A persistent 10 percent increase in funding at the school level has been found to improve learning outcomes (Lafortune, Rothstein, and Schanzenbach forthcoming). Earlier research however has not found support for the idea that money matters (Hanushek (etc.) and Leuven, Lindahl, Oosterbeek, and Webbink (2007b) for examples). These differences in results suggest that the importance of money matters is highly context dependent. This is seen more clearly when we look at the broader literature where some interventions are found effective in some circumstance but not in others. The effects of class size reductions for example, depend strongly on the size of the reference class. Moreover, class size reductions have been found to be more effective with disadvantaged students (see for example Krueger (1999), Krueger (2003) and Fredriksson, Öckert, and Oosterbeek (2013)). This makes sense theoretically as there is some optimal class size (in a cost benefit sense), where the optimum depends on the characteristics of the student population.

Table 4 reports on the impact of extra funding around the cutoff level for impulse areas. At the end of the sample period, at the end of the 2015/16 school year, students have benefited from higher funding levels for 4-5 consecutive years. Test scores however were not measurably affected. Once again we consider the full sample in Table 4A and the *high impact* selection in Table 4B. For all student combined we find point estimates of around zero, with standard errors of around 0.05 SD for the full sample and around 0.07 for the high impact sample. Average effects on test scores bigger than 0.10-0.15σ are statistically rejected. This result does not differ much between the panel A and panel B results.

For disadvantaged students we find more somewhat more variation in the point estimates and slightly bigger standard errors. Still, based on these sets of results, learning effects of extra subsidies bigger than 0.18-0.20σ are statistically rejected. Notice also that the result for the 2015/16 school year are 5-year treatment effects, whereas the literature often reports one-year treatment effects. Experimental estimates of persistence parameters indicate that only 30-70 percent of (nonzero) treatment effects are still measurable two years after the intervention. Based on this range of estimates, a 5-year treatment effect of 0.20 SD's, implies a one-year treatment effect of 0.06-0.14σ. These effect sizes are smaller than effect sizes reported by (Taylor 2015), reporting on the effects of double-dose math teaching, also below effects reported by (Krueger 1999) who report on the effects of class size reductions from 22 to 15 students, and much lower than those reported by Cook, Dodge, Farkas, Fryer, Guyan, Ludwig, Mayer, Pollack, and Steinberg (2014) on high-dosage tutoring.

[Next draft provides cost-benefit calculations and a more complete comparison with the available literature.]

4 Conclusions

In this paper we show that extra school funding meant to support disadvantaged students, i.e. students with poorly educated parents. These students roughly are on average 0.7σ behind the mean and it is plausible that some of this is due to a less supportive home environment. The fact that we do not find any measurable impact and that we can rule out modest impacts on test scores, as well as other measures of school success, suggest that this compensatory funding scheme does not reduce these inequalities.

It remains unclear why *more money does not seem to work*. The usual caveats apply: it may be that schools need even more money to successfully set up programs to reduce inequality. At the margin we evaluate the effects of 10% extra funding per student on average. Some school receive much more than that, which might help. Also we might not measure the right things (test scores and teacher secondary school recommendation[9]). These variables however are very important for the success of students in secondary education. The test combined with the teacher recommendation determine eligibility for educational tracks in secondary school.

It could be that schools do not know precisely how to effectively support disadvantaged students. On the other hand, however they, as our model predicts under some conditions, prefer to not specifically target disadvantaged students (when the benefits do not outweigh the cost). But providing the above mentioned caveats, the results indicate that if governments want to reduce social inequalities in achievement, they need to amend the system. One possibility in this regard is to expand the mandate of the inspection. Today, the inspection evaluates school performance on the basis of test score averages (conditional on characteristics of the school population). If conditional performance falls below a minimum, schools receive a warning. Multiple warnings can mean school closure. The inspection could do the same thing based on achievement scores of disadvantaged students to incentivize schools to limit inequalities.

[9]This is a measure of that correlates with noncognitive abilities once we control for test scores.

A Economic model of teacher time allocation

In this model, schools maximize mean test scores subject to a money constraint. The idea that schools maximize mean test scores makes sense as a baseline model for two reasons. First, the inspection flags schools as potentially underperforming, based on (conditional) mean test scores. Second, mean test scores provide a quality signal to parents (see Koning and van der Wiel (2013) for example who show that this matters for take-up).

Boards allocate money such that mean achievement scores of all students are maximized:

$$\begin{aligned}
Y^* &= \frac{1}{N}\sum_i y_i^* && (3)\\
&= \frac{1}{N}\sum_s \sum_{i\in s} y_i^* && (4)\\
&= \frac{1}{N}\sum_s N_s \left(\frac{1}{N_s}\sum_{i\in s} y_i^*\right) && (5)\\
&= \frac{1}{N}\sum_s N_s Y_s^* && (6)\\
&= \frac{N_1}{N}Y_1^* + \frac{N_2}{N}Y_2^* + ... + \frac{N_S}{N}Y_S^* && (7)
\end{aligned}$$

where each of the individual achievement scores is a function g_i of *expenditures on learning producing activities*. Without loss of generality we consider K different learning generating activities, where E_{ks} the amount of money spent on learning producing activity k. Achievement $y_{i\in s}^*$ is an individual specific function of investments in these activities:

$$y_{i\in s}^* = g_{i\in s}\left(E_{1s}, E_{2s}, ..., E_{Ks}\right) \tag{8}$$

The board maximizes Y^* by allocating total funding X across S schools, and, within each school, across K activities. The board's budget constraint can be written as:

$$\begin{aligned} X &= \sum_s X_s && (9) \\ &= \sum_s \sum_k E_{ks} && (10) \end{aligned}$$

where X_s is the total budget allocated to school s.

Because the education production function can be partitioned in segments (where only the E_{ks} are grouped by school) – separability assumption is imposed –, we can analyze the two stages of this maximization problem separately (see for example Deaton and Muellbauer (1980) p124. for more on this). In the first stage, we investigate how the totality of funding X is allocated between schools. In the second stage, we investigate how schools (supervised by the board) allocate money across the different learning generating activities, conditional on the totality of funding X_s that is allocated to school s.

A.1 The first stage: the model of the school board

The school board maximizes average student achievement across all schools under its management, by allocating all funds across schools.

Boards maximize:

$$Y^* = \frac{1}{N} \sum_s N_s Y_s^* (X_s) \tag{11}$$

subject to the totality of funds under the school board's management:

$$X = \sum_s X_s \tag{12}$$

The Lagrangian for this maximization problem is:

$$L = \frac{1}{N}\sum_s N_s Y_s^* (X_s) + \lambda \left(X - \sum_s X_s \right) \tag{13}$$

The first order conditions of this maximization problem are (in addition to the budget constraint):

$$\frac{N_s}{N}\frac{\partial Y_s^*}{\partial X_s} - \lambda = 0 \;\; \forall s \tag{14}$$

The conditions can be written in terms of per capita expenditures:

$$\frac{1}{N}\frac{\partial Y_s^*}{\partial x_s} - \lambda = 0 \;\; \forall s \tag{15}$$

In other words, school boards maximize mean achievement scores by allocating money such that marginal returns of per capita spending is equalized across schools (and all the money is spent). An important implication of this (not very surprising) result is that when the government decides to reward one specific school under the board with extra funding, boards would react optimally by dividing the money equally (on a per student basis) across all schools. Thereby deviating from what the government had in mind.

A.2 Second stage: the model of the school

In the second stage schools and boards allocate the school-level funding across the learning producing activities. It is possible to analyze this process in general terms. But in this paper we are interested in particular if (and when) it is optimal for a school to invest dis-proportionally in activities specifically directed to disadvantaged students – such as remedial education. And what the impact of extra money or extra personnel might be on time spent on these activities.

The model we develop is inspired by Lazear (2001)'s disruption model of educational production, as well as by Pritchett and Beatty (2012) and Cook, Dodge, Farkas, Fryer, Guyan, Ludwig, Mayer, Pollack, and Steinberg (2015). Much of teaching is has aspects of a public good. Lazear argues that "A class-room almost defines what is meant by a public good". Our model recognizes this. As opposed to Lazear (2001) our model takes class size as exogenous, and focusses on teacher time allocation across students with different starting levels of achievement.

Taking class size as exogenous seems justified by some of our empirical results. More money, at the margin, leads to more teachers being hired, but does not lead to any sizable class size reductions. This suggest that some optimal level of class size is reached and that investments in alternative ways of using time are more profitable. The question we analyse is whether remedial education is economically sensible and what the effects of more personnel are on the amount of remedial education that schools engage in. Notice that because classroom teaching is a public good for a big part remedial education of smaller subgroups have large costs (in terms of forgone learning of the main group).

The model we develop below is about how teachers (supervised by the school leadership) allocate time across different types of students, in an attempt to maximize the mean net learning time of students. Based on this model we analyze how the allocation of teacher time is affected when the budget constraint is relaxed. [*The analysis is very preliminary*]

Suppose that there two types of students, *B* students are behind, and *A* students are on track. *B* students do not learn as much from standard classroom activities, but only a fraction x of what *A* students learn. Remedial education would help them keep up. Teachers can spend part of their time on remedial education – specifically on supporting B students.

t_A is the fraction of available teacher time spent on general classroom activities. t_B is the fraction of teacher time spent on remedial education of students who are behind, *B* students. The total amount of time available to a classroom is $1 + \pi$, where 1 represents the classroom teacher, and π is the amount of extra available teacher time in the school. It is not uncommon for example, that a primary school operates classes but 12 fte teacher time available, so that π would be $\pi = 2/10 = 0.2$. As at the margin, class size is exogenous, and so are teacher's wages, the fraction of extra money available to the school practically corresponds to π.

A student benefit in full from a unit of time spent on general classroom activities. A student do not benefit from remedial education. *B* students learn only a fraction $0 \leq x \leq 1$ from a unit of time spent on general classroom activities (some fraction $1-x$ of it is too difficult for them). *B* students benefit in full from a unit of time spent on remedial education. A key element in our model is that remedial education affects *B* students directly, but also indirectly, as they will benefit more from other activities as well (such as lecturing). The parameter θ measures the fraction of time teachers have to spend on remedial education in order to fully prepare *B* students for the general classroom activities.

If teachers spend a fraction $t_B \leq \theta$, B students learn t_B from remedial education itself *plus* a fraction $\left(x + \frac{t_B}{\theta}(1-x)\right)$ of the time spent on the rest of the activities $(1+\pi-t_B)$. *A* students learn $1+\pi-t_B$, all the time available $1+\pi$, minus the amount of teacher time that is spent on remedial education. Extra personnel means more time available to divide tasks. In this section we evaluate how extra personnel matters for the decision to engage in remedial education. Total time available is now $1+\pi$.

Mean learning time per student as a function of t_B is:

$$MT(t_B) \quad = \quad f_B\left(t_B + \left(x + \frac{t_B}{\theta}(1-x)\right)(1+\pi-t_B)\right) + (1-f_B)(1+\pi-t_B) \quad (16)$$

where f_A and f_B are the fractions of *A* and *B* students.[10]

First order condition, based on derivative w.r.t t_B:

$$MT'(t_B) \quad = \quad f_B + f_B\frac{1}{\theta}(1-x)(1+\pi-t_B) - f_B(x + \frac{t_B}{\theta}(1-x)) - (1-f_B) = 0 \quad (17)$$

[10]The function is quadratic in t_B, with $MT'' < 0$. The function, therefore, has a maximum, somewhere on the t_B domain.

Solve for t_B:

$$\frac{t_B^*}{\theta} = \frac{1}{2}\left[\frac{\theta+1+\pi}{\theta} - \frac{1-f_B}{f_B}\frac{1}{1-x}\right] \tag{18}$$

OR:

$$t_B^* = \frac{1}{2}\left[1+\pi+\theta\left(1-\frac{1-f_B}{f_B}\frac{1}{1-x}\right)\right] \tag{19}$$

The optimal amount of time spent on remedial education depends positively on π, the fraction of extra teacher time available to the classroom.

Suppose parameters are such that there is internal solution for $0 \leq t_B^* \leq \theta$, both for the reference situation, i.e. the control group just to the left of the threshold for impulse areas, and for the treated group just to the right.[11] Take π^C for control and π^T for treatment, where $\pi^C < \pi^T$. The differential tutoring time in equilibrium is:

$$t_B^*(T) - t_B^*(C) = \frac{1}{2}\left(\pi^t - \pi^c\right) \tag{20}$$

B students in treatment schools receive more remedial education, i.e. half of the extra time available in treatment schools. But not only do *B* students in treatment receive more support, because of it, they also benefit more from general classroom activities (such as lectures). The full learning differential between *B* students in treatment and control is bigger therefore than the amount of remedial education they receive. Further versions of this paper go into the comparative statics of this model.

[11] In fact corner solutions are important here. In many situations $t_B^* < 0$, or $t_B^* > \theta$, so that either there's full remedial education or no remedial education at all. In such situations, changes in π, would not affect the optimal teaching strategies – they would just do more of them.

B Construction indicator of noncognitive skills, based on teacher's recommendation

The standardized score on the CITO end-of-school test corresponds to a recommendation for a track in secondary school. 501-523 corresponds to vmbo basisberoeps, 524-529 corresponds to vmbo kaderberoeps (thresholds change over time sometimes), 530-536 corresponds to vmbo gemengd/theoretisch, 537-544 corresponds to havo, and 544-550 corresponds to vwo.

Teachers however do not need to follow this implied recommendation (otherwise, the recommendation would be useless). They may adjust upward or downward, based on additional knowledge he/she has about the child. One of the components here may be related to character, self-confidence, grit, or other "noncognitive" skills. In the data we observe quite quite frequently indeed that teachers deviate from the recommendation that is implied by the test-score.

Teacher recommendations are not always fully precise and may reflect doubt. Teachers for example my recommend havo-vwo, for example, when they are not quite sure. As teachers may indicate a range of school types, we construct two indicators (dummy variables). One dummy equals 1 if a student has an implied recommendation (by the test score) that is lower than the lowest teacher recommendation in the indicated range. The second dummy is 1 when a student has a test score that is higher than the highest teacher recommendation. It happens quite often that teachers "correct" the implied recommendation. In the school year 2013/2014 for example, 12% of students received a teacher recommendation below the test score recommendation, and 21% received a teacher recommendation above the test score recommendation.

Table 1: Balance tests

	(1) 2010/11	(2) 2011/12	(3) 2012/13	(4) 2013/14	(5) 2014/15
	Panel A: Full sample				
school-level					
Number of students per school	9.71	13.12	8.77	12.90	25.53
	(15.55)	(15.46)	(15.49)	(16.08)	(15.81)
Percentage disadvantaged students per school	2.17	2.30	3.18**	2.77*	1.57
	(1.46)	(1.48)	(1.50)	(1.42)	(1.11)
student-level					
Parents born in The Nether1ands	-0.01	-0.03	-0.04*	-0.04	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Disposable household income	639.97	338.08	541.16	1252.84	1203.10
	(996.49)	(881.54)	(928.71)	(1127.49)	(1131.25)
disadvantaged student-level					
Parents born in The Nether1ands (disadv. students)	-0.00	-0.01	-0.02	-0.03	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Disposable househo1d income (disadvantaged students)	-89.77	-234.32	-463.06	-619.09	-459.30
	(534.89)	(509.88)	(520.41)	(543.03)	(569.19)
	Panel B: Percentage disadvantaged students above median				
school-level					
Number of students per school	6.25	10.98	-12.01	-21.16	3.11
	(18.17)	(18.25)	(18.46)	(19.52)	(19.22)
Percentage disadvantaged students per school	3.21	3.82*	4.47**	4.10**	1.82
	(2.23)	(2.12)	(2.11)	(1.98)	(1.54)
student-level					
Parents born in The Nether1ands	-0.01	-0.04	-0.04	-0.03	-0.00
	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
Disposab1e househo1d income	-153.79	-264.48	-1027.15	-689.30	-2.41
	(896.65)	(1018.73)	(1079.56)	(1175.89)	(1146.48)
disadvantaged student-level					
Parents born in The Nether1ands (disadv. students)	0.03	-0.00	-0.01	0.00	0.01
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Disposable househo1d income (disadvantaged students)	-326.02	-776.40	-427.25	-887.17	-644.74
	(665.01)	(607.69)	(676.31)	(661.31)	(632.51)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Effect sizes reported based on RDD. Balance variables: number of students, percentage disadvantaged students, log household income, fraction immigrant. [FOR SOME OF THESE VARIABLES WE CAN GO BACK FURTHER IN TIME.] Student-level models are estimated on school level averages, using the number of students per school as weights. Disadvantaged student-level models are estimated on school-level averages over disadvanged students, using the number of disadvantaged students per school as weights. [We use a bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014).]

Table 2: Process variables

	(1) 2010/11	(2) 2011/12	(3) 2012/13	(4) 2013/14	(5) 2014/15
	Panel A: full sample				
student-level					
log total funding per student	0.05***	0.04**	0.05***	0.03**	0.04***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Funding o.a.b. per student	279.69***	274.49***	279.70***	250.21***	232.09***
	(47.51)	(47.04)	(47.88)	(43.93)	(40.41)
disadvantaged student-level					
log funding o.a.b. per disadvantaged student	0.82***	0.87***	1.08***	0.92***	0.85***
	(0.10)	(0.11)	(0.15)	(0.12)	(0.11)
log total funding per disadvantaged student (under targeting)	0.30***	0.30***	0.30***	0.31***	0.29***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Funding o.a.b. per disadvantaged student	1629.44***	1598.97***	1640.23***	1710.54***	1704.96***
	(126.92)	(128.30)	(130.66)	(158.74)	(132.39)
	Panel B: percentage disadvantaged students above median				
student-level					
log total funding per student	0.10***	0.09***	0.10***	0.11***	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Funding o.a.b. per student	469.10***	456.91***	488.47***	482.87***	429.70***
	(75.98)	(75.36)	(73.08)	(73.02)	(65.48)
disadvantaged student-level					
log funding o.a.b. per disadvantaged student	0.79***	0.81***	1.05***	1.00***	0.89***
	(0.10)	(0.09)	(0.13)	(0.12)	(0.11)
log total funding per disadvantaged student (under targeting)	0.27***	0.26***	0.29***	0.31***	0.28***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Funding o.a.b. per disadvantaged student	1565.63***	1515.44***	1672.59***	1756.15***	1732.59***
	(144.12)	(150.44)	(141.04)	(158.19)	(154.89)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Effect sizes reported based on RDD. Process variables are: fraction impulse areas, log disbursed funds per pupil, ... [We use a bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014). Results don't change much if we use Imbens Kalyanaraman (2012)]

Table 3: Intermediate effects

	(1) 2010/11	(2) 2011/12	(3) 2012/13	(4) 2013/14	(5) 2014/15
	Panel A: full sample				
student-level					
log total salary payments per student	-0.00	0.03*	0.06***	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log number of total personnel (fte) per student	0.01	0.05**	0.07***	0.05**	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log number of teaching personnel (fte) per student	0.01	0.04**	0.06***	0.04**	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log number of management personnel (fte) per student	-0.03	0.05	0.10	0.10	0.04
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)
log number of support personnel (fte) per student	-0.06	0.00	0.09	-0.03	-0.04
	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)
group size (based on median)	1.78*	0.57	0.53	-0.56	-0.25
	(0.92)	(0.74)	(0.77)	(0.83)	(0.79)
group size (based on mean)	1.55**	0.11	0.54	-0.33	-0.20
	(0.75)	(0.60)	(0.66)	(0.74)	(0.69)
	Panel B: percentage disadvantaged students above median				
student-level					
log total salary payments per student	0.02	0.07***	0.10***	0.09***	0.07*
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
log number of total personnel (fte) per student	0.02	0.07***	0.11***	0.10***	0.08**
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
log number of teaching personnel (fte) per student	0.03	0.07***	0.09***	0.07**	0.07*
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)
log number of management personnel (fte) per student	0.07	0.12	0.27***	0.30***	0.17*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
log number of support personnel (fte) per student	-0.02	0.08	0.24	0.12	0.08
	(0.15)	(0.16)	(0.15)	(0.14)	(0.14)
group size (based on median)	1.76	-0.20	-0.14	0.00	-0.45
	(1.35)	(0.80)	(1.01)	(1.12)	(1.10)
group size (based on mean)	1.82*	-0.45	-0.26	0.12	-0.13
	(1.09)	(0.63)	(0.82)	(1.01)	(0.96)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Effect sizes reported based on RDD. Intermediate effects variables are: log salaries per student, log fte teaching personnel per student, class size, support personnel, management personnel. [We use a bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014). Results don't change much if we use Imbens Kalyanaraman (2012), except that the effects on salary payments are more stable. Bigger bandwidths yield stabler results.]

Table 4: Effects on student performance

	(1) 2008/09	(2) 2009/10	(3) 2010/11	(4) 2011/12	(5) 2012/13	(5) 2013/14	(5) 2014/15	(5) 2015/16
	Panel A: full sample							
student-level								
Standardized test scores	0.01	-0.05	-0.01	-0.01	-0.04	-0.01	-0.06	0.01
	0.05	0.05	0.06	0.04	0.04	0.04	0.05	0.05
disadvantaged student-level								
Standardized test scores	-0.04	-0.07	0.01	0.01	-0.03	0.03	-0.02	0.02
	0.08	0.08	0.08	0.07	0.06	0.07	0.08	0.08
	Panel B: percentage disadvantaged students above median							
student-level								
Standardized test scores	-0.05	-0.07	-0.10	-0.07	$-0.15^{**\dagger}$	0.01	-0.10	-0.00
	0.07	0.08	0.07	0.07	0.07	0.07	0.08	0.07
disadvantaged student-level								
Standardized test scores	-0.06	0.06	0.02	-0.06	-0.10	0.09	-0.06	-0.05
	0.08	0.12	0.10	0.09	0.08	0.09	0.11	0.11

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We use a bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014). Two options for obtaining more precision.

References

COOK, P. J., K. DODGE, G. F. FARKAS, R. FRYER, J. GUYAN, J. LUDWIG, S. MAYER, H. POLLACK, AND L. STEINBERG (2014): "The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: results from a randomized experiment in Chicago," *NBER*, (19862).

——— (2015): "Not too late: Improving academic outcomes for disadvantaged youth," *IPR working paper series*, (15-01).

CORTES, K. E., J. S. GOODMAN, AND T. NOMI (2015): "Intensive math instruction and educational attainment long-run impacts of double-dose algebra," *Journal of Human Resources*, 50(1), 108–158.

DEATON, A., AND J. MUELLBAUER (1980): "Economics and consumer behavior," *Cambridge university press*.

FREDRIKSSON, P., B. ÖCKERT, AND H. OOSTERBEEK (2013): "Long-Term Effects of Class Size," *Quarterly Journal of Economics*.

GELMAN, A., AND G. IMBENS (2014): "Why high-order polynomials should not be used in regression discontinuity designs," *NBER*.

HANUSHEK, E. A., S. LINK, AND L. WOESSMANN (2013): "Does school autonomy make sense everywhere? Panel estimates from PISA," *Journal of Development Economics*, 104, 212–232.

HINES, J. R., AND R. H. THALER (1995): "The flypaper effect," *Journal of Economic Perspectives*, 9(4).

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*.

KARSTEN, S. (2006): "Policies for disadvantaged children under scrutiny: the Dutch policy compared with policies in France, England, Flanders and the USA," *Comparative Education*, 42(02), 261–282.

KONING, P., AND K. VAN DER WIEL (2013): "Ranking the schools: How school-quality information affects school choice in the Netherlands," *Journal of the European Economic Association*, 11(2).

KRUEGER, A. B. (1999): "Experimental estimates of education production functions," *Quarterly Journal of Economics*.

——— (2003): "Economic considerations and class size," *Economic Journal*.

LAFORTUNE, J., J. ROTHSTEIN, AND D. SCHANZENBACH (forthcoming): "School Finance Reform and the Distribution of Student Achievement," *American Economic Journal: Applied Economics*.

LAZEAR, E. P. (2001): "Educational production," *Quarterly Journal of Economics*.

LEUVEN, E., M. LINDAHL, H. OOSTERBEEK, AND D. WEBBINK (2007a): "The Effect of Extra Funding for Disadvantaged Pupils on Achievement," *The Review of Economics and Statistics*, 89(4).

——— (2007b): "The Effect of Extra Funding for Disadvantaged Pupils on Achievement," *Review of Economics and Statistics*, 89(4).

OECD (2011): "Divided we stand: why inequality keeps rising," *OECD Publishing*.

PRITCHETT, L., AND A. BEATTY (2012): "The negative consequences of overambitious curricula in developing countries," *CGD working paper*, (293).

REARDON, S. F., AND J. P. ROBINSON (2012): "Regression Discontinuity Designs With Multiple Rating-Score Variables," *Journal of Research on Educational Effectiveness*, 5(1).

TAYLOR, E. (2015): "Spending more of the school day in math class: Evidence from a regression discontinuity in middle school," *Journal of Public Economics*, 117(2).