ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Ooghe, Erwin; Schokkaert, Erik

**Working Paper**

# School Accountability: Can We Reward Schools and Avoid Pupil Selection?

IZA Discussion Papers, No. 7420

**Provided in Cooperation with:**
IZA – Institute of Labor Economics

*Suggested Citation:* Ooghe, Erwin; Schokkaert, Erik (2013) : School Accountability: Can We Reward Schools and Avoid Pupil Selection?, IZA Discussion Papers, No. 7420, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/80587

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 7420

# School Accountability: Can We Reward Schools and Avoid Pupil Selection?

Erwin Ooghe
Erik Schokkaert

May 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# School Accountability: Can We Reward Schools and Avoid Pupil Selection?

**Erwin Ooghe**
*KU.Leuven and IZA*

**Erik Schokkaert**
*KU.Leuven and CORE*

Discussion Paper No. 7420
May 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 7420
May 2013

# ABSTRACT

# School Accountability: Can We Reward Schools and Avoid Pupil Selection?*

School accountability schemes require measures of school performance, and these measures are in practice often based on pupil test scores. It is well-known that insufficiently correcting these test scores for pupil characteristics may provide incentives for inefficient pupil selection. We show that the trade-off between reward and pupil selection is not only a matter of sufficient information. A school accountability scheme that rewards school performance will create incentives for pupil selection, even under perfect information, unless the educational production function satisfies an (unrealistic) separability assumption. We propose different compromise solutions and discuss the resulting incentives in theory. The empirical relevance of our analysis – i.e., the rejection of the separability assumption and the magnitude of the incentives in the different compromise solutions – is illustrated with Flemish data.

JEL Classification: H52, I22, I24

Keywords: school accountability, cream-skimming, educational production function

Corresponding author:

Erwin Ooghe
Department of Economics
KU.Leuven
Naamsestraat 69 - bus 3500
3000 Leuven
Belgium
E-mail: erwin.ooghe@kuleuven.be

* We would like to thank Ides Nicaise and Jan Van Damme for their permission to use the SiBO-data, Frederik Maes and Peter Helsen for their valuable help with the data, and Dolors Berga, Geert Dhaene, Carmen Herrero, Iñigo Iturbe-Ormaetxe, Dirk Van de gaer, Frank Vandenbroucke, Carine Van de Voorde, and seminar participants in Alicante, Leuven, Louvain-la-Neuve, Oxford, and Rome for useful comments. The usual disclaimer applies.

# 1 Introduction

Public education used to have some common features around the world. Schools often received funding per pupil and had limited autonomy, inspectors controlled the quality of education, and school choice by parents was often restricted. Critics argued that these features explained the poor performance of (some) public schools.

School accountability increased in several countries to improve student learning. In the U.S., for example, the 'No Child Left Behind Act of 2001' forced all states to set up an accountability system for public schools. In some states schools had to publish report cards—information about their performance based on pupil test scores—to inform parental school choice. Other states used bonuses or sanctions depending on school performance. School accountability improved pupil test scores, but it is unclear whether explicit financial bonuses and sanctions are necessary (Wössmann, 2003; Hanushek and Raymond, 2004, 2005; Jacob, 2005; Figlio and Rouse, 2006; West and Peterson, 2006; Burgess et al., 2007; Chiang, 2009). Accountability can also result in potentially undesirable strategic reactions such as teaching to the rating, student retainment, removal of low-achieving students, and even adapting the caloric content of the school lunches at the testing date (Jacob, 2005; Figlio and Winicki, 2005; Burgess et al., 2005; Reback, 2008). In a nutshell, the overall success of incentive-based reforms crucially depends on their design.

In this paper we focus on another strategic reaction of schools, pupil selection. The average test score in a school strongly depends on the characteristics of the pupil population. Insufficiently correcting for pupil characteristics may lead to a biased evaluation of school performance (Meyer, 1997; Ladd and Walsh, 2002; Hanushek and Raymond, 2003; Taylor and Nguyen, 2006; Neal, 2008). Moreover, it can seduce schools to appear more attractive for specific student groups. Pupil selection may then improve the measured performance of a school without adding real skills.

In this paper we focus on school financing schemes, but the question is equally relevant for the design of report cards.[1] In section 2 we show that it is possible to devise a school financing scheme that rewards higher average test scores without creating incentives for pupil selection, but only if the educational

[1] It could also be relevant for the design of differentiated vouchers (Epple and Romano, 2008).

production function satisfies an (unrealistic) separability assumption. In general a trade-off is inevitable and we therefore propose some compromise solutions. One family of solutions rewards schools for good administration, but does not necessarily eliminate all pupil selection. The well-known value-added scheme is a special case. The other family avoids pupil selection, but does not necessarily reward schools with higher test scores. In section 3 we illustrate the empirical relevance of the trade-off and simulate the incentives provided by the different compromise solutions with empirical data for Flanders (the northern part of Belgium). Section 4 concludes.

# 2 Accountability and incentives

We construct a simple model to show the incompatibility between on the one hand creating incentives for higher test scores and on the other hand avoiding incentives for pupil selection. To bring the key trade-off into focus we start from the most favorable informational assumptions. We assume that sufficient data are available at the pupil level, as it is already well known that informationally less demanding accountability schemes cannot sufficiently correct for differences in pupil characteristics (Meyer, 1997; Hanushek and Raymond, 2003). Furthermore the selection of relevant pupil test scores and its aggregation (over different dimensions and pupils) into a cardinal and comparable indicator of school output is assumed to be settled (Cawley et al., 1999; Neal, 2008). We also neglect the problem that school output measures are typically less reliable for small schools (Kane and Staiger, 2002). Introducing these optimistic assumptions strengthens our impossibility result.

## 2.1 Preliminaries

The agreed measure of school output $y \in \mathbb{R}$ is a function of school variables denoted by $x \in X$; we write $y = f(x)$. School variables consist of administration variables $a \in A$ and background variables $b \in B$; we write $x = (a, b)$, and define the set $X$ as the product $A \times B$. The classification of a school variable as an administration or background variable is simple in theory. Endogenous variables that can be chosen by a school are attributed to administration; for example, the number of instruction hours, the level of remediation per pupil, and teacher motivation. Exogenous variables—variables that cannot be changed

by a school, but whose distribution at school can possibly be altered by selection mechanisms—belong to background; think of initial test scores, innate intelligence, and socioeconomic status of the pupils. Because most background variables are directly related to characteristics of the pupils, we use the term pupil background from now on.

The distinction between administration and background variables is less evident in empirical applications. Usually, the function $f$ will be estimated via a standard explanatory model of test scores; see, e.g., Hanushek (2006) for an overview. A typical estimation includes observable characteristics, unobserved pupil and school effects, and idiosyncratic error terms. Each of these right-hand side variables, observed and unobserved, must be classified as an administration or a background variable. We will make a specific proposal in the next (empirical) section, but for the theoretical analysis it is sufficient to assume that *some* classification is made.

We do not explicitly model school behavior (as, e.g., in Barlevy and Neal, 2011). The output function $f$ is a reduced form equation that reflects the educational production technology. We assume that $f$ does not change under the incentive scheme. Changes in subsidies can of course motivate schools to be more effective, otherwise the whole exercise would be meaningless. Changes may also induce incentives for pupil selection. Both effects are fully captured in our framework by a change in the administration variables in $a$ and pupil background characteristics in $b$.

We use subscripts $j = 1, 2, \ldots, J$ to denote schools. A school subsidy scheme $s : X^J \to \mathbb{R}^J$ maps all information about the different schools $\mathbf{x} = (x_1, x_2, \ldots, x_J)$ into a vector of school subsidies $s(\mathbf{x}) = (s_1(\mathbf{x}), s_2(\mathbf{x}), \ldots, s_J(\mathbf{x}))$. We look for a subsidy scheme that rewards schools for better output without providing incentives to attract or refer pupils with specific characteristics. What form should $s(\mathbf{x})$ take? To answer this question, we do not start from an overall social objective function, but we model the requirements to be imposed on the school financing scheme directly in terms of two basic principles. These principles, and some of the results later on, are inspired by the theory of fair allocation (see, e.g., Fleurbaey, 2008, for an overview) and its application to health insurance (Schokkaert et al., 1998; Schokkaert and Van de Voorde, 2004). We first formulate the basic principles, then show that they are incompatible in general, and finally introduce some possible compromise solutions.

## 2.2 Getting the incentives right

We start with the reward principle. A change in the output of a school that is *only* caused by a change in administration should change the school subsidy in the same direction. Let $(\mathbf{a}, \mathbf{b})$ be the decomposition of $\mathbf{x}$ (with obvious notation).

INCENTIVES FOR GOOD ADMINISTRATION: For all $\mathbf{x}, \mathbf{x}'$ in $X^J$, for all $j = 1, 2, \ldots, J$, if $a_k = a'_k$ for each school $k \neq j$, and $\mathbf{b} = \mathbf{b}'$, then there exists a strictly increasing function $\phi$, with $\phi(0) = 0$, such that $s_j(\mathbf{x}') - s_j(\mathbf{x}) = \phi(y'_j - y_j)$.

The axiom does not say that the subsidy increase should be sufficiently large to make the cost (if any) of the change in administration worthwhile. It simply says that good administration should be financially encouraged. It can be interpreted as a minimalist necessary condition for efficiency. For later use, if the subsidy functions $s_j$ and the output function $f$ are differentiable with respect to some administration variable $a_{jk}$—an element in $a_j = (\ldots, a_{jk}, \ldots)$—then the axiom relates the marginal subsidy to the marginal output, or

$$\partial s_j(\mathbf{x})/\partial a_{jk} = \phi(\partial f(x_j)/\partial a_{jk}), \tag{1}$$

for all profiles and schools, with $\phi$ as defined before.

We now turn to pupil selection. Changes in the background of pupils without changes in administration efficiency should not be rewarded in the funding scheme. Otherwise schools would have an incentive to attract or refer pupils with a specific background.

NO INCENTIVES FOR PUPIL SELECTION: For all $\mathbf{x}, \mathbf{x}'$ in $X^J$, for all $j = 1, 2, \ldots, J$, if $\mathbf{a} = \mathbf{a}'$, and $b_k = b'_k$ for each school $k \neq j$, then $s_j(\mathbf{x}') = s_j(\mathbf{x})$.

The principle clearly wipes out all financial incentives for pupil selection. In general such pupil selection is undesirable, as it may lead to unequal treatment of pupils within schools, to segregation, and to restrictions on the freedom of choice of pupils and parents. Note however that a normative trade-off can arise if the segregation or integration of pupils over schools would increase average school output. Incentives for pupil selection could then be desirable from an efficiency point of view. We will come back to this issue when we discuss compromise solutions that allow for pupil selection. Given differentiability with respect to a pupil background variable $b_{jk}$—an element in $b_j = (\ldots, b_{jk}, \ldots)$—the axiom

imposes no subsidy changes at the margin, or

$$\partial s_j(\mathbf{x})/\partial b_{jk} = 0, \tag{2}$$

for all profiles and schools.

## 2.3 Performance incentives create selection incentives

The two principles seem minimal requirements to be imposed if the aim is to create incentives for good administration and to avoid pupil selection at the same time. It is therefore striking that it is not possible to design a funding scheme that satisfies both principles in general, i.e., for all possible output functions $f$. This impossibility result is well known (in many variants) in the social choice literature (Fleurbaey, 2008). Yet it remained largely unnoticed in the literature on school accountability. Meyer (1997) is the only one, as far as we know, that has drawn attention to the fact that schools cannot be ranked unambiguously according to performance, if the effect of background variables on output differs between them, but he did not integrate this observation in a general theoretical framework.[2]

We provide a simple proof of the incompatibility between the two incentive axioms. We focus on an arbitrary school, keeping information on all other schools constant. We suppress subscripts, and, with a slight abuse of notation, we denote the output and the subsidy of the school by $f(a, b)$ and $s(a, b)$. Let $b \in B = \mathbb{R}$ be an index of pupil background at the school. Figure 1 presents school output as a function of pupil background for two types of administration $a$ and $a'$.

**Figure 1**

Start at situation 1 with administration $a$ and pupil background $b$. An increase in the background index from $b$ to $b'$ leads us to situation 2. The axiom NO INCENTIVES FOR PUPIL SELECTION requires the same subsidy in both situations, thus $s(a, b) = s(a, b')$. If the school would now change administration from $a$ to $a'$, ceteris paribus, then we go from situation 2 to 3 with a lower output. The axiom INCENTIVES FOR GOOD ADMINISTRATION requires a lower subsidy

[2] Moreover, Meyer (1997) claims that the empirical relevance of his observation is limited because "the assumption that slopes do not vary across schools is often a very reasonable assumption." In the next section, we falsify this claim with Flemish data.

leading to $s(a, b') > s(a', b')$. If the school sticks to administration $a'$, but the pupil background index changes back to $b$, then we arrive in situation 4. Again the same subsidy should apply, so $s(a', b') = s(a', b)$. Finally, a change in administration back to $a$, ceteris paribus, lowers output again, and the subsidy must follow, or $s(a', b) > s(a, b)$. All things together we get a cycle. Proposition 1 summarizes this finding.

PROPOSITION 1. There is no subsidy scheme that satisfies INCENTIVES FOR GOOD ADMINISTRATION and NO INCENTIVES FOR PUPIL SELECTION in general, i.e., for each possible output function $f$.

Proposition 1 has to be interpreted carefully: the general impossibility result only holds if we look for a subsidy scheme satisfying both axioms for *all* possible output functions $f$. It is obvious that the incompatibility disappears in Figure 1 if the lines would not intersect. Proposition 2 generalizes this observation (a proof can be found in the appendix).

PROPOSITION 2. A subsidy scheme can satisfy INCENTIVES FOR GOOD ADMINISTRATION and NO INCENTIVES FOR PUPIL SELECTION if and only if there exist functions $g : \mathbb{R} \times B \to \mathbb{R}$ and $h : A \to \mathbb{R}$, with $g$ strictly increasing in its first argument, such that $f(a, b) = g(h(a), b)$, for all $x = (a, b)$ in $X$.

The intuition is easy. The separability condition in Proposition 2 allows to unambiguously classify schools according to the performance index $h(a)$: a higher index corresponds with a higher output, irrespective of the pupils' background. If we define each subsidy $s_j(\mathbf{x})$ to be a strictly increasing function of the performance index $h(a_j)$ only, then both requirements will be satisfied by the resulting subsidy scheme.

It is an empirical question whether this separability condition is satisfied by the data. It is implicitly imposed by the simple linear models that are often used to estimate educational production functions. However, proposition 1 suggests that falsely assuming a linear form may have undesirable consequences in terms of school administration and pupil selection. It is therefore important not to simply assume separability, but to test if it holds. In our empirical illustration we discuss a case where it is strongly rejected.

## 2.4 Compromise solutions

If there is a conflict between the two principles, we have to formulate compromise solutions. We can keep the incentives for good administration intact, but then we may introduce incentives for selecting pupils with a certain background. Or we can make sure that we avoid selection, but then the incentives to improve pupil learning can be very different for different pupils and may even become negative.

For ease of exposition, we suppress in our notation the dependency of the subsidies on the profile $\mathbf{x}$. We use linear subsidy schemes from now on and write the per pupil subsidy for school $j$ as

$$s_j = \texttt{constant} + \underbrace{\texttt{slope}}_{>0} \times \widetilde{y}_j, \tag{3}$$

with $\widetilde{y}_j$ the (possibly corrected) output of school $j$ that will be defined later on. In this section we leave the choice of `constant` and `slope` open. A natural choice would be to choose `constant` so as to satisfy the budget constraint of the regulator, and `slope` so as to guarantee a minimal subsidy to all schools. This is the approach that will be followed in the empirical application, but for the theoretical analysis the choice of `constant` and `slope` is irrelevant.

Before we propose two families of compromise solutions, we discuss two benchmark subsidy schemes: per capita ($PC$) and uncorrected output ($UO$) funding. In many countries school funding is simply per capita, i.e.,

$$s_j^{PC} = \texttt{constant}. \tag{4}$$

A per capita scheme does not provide any incentives, neither for good administration, nor for pupil selection. An uncorrected output scheme fully rewards schools for output increases, without any correction for pupil background. The subsidy is equal to

$$s_j^{UO} = \texttt{constant} + \texttt{slope} \times \underbrace{f(a_j, b_j)}_{y_j}. \tag{5}$$

The scheme gives incentives for good administration, because changes in administration that lead to higher output clearly will be rewarded. Assuming differentiability we obtain

$$\partial s_j^{UO} / \partial a_{jk} = \texttt{slope} \times \partial f(a_j, b_j) / \partial a_{jk}, \tag{6}$$

and, given that $\texttt{slope} > 0$, condition (1) is indeed satisfied. For the same reason however, also changes in background that lead to higher output will be rewarded. Schools have an incentive to attract pupils with a background that is 'favorable' to output. Given differentiability the subsidy change is equal to

$$\partial s_j^{UO}/\partial b_{jk} = \texttt{slope} \times \partial f(a_j, b_j)/\partial b_{jk}, \tag{7}$$

violating condition (2) if $\partial f(a_j, b_j)/\partial b_{jk}$ differs from zero.

A first family of compromise solutions is based on a reference administration ($RA$) level, denoted $\widetilde{a}$, to correct output. Define corrected output as $\widetilde{y_j} = y_j - f(\widetilde{a}, b_j)$; the subsidy is then equal to

$$s_j^{RA} = \texttt{constant} + \texttt{slope} \times ( \underbrace{f(a_j, b_j)}_{y_j} - f(\widetilde{a}, b_j) ). \tag{8}$$

Schools are rewarded if their output is higher than the hypothetical output that would result if the school had chosen the reference administration level, ceteris paribus. The scheme creates incentives for good administration, because changes in administration that are favorable to output translate into higher subsidies. Assuming differentiability of the reference scheme, the incentive for good administration $\partial s_j^{RA}/\partial a_{jk}$ is exactly equal to the one for uncorrected output in equation (6). Reference administration schemes may lead to selection incentives, however. The selection incentive depends on

$$\partial s_j^{RA}/\partial b_{jk} = \texttt{slope} \times (\partial f(a_j, b_j)/\partial b_{jk} - \partial f(\widetilde{a}, b_j)/\partial b_{jk}), \tag{9}$$

and will typically be different from zero, thus violating (2). But, comparing (7) and (9), if the derivatives $\partial f(a_j, b_j)/\partial b_{jk}$ and $\partial f(\widetilde{a}, b_j)/\partial b_{jk}$ are similar in magnitude, then $|\partial s_j^{RA}/\partial b_{jk}|$ will be smaller than $|\partial s_j^{UO}/\partial b_{jk}|$. Summing up, reference administration schemes provide similar incentives for good administration compared to uncorrected output schemes, but can be expected to provide lower incentives for pupil selection.

The mirror image of the previous scheme is to choose a reference pupil background ($RB$), say $\widetilde{b}$. If we define corrected output as $\widetilde{y_j} = f(a_j, \widetilde{b})$, then the school will be rewarded on the basis of the hypothetical output that would arise if its actual administration were applied to the reference pupil population. This yields

$$s_j^{RB} = \texttt{constant} + \texttt{slope} \times f(a_j, \widetilde{b}). \tag{10}$$

The subsidy $s_j^{RB}$ does not depend on the school background $b_j$ anymore, which removes selection incentives. With differentiability, we indeed get $\partial s_j^{RB}/\partial b_{jk} = 0$ as required by (2). But actual output does not appear in equation (10) either. We can immediately derive that

$$\partial s_j^{RB}/\partial a_{jk} = \texttt{slope} \times \partial f(a_j, \widetilde{b})/\partial a_{jk}, \tag{11}$$

and condition (1) is no longer satisfied if the change in the true output, being $\partial f(a_j, b_j)/\partial a_{jk}$, has a different sign compared to the change in the hypothetical output $\partial f(a_j, \widetilde{b})/\partial a_{jk}$. Because we expect that the signs of both derivatives often coincide, the reference background scheme will provide more incentives for good administration compared to a per capita scheme. Summing up, reference background schemes provide no incentives for pupil selection, as is the case in a per capita scheme, but in addition they can be expected to provide some incentives for good administration.

Table 1 summarizes the different schemes and their properties, i.e., is the axiom satisfied, how large do we expect the incentives to be, and for how many schools will the axioms be satisfied? Per capita schemes do not give incentives for good administration nor incentives for pupil selection to *any* school. Uncorrected output schemes give both incentives to *all* schools. We expect the reference schemes to do better. More precisely, reference administration schemes outperform the uncorrected output schemes, because they give the same incentives for good administration to all schools, but with a lower incentive for pupil selection. Reference background schemes outperform per capita schemes, because they provide no incentives for pupil selection, but incentives for good administration to some schools.

**Table 1**

While we expect reference administration and reference background schemes to be better than the simple benchmark schemes, they also require more information. More importantly, the extra information required by the schemes in equation (8) and (10) is also different. To implement these schemes, the regulator must have (an estimate of) the educational production function $f$. In addition, a reference administration scheme requires information about output $y_j$ and background variables $b_j$, while a reference background scheme needs information about the administration variables $a_j$. These different requirements

may have practical consequences. A reference background scheme offers scope for strategic behavior, e.g., increasing instruction time without any real results. Even worse, it may create incentives for misreporting variables, like instruction time, that are difficult to verify. Strategic behavior is less problematic in a reference administration scheme. Test scores are collected in a standardized way and background variables typically consist of pupil characteristics that can (more) easily be controlled by the regulator.

A final note. The reference administration and reference background schemes are "families" of solutions, since we obtain another scheme for each specific choice of reference. The implications that have been described in this section hold for the complete family, but this does not mean that the choice of reference values is irrelevant. We will return to this issue in our empirical illustration.

# 3 Empirical illustration

For our empirical application we used the data collected in Flanders (the northern part of Belgium) through the 'SiBO'-project. The aim of the project is to describe and explain differences in the primary school curriculum of Flemish pupils. Pupils were tested in mathematics at the start of the first grade (in September-October 2003 when (most) pupils were 6 years old) and at the end of grades 1 and 2 (in May-June of 2004 and 2005).[3] Other pupil data include the gender of the pupil, the language they speak with each of the parents, and the education level of the parents.[4] Classroom data consist of the total experience of the teacher, the class size, the instruction time for mathematics, and the number of teachers in a class. We also include the average initial test score of the peers, defined as all pupils in the same class. Table 2 provides a description of all variables.

**Table 2**

We restrict ourselves to schools with at least 10 pupils tested in each grade. We have 5817 pupil-time observations—2239 pupils appearing in both grades, 628 in grade 1 only, and 711 in grade 2 only—distributed over 111 schools.

[3] The math tests consists of between 40 and 80 questions (depending on the grade). The score distributions are well-behaved, showing no floor and only limited ceiling effects.

[4] Note that Dutch is the official language in Flanders.

The main reason for attrition and replenishment is student retainment. We come back to this potential source of selection bias. Tables 3a and 3b contain summary statistics for the pupil and classroom data.

**Table 3a and 3b**

## 3.1 Explaining test scores: a linear exploration of the data

Let $y_{ijt}$ be the (standardized) math test score of pupil $i$ at school $j$ at time $t$ and let $z_{ijt}$ be the vector of observable regressors. To explore the data, we start with a standard linear panel model, i.e.,

$$y_{ijt} = \beta' z_{ijt} + u_i + v_j + w_{ijt}, \tag{12}$$

with $u_i$ a 'random' pupil-level effect, $v_j$ a 'fixed' school level effect, and $w_{ijt}$ an idiosyncratic error term. The specification (12) satisfies the separability condition in (2), irrespective of how the right-hand variables are classified into administration or background variables. We come back to the classification later on.

Because of attrition and replenishment in the data, we must check and eventually also correct estimates for selection bias. To check for selection bias we use a variable addition test; see, Verbeek and Nijman (1992) and Wooldridge (1995). The results indicate that missingness might be informative, especially for attrition. To check whether a selection correction influences the estimation results, we add a selection equation to each period in the spirit of Hausman and Wise (1979); we allow for correlation between the individual level effects in the selection and the output equations. The corrected estimates do not statistically differ from the uncorrected estimates, allowing us to ignore selection issues in the sequel.[5]

**Table 4**

Table 4 reports estimates of the parameters in equation (12).[6] The initial test score plays an important role in all models. Its coefficient is rather robust

[5] The selection correction model assumed random (rather than fixed) school effects, leading to a so-called multilevel model. An attempt with fixed school effects did not converge, probably due to the high number of dummies in the selection equation.

[6] We add a dummy 'missing' to each covariate group (to limit the reduction in total sample size). We do not report the corresponding estimates which are, as expected, never significant.

and smaller than 1, indicating that the gain in test scores, is larger for pupils with a lower initial test score. The background variables play a more modest role and their effects depend on whether or not the initial test score is taken up as a covariate. In model (b) without initial test scores, boys do better than girls; being ahead of age is not significant while lagging behind is correlated with a lower math performance; having Dutch-speaking and more educated parents improve test scores and these effects are stronger and more significant for mothers compared to fathers. In model (c) with initial test scores as an additional regressor, some of the estimated coefficients for the background variables change in magnitude and even in sign. We provide two striking examples.

First, speaking Dutch with your parents has a negative coefficient once we correct for initial test scores. Indeed, pupils that do not speak Dutch at home have a worse preparation on average before starting primary education. Therefore their initial test score underestimates their potential, leading to a catching-up effect in the first grades. Second, the effect of fathers' education level is now stronger than that of mothers. This is consistent with the hypothesis that mothers have a larger effect on initial test scores during the pre-primary education period, while fathers have a larger effect on the primary education growth of their children.

Comparing models (c) and (d) shows that adding class level variables has a minor effect on the estimated coefficients for the background variables. Among the class variables (introduced in column (d)), instruction time and class size have a significant and positive effect, while having two teachers reduces test scores.

### 3.2 Testing separability

A linear specification is common in the literature on educational production functions. For the derivation of a financing scheme however, it is essential to test explicitly whether the separability assumption in proposition 2 holds. To do so, we split the observables $z_{ijt}$ into administration and background variables, denoted $z_{a,ijt}$ and $z_{b,ijt}$. It is natural to assign variables at the class and school level—including the school-specific constant, but excluding the peer variable—to administration. All other variables—the pupil-level variables, the time dummy, and the peer variable—are classified as background. To test separability we generalize (12), allowing the pupil background coefficients to vary over schools,

i.e.,

$$y_{ijt} = \beta_a' z_{a,ijt} + \beta_{b,j}' z_{b,ijt} + u_i + v_j + w_{ijt}, \tag{13}$$

For the purpose of illustration, we define school output as the expected average pupil output. If a bar denotes an average (with the subscript indicating at which level the average is taken) and a hat indicates an estimate, then school output is equal to

$$\overline{y}_j = \widehat{\beta}_a' \overline{z}_{a,j} + \widehat{\beta}_{b,j}' \overline{z}_{b,j} + \widehat{v}_j, \tag{14}$$

where $\overline{y}_j$ is the average test score in school $j$. Since the slope coefficients $\widehat{\beta}_{b,j}$ tell us how pupils with a certain background perform at school $j$, it is natural to assign these coefficients to administration.[7] We then get:

$$\overline{y}_j \quad = \quad \underbrace{\widehat{\beta}_a' \overline{z}_{a,j} \; + \; \widehat{v}_j}_{\text{pure administration}} \quad + \quad \underbrace{\widehat{\beta}_{b,j}' \overline{z}_{b,j}}_{\text{mixture}} .$$

The non-linear terms in $\widehat{\beta}_{b,j}' \overline{z}_{b,j}$ mix administration and background. They are crucial to test the separability condition in proposition 2. More precisely, separability is satisfied if the slope coefficients $\widehat{\beta}_{b,j}$ are the same for all schools. Table 5 summarizes the separability tests based on model (13).

**Table 5**

The 'equal slope'-hypothesis is statistically rejected, for each background variable separately as well as for all background variables jointly. The theoretical consequences of this non-separability have been described before. Incentives for good administration may create incentives for pupil selection and, vice-versa, removing incentives for pupil selection may create incentives for bad administration. We now turn to the empirical relevance of this incompatibility. In particular, we will check to what extent the different subsidy schemes defined above satisfy the principles of INCENTIVES FOR GOOD ADMINISTRATION and NO INCENTIVES FOR PUPIL SELECTION.

[7] To avoid confusion, we stress that the subscript $b$ in the estimated slope vector $\widehat{\beta}_{b,j}$ indicates that it is a slope vector for the background variables. Still, these background slopes are at the school level and therefore classified as administration variables.

### 3.3 The trade-off in practice

Recall the linear subsidy scheme defined in equation (3). Let us first operationalize its different interpretations within the context of the non-separable model (13). For this empirical application we have to specify the constant and the slope. The `constant` can be fixed by introducing the budget constraint faced by the regulator, i.e., by imposing that the average subsidy per pupil has to be equal to the available budget per pupil. If we normalize the available budget to be 1 unit per pupil, the per-pupil subsidy at school $j$ becomes

$$s_j = \texttt{1} + \texttt{slope} \times ( \widetilde{y}_j - \overline{\widetilde{y}} ),$$

with $\overline{\widetilde{y}}$ the average (corrected) output.[8]

To fix the `slope`, a natural constraint is to guarantee each school a minimal subsidy per pupil, say $\underline{s}$, with $0 < \underline{s} < 1$. A minimal subsidy requirement imposes an upper bound on the `slope`:

$$\texttt{slope} \leq ( 1 - \underline{s} ) / ( \overline{\widetilde{y}} - \min \widetilde{y}_j ). \tag{15}$$

With $\underline{s}$ fixed, equation (15) will yield a different upper bound for `slope`, depending on the subsidy scheme used. To ease the comparison of the results for the different schemes, we choose `slope` to be equal to the lowest upper bound over the different schemes. Taking $\underline{s} = 0.5$, we get a coefficient `slope` ($\approx 0.4$) that is common to all schemes, but the minimal subsidy per pupil will differ slightly between the different schemes.

We can now provide a formula for each subsidy scheme. Denoting as before the reference levels by a tilde, we get (the derivations can be found in the appendix):

$$\begin{aligned}
s_j^{PC} &= 1, \\
s_j^{UO} &= 1 + \texttt{slope} \times (\overline{y}_j - \overline{y}), \\
s_j^{RA} &= 1 + \texttt{slope} \times \{(\overline{y}_j - \overline{y}) - \widetilde{\beta}_b'(\overline{z}_{b,j} - \overline{z}_b)\}, \\
s_j^{RB} &= 1 + \texttt{slope} \times \{(\overline{y}_j - \overline{y}) - \widetilde{\beta}_{b,j}'(\overline{z}_{b,j} - \widetilde{z}_b) + \overline{\widehat{\beta}_{b,j}'(\overline{z}_{b,j} - \widetilde{z}_b)}\}.
\end{aligned}$$

As noted before, the reference levels in reference administration and reference background models can be chosen. These choices will have different implications

[8] Remember that if we define an average without a subscript, this is the average over the whole population, i.e. over all schools.

for incentives. In the empirical illustration, the reference levels for slopes $\widetilde{\beta}_b$ and averages $\widetilde{z}_b$ are based on the distribution of the estimated coefficients $\widehat{\beta}_{b,j}$ and the averages $\overline{z}_{b,j}$ over the different schools. We choose the 5th percentile (low), the median (mid), and the 95th percentile (high).

It is instructive to compare these solutions with what could be described as the common practice of using a value added ($VA$) model. If one sticks to the (rejected) separable model

$$y_{ijt} = \beta_a^{VA\prime} z_{a,ijt} + \beta_b^{VA\prime} z_{b,ijt} + u_i^{VA} + v_j^{VA} + w_{ijt}^{VA}, \tag{16}$$

then school output is equal to

$$\overline{y}_j = \widehat{\beta}_a^{VA\prime} \overline{z}_{a,j} + \widehat{\beta}_b^{VA\prime} \overline{z}_{b,j} + \widehat{v}_j^{VA}.$$

The part $\widehat{\beta}_a^{VA\prime} \overline{z}_{a,j} + \widehat{v}_j^{VA}$ is usually considered to be the value-added of the school; see, e.g., Meyer (1997). If we define corrected output $\widetilde{y}_j$ as value added, the per pupil subsidy reduces to (see appendix)

$$s_j^{VA} = 1 + \texttt{slope} \times \{(\overline{y}_j - \overline{y}) - \widehat{\beta}_b^{VA\prime} (\overline{z}_{b,j} - \overline{z}_b)\}.$$

Comparing this value-added scheme with the reference administration subsidy scheme immediately shows that the former is a special case of the latter in which the reference coefficients $\widetilde{\beta}_b$ are chosen to be the coefficients $\widehat{\beta}_b^{VA}$, estimated with the wrongly specified model (16). Since $\widehat{\beta}_b^{VA}$ will not be very different from the median value for $\widehat{\beta}_{b,j}$, the value-added scheme will have similar properties as the corresponding (median) reference administration scheme. It will create incentives for efficiency, but also incentives for pupil selection.

To check whether these schemes satisfy our two basic principles, we consider two simulations, each for two background variables. Since it follows from the results in Table 4 that especially initial test scores and parental education strongly correlate with final test scores, we select these two variables for further analysis. The distribution of their slopes and averages over schools is given in Table 6.[9] The simulation results are given in Tables 7 and 8 respectively.

**Tables 6, 7, and 8**

[9] The reported slopes for average initial test scores include the peer effect. For parental education we include both the education of the mother and of the father.

A first simulation focuses on the subsidy change that might result from changing the administration of a school even if school output does not change. This can be realized by increasing (or decreasing) the slope of the background variable $\widehat{\beta}_{b,jk}$ by one standard deviation and simultaneously decreasing (increasing) the school fixed effect $\widehat{v}_j$ such that $\Delta\overline{y}_j = \Delta\widehat{\beta}_{b,jk}\overline{z}_{b,jk} + \Delta\widehat{v}_j = 0$. One could interpret this as reflecting a school policy that shifts teaching effort between different group of pupils, without changing school output. Since the output effect is zero by construction, the first principle (INCENTIVES FOR GOOD ADMINISTRATION) obviously requires that this policy should not lead to an increase in the school subsidy. Table 7 contains the effects of the simulated change on the school subsidies per pupil for the different subsidy schemes and for both background variables. To interpret the numbers, recall that the subsidies are normalized to be 1 on average. So, -0.08 or 0.06 can be interpreted as a loss or a gain equal to 8% or 6% of the average school subsidy. We do not report the results for the per-capita scheme, because per-capita subsidies obviously do not respond to the simulated changes. Notice that all subsidy changes in the tables should be interpreted as short-term changes, i.e., assuming that other schools do not change policy.

Table 7 shows that the $RA$ schemes—including the value added scheme—and the uncorrected ouput scheme satisfy the reward principle. If output does not change, the subsidy does not change. However, this is not true for the reference background ($RB$) schemes, for which the value of $\Delta s_j$ depends on the choice of reference values and can be negative or positive. If one chooses a low value for the reference $\widetilde{z}$, the right-hand panel of Table 7 shows that 96% (resp. 100%) of the schools have an incentive to distort their effort choices towards those with low initial test scores (parental education). If one chooses a high value for the reference $\widetilde{z}$, about 95% (100%) of the schools have an incentive to increase the slope coefficient for initial test scores (parental education).

The problem of the $RB$ schemes is in fact more severe. Recall that the regulator can only apply these $RB$ schemes, if she gets the necessary information about the administration variables from the schools themselves. This information is difficult to verify and easy to manipulate. It is therefore clear that there is a real danger that the $RB$ schemes are manipulated by the schools to receive a higher subsidy without a better performance.

A second simulation looks at the subsidy change resulting from a change in

the pupil distribution of a school. We implement this hypothetical change by simulating the effects of increasing or decreasing $\overline{z}_{b,jk}$ by one standard deviation. Ideally, NO INCENTIVES FOR PUPIL SELECTION requires that the subsidy should not change, since schools should not be rewarded for pupil selection if there is no increase in administration efficiency. The results for the different subsidy schemes are given in Table 8. With the reference background schemes, schools are not rewarded when changing their pupil composition (without changing their administration efficiency). However, the reference administration schemes and the uncorrected output scheme provide incentives for pupil selection. In case of uncorrected output for example, increasing the average initial test scores at school by one standard deviation may increase the subsidy up to 18% of the average school subsidy. The gains and losses in case of the reference administration schemes are typically smaller and depend on the reference slope. If the $RA$ scheme is implemented with a low (high) reference value for the pupil background, almost all schools gain by attracting pupils with higher (lower) initial test scores or better (worse) parental education.

The tables suggest that choosing a median reference level minimizes the absolute magnitude of the selection incentives. Since this is very close to the value added ($VA$) model, the latter also performs satisfactorily in this respect. Moreover, the reference level will play a role for behavior and can now be used to steer selection incentives. A low reference implies that most schools benefit from attracting stronger pupils, while a high reference implies that most schools gain from attracting weaker pupils. An intermediate level—as in the value added model—imply that some schools gain and other schools lose from attracting better students. Interestingly, this may lead to efficient sorting. Schools with a higher slope than the reference slope perform better for stronger pupils and also have an incentive to attract the stronger pupils and refer the weaker ones; and schools with a lower slope than the reference slope perform better for weaker pupils and also get an incentive to attract them and to refer better pupils. Of course, stronger segregation of pupils will result. The simulation results clearly illustrate the trade-off between different objectives that was already mentioned when we introduced the principle of NO INCENTIVES FOR PUPIL SELECTION.

# 4 Conclusion

Recent experiences have shown that introducing school accountability may create incentives for efficiency. It may also have undesirable side-effects, however, like pupil selection, even if test scores can be perfectly corrected for pupil characteristics. We have shown that a school financing scheme that rewards output also creates incentives for cream-skimming, unless the educational production function satisfies an (unrealistic) separability assumption. It is therefore necessary to consider explicitly the trade-off between the two objectives of improving performance and avoiding selection. We discuss the pros and cons of different compromise solutions and we have shown how information from the empirical educational production literature can be integrated in a coherent normative framework.

The empirical relevance of this analysis is illustrated with data on Flemish primary schools, for which the separability assumption is strongly rejected. Given the manipulability of schemes that rely on information about the policy decisions of schools, one could argue in favor of what we have called "reference administration" schemes. To implement such schemes the regulator only needs information on the characteristics of the pupils. Our empirical results show the importance of choosing a "correct" reference value for the administration variable. Choosing a low (high) reference value for the slope variable creates incentives for better treatment and selection of pupils with "better" ("worse") characteristics in terms of initial test scores and parental education. Picking an intermediate value (as is implicitly done by the value added model) will minimize the selection incentives and may even induce a kind of efficient sorting.

We have interpreted our axioms and results in terms of a funding scheme. This is not the only possible interpretation, however. One could as well argue that $s(\mathbf{x})$ represents only a performance measure, rather than a subsidy. Our principles remain valid in this measurement interpretation—INCENTIVES FOR GOOD ADMINISTRATION could be rebaptized as "performance sensitivity" and NO INCENTIVES FOR PUPIL SELECTION as "correction for pupil characteristics"—and the impossibility result remains relevant in this setting. Even if the regulator is not willing to introduce accountability in the system (which may be the case in many European countries) and sticks to the idea of quality norms and control, the framework remains valid. It is natural to financially compensate

schools with a socially disadvantaged pupil population, as it is more difficult for them to realize the required quality norms. In general however, it is not possible to compensate schools for their pupil population, while leaving the school autonomy to meet the quality criteria unaffected. The current framework can therefore shed a light on this question as well.

# References


[1] Barlevy, G., and D., Neal, 2012, Pay for percentile, *American Economic Review* 102(5), 1805-1831.

[2] Burgess, S., C., Propper, H., Slater, and D., Wilson, 2005, Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools. CMPO, Bristol: Working Paper 05/128.

[3] Burgess, S., C., Propper, and D., Wilson, 2007, The impact of school choice in England. *Policy Studies* 28(2), 129-143.

[4] Cawley, J., J., Heckman, and E., Vytlacil, 1999, On policies to reward the value added by educators, *Review of Economics and Statistics* 81(4), 720-727.

[5] Chiang, H., 2009, How accountability pressure on failing schools affects student achievement, *Journal of Public Economics* 93, 1045-1057.

[6] Epple, D., and R., Romano, 2008, Educational vouchers and cream skimming, *International Economic Review* 49(4), 1395-1435.

[7] Figlio, D., and C., Rouse, 2006, Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics* 90, 239-255.

[8] Figlio, D., and J., Winicki, 2005, Food for thought: the effects of school accountability plans on school nutrition, *Journal of Public Economics* 89, 381-394.

[9] Fleurbaey, M., 2008, *Fairness, Responsibility and Welfare*, Oxford University Press.

[10] Hanushek, E., 2006, School resources, chapter 14 in, E., Hanushek, and F., Welch, eds., *Handbook of the Economics of Education* vol 2, Elsevier.

[11] Hanushek, E., and M., Raymond, 2003, Lessons about the design of state accountability systems, in, P., Peterson, and M., West, eds., *No Child Left Behind? The Politics and Practice of Accountability*, Brookings.

[12] Hanushek, E. and M., Raymond, 2004, The effect of school accountability systems on the level and distribution of student achievement, *Journal of the European Economic Association* 2(2/3), 406-415.

[13] Hanushek, E. , and M., Raymond, 2005, Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management* 24(2), 297-327.

[14] Hausman, J, and D., Wise, 1979, Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica* 47(2), 455-473.

[15] Jacob, B., 2005, Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools, *Journal of Public Economics* 89, 761-796.

[16] Kane, T., and D., Staiger, 2002, The promise and pitfalls of using imprecise school accountability measures, *Journal of Economic Perspectives* 16(4), 91-114.

[17] Ladd, H., and R., Walsh, 2002, Implementing value-added measures of school effectiveness: getting the incentives right, *Economics of Education Review* 21, 1-17.

[18] Meyer, R., 1997, Value-added indicators of school performance: a primer, *Economics of Education Review* 16(3), 283-301.

[19] Neal, D., 2008, Designing incentive systems for schools, in, M., Springer, ed., *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings.

[20] Reback, R., 2008, Teaching to the rating: school accountability and the distribution of student achievement, *Journal of Public Economics* 92, 1394-1415.

[21] Schokkaert, E., G., Dhaene, and C., Van de Voorde, 1998, Risk adjustment and the trade-off between efficiency and risk selection: an application of the theory of fair compensation. *Health Economics* 7, 465-480.

[22] Schokkaert, E., and C., Van de Voorde, 2004, Risk selection and the specification of the conventional risk adjustment formula, *Journal of Health Economics* 23, 1237-1259.

[23] Taylor, J., and A., Ngoc Nguyen, 2006, An analysis of the value added by secondary schools in England: is the value added indicator of any value? *Oxford Bulletin of Economics and Statistics* 68(2), 203-224.

[24] Verbeek, M., and T., Nijman, 1992, Testing for selectivity bias in panel data models, *International Economic Review* 33, 681-703.

[25] West, M., and P., Peterson, 2006, The efficacy of choice threats within school accountability systems: results from legislatively induced experiments, *Economic Journal* 116, 46-62.

[26] Wooldridge, J., 1995, Selection corrections for panel data models under conditional mean independence assumptions, *Journal of Econometrics* 68, 115-132.

[27] Wössmann, L., 2003, Schooling resources, educational institutions and student performance: the international evidence, *Oxford Bulletin of Economics and Statistics* 65(2), 117-170.

## Proof of proposition 2

A subsidy scheme can satisfy INCENTIVES FOR GOOD ADMINISTRATION and NO INCENTIVES FOR PUPIL SELECTION if and only if there exist functions $g : \mathbb{R} \times B \to \mathbb{R}$ and $h : A \to \mathbb{R}$, with $g$ strictly increasing in its first argument, such that $f(a, b) = g(h(a), b)$, for all $x = (a, b)$ in $X$.

If the separability condition holds, it is possible to define a subsidy scheme $s$ such that each school subsidy $s_j$ is a strictly increasing function of $h(a_j)$ *only*. Such a scheme satisfies both axioms. We show the opposite.

Consider a subsidy scheme that satisfies INCENTIVES FOR GOOD ADMINISTRATION and NO INCENTIVES FOR PUPIL SELECTION. We show that, for arbitrary administrations $a, a' \in A$ and backgrounds $b, b' \in B$, we have

$$f(a, b) \geq f(a', b) \Leftrightarrow f(a, b') \geq f(a', b'). \tag{17}$$

This would indeed allow to properly define functions

1. $h : A \to \mathbb{R}$ with $h(a) \geq h(a')$ if $f(a, b) \geq f(a', b)$ for some $b \in B$, and
2. $g : \mathbb{R} \times B \to \mathbb{R}$ with $g(h(a), b) = f(a, b)$ for all $x = (a, b)$,

and $g$ will be strictly increasing in its first argument.

We proceed by contradiction. Suppose equation (17) does not hold, e.g., both $f(a, b) \geq f(a', b)$ and $f(a, b') < f(a', b')$ are true for some $a, a' \in A$ and $b, b' \in B$. (It is easy to verify the other direction using the same logic.) We can use these $a, a' \in A$ and $b, b' \in B$ to construct four states—$(a, b)$, $(a', b)$, $(a, b')$, and $(a', b')$—for some school (tacitly assuming that school information remains constant for all other schools). We suppress subscripts and use $f(a, b)$ and (with slight abuse of notation) $s(a, b)$ to refer to the output and the subsidy of the school under consideration. Applying INCENTIVES FOR GOOD ADMINISTRATION twice, we must have

$$s(a, b) - s(a', b) \geq 0 \quad \text{and} \quad s(a, b') - s(a', b') < 0. \tag{18}$$

Applying NO INCENTIVES FOR PUPIL SELECTION twice, we obtain

$$s(a, b) = s(a, b') \quad \text{and} \quad s(a', b) = s(a', b'),$$

and, subtracting both equations, we get:

$$s(a, b) - s(a', b) = s(a, b') - s(a', b'). \tag{19}$$

Equation (18) and (19) are incompatible, a contradiction.

# A derivation of the empirical subsidy schemes

The per-capita and uncorrected output schemes are straightforward. We discuss the reference administration, reference background and value added scheme. A subsidy scheme is defined as

$$s_j = \texttt{1} + \texttt{slope} \times (\widetilde{y}_j - \overline{\widetilde{y}}),$$

with the slope defined by (15) for each scheme. We focus here on the difference $\widetilde{y}_j - \overline{\widetilde{y}}$.

We start from the empirical model

$$\begin{aligned} \overline{y}_j &= \widehat{\beta}'_a \overline{z}_{a,j} + \widehat{v}_j + \widehat{\beta}'_{b,j} \overline{z}_{b,j} &&= f(\underbrace{\overline{z}_{a,j}\ ,\ \widehat{v}_j\ ,\ \widehat{\beta}'_{b,j}}_{a_j}\ ,\ \underbrace{\overline{z}_{b,j}}_{b_j}), \\ &&&= f(a_j\ ,\ b_j). \end{aligned}$$

The RA models use a reference administration, say $\widetilde{a} = (\widetilde{z}_a, \widetilde{v}, \widetilde{\beta}_b)$, to define the hypothetical output as

$$\widetilde{y}_j = \overline{y}_j - f(\widetilde{a}, b_j) = \overline{y}_j - (\beta'_a \widetilde{z}_a + \widetilde{v} + \widetilde{\beta}'_b \overline{z}_{b,j}).$$

The average hypothetical output is equal to

$$\overline{\widetilde{y}} = \overline{y} - (\beta'_a \widetilde{z}_a + \widetilde{v} + \widetilde{\beta}'_b \overline{z}_b),$$

and the difference $\widetilde{y}_j - \overline{\widetilde{y}}$ is indeed equal to

$$(\overline{y}_j - \overline{y}) - \widetilde{\beta}'_b (\overline{z}_{b,j} - \overline{z}_b).$$

Starting from the same empirical model, the RB models replace $\overline{z}_{b,j}$ by a reference background $\widetilde{b} = \widetilde{z}_b$ to get

$$\widetilde{y}_j = f(a_j, \widetilde{b}) = \widehat{\beta}'_a \overline{z}_{a,j} + \widehat{v}_j + \widehat{\beta}'_{b,j} \widetilde{z}_b.$$

The OLS estimate for $\widehat{v}_j$ is

$$\widehat{v}_j = \overline{y}_j - \widehat{\beta}'_a \overline{z}_{a,j} - \widehat{\beta}'_{b,j} \overline{z}_{b,j},$$

and we can rewrite the hypothetical output as

$$\widetilde{y}_j = \overline{y}_j - \widehat{\beta}'_{b,j} (\overline{z}_{b,j} - \widetilde{z}_b).$$

The average is given by

$$\overline{\widetilde{y}}_j = \overline{y} - \overline{\widehat{\beta}'_{b,j} (\overline{z}_{b,j} - \widetilde{z}_b)},$$

and the difference $\widetilde{y}_j - \overline{\widetilde{y}}_j$ indeed becomes

$$(\overline{y}_j - \overline{y}) - \widehat{\beta}'_{b,j}(\overline{z}_{b,j} - \widetilde{z}_b) + \overline{\widehat{\beta}'_{b,j}(\overline{z}_{b,j} - \widetilde{z}_b)}.$$

Finally, for the value-added (VA) model we have

$$\widetilde{y}_j = \widehat{\beta}_a^{VA\prime}\overline{z}_{a,j} + \widehat{v}_j^{VA},$$

with the OLS estimate of $v_j^{VA}$ in (16) given by

$$\widehat{v}_j^{VA} = \overline{y}_j - \widehat{\beta}_a^{VA\prime}\overline{z}_{a,j} - \widehat{\beta}_b^{VA\prime}\overline{z}_{b,j}.$$

Plugging in the OLS estimate, corrected output becomes

$$\widetilde{y}_j = \overline{y}_j - \widehat{\beta}_b^{VA\prime}\overline{z}_{b,j}.$$

Averaging the corrected output, we get

$$\overline{\widetilde{y}} = \overline{y} - \widehat{\beta}_b^{VA\prime}\overline{z}_b,$$

and the difference $\widetilde{y}_j - \overline{\widetilde{y}}$ indeed reduces to

$$(\overline{y}_j - \overline{y}) - \widehat{\beta}_b^{VA\prime}(\overline{z}_{b,j} - \overline{z}_b).$$

# Figures and tables

**Figure 1.** Aligning performance and selection incentives: mission impossible

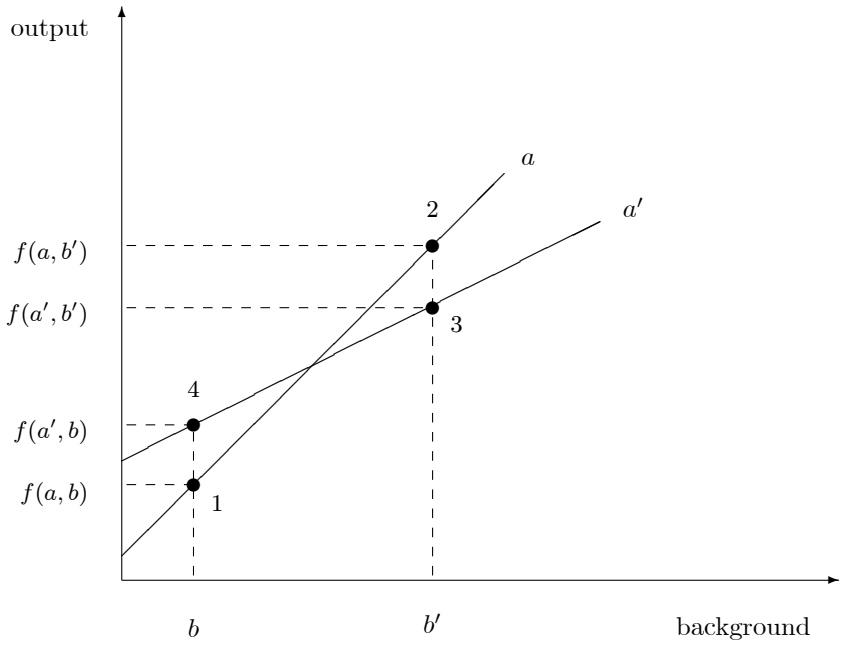

**Table 1.** Different schemes provide different incentives (in theory)

| incentive | for good administration | | | for pupil selection | | |
|---|---|---|---|---|---|---|
| | always | magnitude | # schools | never | magnitude | # schools |
| $PC$ | | 0 | 0 | $\surd$ | 0 | 0 |
| $RB$ | | low | some | $\surd$ | 0 | 0 |
| $RA$ | $\surd$ | high | all | | low | some |
| $UO$ | $\surd$ | high | all | | high | all |

**Table 2.** Abbreviation and description of the variables

| | | |
|---|---|---|
| math | | test score result in mathematics at the end of grade 1 and 2 |
| time2 | = 1, | if pupil is in second grade, 0 otherwise |
| $math_0$ | | initial test score result in mathematics at the start of grade 1 |
| girl | = 1, | if girl, 0 otherwise |
| m_dutch/f_dutch | = 1, | if mother/father speaks Dutch with pupil, 0 otherwise |
| m_edu_sec/f_edu_sec | = 1, | if mother/father has a secondary education degree, 0 otherwise |
| m_edu_high/f_edu_high | = 1, | if mother/father has a tertiary (short type) education degree, 0 otherwise |
| m_edu_uni/f_edu_uni | = 1, | if mother/father has a tertiary (long type) education degree, 0 otherwise |
| duo | = 1, | if there are two teachers, 0 otherwise |
| peer | | average initial test score of peers, i.e., fellow pupils in the class |
| time_math | | mathematics instruction in the classroom in hours per week |
| experience | | (average) teaching experience of teacher(s) in years |
| class_size | | number of pupils in the classroom |

**Table 3a.** Summary statistics for pupil variables.

| math score | mean | std.dev. | p10 | median | p90 |
|---|---|---|---|---|---|
| grade 1 | 8.75 | 1.00 | 7.44 | 8.77 | 10.06 |
| grade 2 | 9.71 | 1.00 | 8.43 | 9.71 | 11.04 |
| initial math score | mean | std.dev. | p10 | median | p90 |
| grade 1 | 8.05 | 1.02 | 6.71 | 8.13 | 9.31 |
| grade 2 | 8.16 | 0.97 | 6.85 | 8.22 | 9.37 |
| sex | = boy | = girl | | | |
| grade 1 | 50.54% | 49.46% | | | |
| grade 2 | 50.58% | 49.15% | | | |
| language mother | = dutch | $\neq$ dutch | miss. | | |
| grade 1 | 86.82% | 8.61% | 4.57% | | |
| grade 2 | 85.69% | 9.12% | 5.19% | | |
| language father | = dutch | $\neq$ dutch | miss. | | |
| grade 1 | 84.80% | 10.29% | 4.91% | | |
| grade 2 | 84.54% | 9.90% | 5.56% | | |
| mother's highest degree | $< 2^{\text{ary}}$ | $2^{\text{ary}}$ | $3^{\text{ary}}$ ($\neq$univ.) | $3^{\text{ary}}$ (=univ.) | miss. |
| grade 1 | 19.25% | 33.94% | 29.06% | 8.72% | 9.03% |
| grade 2 | 16.54% | 33.56% | 30.41% | 10.00% | 9.49% |
| father's highest degree | $<2^{\text{ary}}$ | $2^{\text{ary}}$ | $3^{\text{ary}}$ ($\neq$univ.) | $3^{\text{ary}}$ (=univ.) | miss. |
| grade 1 | 19.29% | 34.74% | 21.00% | 12.07% | 12.90% |
| grade 2 | 17.49% | 34.78% | 22.10% | 13.39% | 12.24% |

**Table 3b.** Summary statistics for class variables

| # of teachers | = 1 | = 2 | | | |
|---|---|---|---|---|---|
| grade 1 | 89.33% | 10.67% | | | |
| grade 2 | 86.27% | 13.73% | | | |
| instruction time | mean | std.dev. | p10 | median | p90 |
| grade 1 | 6.17 | 0.86 | 5 | 6 | 7 |
| grade 2 | 6.30 | 0.87 | 5.5 | 6 | 7 |
| total experience | mean | std.dev. | p10 | median | p90 |
| grade 1 | 15.15 | 8.95 | 4 | 15 | 28 |
| grade 2 | 17.67 | 9.37 | 4 | 18 | 30 |
| class size | mean | std.dev. | p10 | median | p90 |
| grade 1 | 20.12 | 3.80 | 15 | 20 | 26 |
| grade 2 | 20.24 | 4.08 | 15 | 20 | 26 |
| peer effect | mean | std.dev. | p10 | median | p90 |
| grade 1 | 8.05 | 0.47 | 7.48 | 8.13 | 8.55 |
| grade 2 | 8.16 | 0.48 | 7.62 | 8.27 | 8.64 |

**Table 4.** Explaining math test scores

| math | model a | | model b | | model c | | model d | |
|---|---|---|---|---|---|---|---|---|
| | coeff. | p>\|t\| | coeff. | p>\|t\| | coeff. | p>\|t\| | coeff. | p>\|t\| |
| constant | 3.35 | 0.00 | 8.34 | 0.00 | 3.75 | 0.00 | 2.15 | 0.00 |
| time2 | 0.82 | 0.00 | 0.82 | 0.00 | 0.82 | 0.00 | 0.80 | 0.00 |
| $math_0$ | 0.67 | 0.00 | | | 0.65 | 0.00 | 0.65 | 0.00 |
| girl | | | -0.26 | 0.00 | -0.24 | 0.00 | -0.24 | 0.00 |
| m_dutch | | | 0.16 | 0.00 | -0.13 | 0.00 | -0.14 | 0.00 |
| f_dutch | | | 0.10 | 0.03 | -0.07 | 0.07 | -0.07 | 0.08 |
| m_edu_sec | | | 0.17 | 0.00 | 0.00 | 0.86 | 0.01 | 0.85 |
| m_edu_high | | | 0.46 | 0.00 | 0.10 | 0.00 | 0.10 | 0.00 |
| m_edu_uni | | | 0.55 | 0.00 | 0.19 | 0.00 | 0.19 | 0.00 |
| f_edu_sec | | | 0.08 | 0.02 | 0.06 | 0.01 | 0.07 | 0.01 |
| f_edu_high | | | 0.25 | 0.00 | 0.14 | 0.00 | 0.14 | 0.00 |
| f_edu_uni | | | 0.37 | 0.00 | 0.24 | 0.00 | 0.24 | 0.00 |
| duo | | | | | | | -0.10 | 0.01 |
| peer | | | | | | | 0.10 | 0.27 |
| time_math | | | | | | | 0.08 | 0.00 |
| experience | | | | | | | 0.00 | 0.25 |
| class_size | | | | | | | 0.01 | 0.03 |

**Table 5.** Educational production is not likely to be separable

| | $F$-value | Prob $> F$ |
|---|---|---|
| initial test score | 4.46 | 0.00 |
| girl | 3.89 | 0.00 |
| mother dutch | 8.46 | 0.00 |
| father dutch | 12.07 | 0.00 |
| education mother | 7.37 | 0.00 |
| education father | 6.16 | 0.00 |
| all variables | $9\times10^7$ | 0.00 |

**Table 6.** The distribution of slopes and averages over schools

| | slopes $\widehat{\beta}_{b,jk}$ | | | | | | averages $\overline{z}_{b,jk}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | p05 | p25 | p50 | p75 | p95 | $\sigma$ | p05 | p25 | p50 | p75 | p95 | $\sigma$ |
| initial test score | 0.57 | 0.63 | 0.76 | 0.87 | 0.98 | 0.15 | 7.31 | 7.91 | 8.17 | 8.43 | 8.62 | 0.44 |
| mother with university degree | -0.67 | 0.43 | 0.17 | 0.46 | 1.21 | 0.58 | 0.00 | 0.03 | 0.08 | 0.15 | 0.24 | 0.08 |
| father with university degree | -0.47 | -0.13 | 0.18 | 0.39 | 0.67 | 0.42 | 0.00 | 0.06 | 0.11 | 0.17 | 0.30 | 0.10 |

**Table 7a.** Incentives for good administration when changing the initial test score slope

| change ideally | | increase slope ($+1\sigma$), but no change in output $\Delta s_j$ should be zero for all schools | | | | | | | decrease slope ($-1\sigma$), but no change in output $\Delta s_j$ should be zero for all schools | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| statistic | | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 |
| $RA$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $VA$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $RB$, with $\widetilde{z}$ | low | -0.08 | -0.07 | -0.05 | -0.04 | 0.00 | 95.9% | 4.1% | 0.00 | 0.04 | 0.05 | 0.07 | 0.08 | 4.1% | 95.9% |
| | mid | -0.03 | -0.02 | 0.00 | 0.02 | 0.05 | 51.3% | 48.7% | -0.05 | -0.02 | 0.00 | 0.02 | 0.03 | 48.7% | 51.3% |
| | high | 0.00 | 0.01 | 0.03 | 0.04 | 0.08 | 5.6% | 94.4% | -0.08 | -0.04 | -0.03 | -0.01 | 0.00 | 94.4% | 5.6% |
| $UO$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |

**Table 7b.** Incentives for good administration when changing the parental education slope

| change ideally | | increase slope ($+1\sigma$), but no change in output $\Delta s_j$ should be zero for all schools | | | | | | | decrease slope ($-1\sigma$), but no change in output $\Delta s_j$ should be zero for all schools | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| statistic | | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 |
| $RA$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $VA$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $RB$, with $\widetilde{z}$ | low | -0.10 | -0.06 | -0.04 | -0.02 | 0.00 | 100% | 00.0% | 0.00 | 0.02 | 0.04 | 0.06 | 0.10 | 0.00% | 100% |
| | mid | -0.07 | -0.03 | 0.00 | 0.02 | 0.03 | 48.1% | 51.9% | -0.03 | -0.02 | 0.00 | 0.03 | 0.07 | 51.9% | 48.1% |
| | high | 0.00 | 0.04 | 0.07 | 0.09 | 0.10 | 04.4% | 95.6% | -0.10 | -0.09 | -0.07 | -0.04 | 0.00 | 95.6% | 4.4% |
| $UO$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |

**Table 8a.** Incentives for pupil selection when changing the initial test score distribution

| change | | attract pupils with a higher initial test scores ($+1\sigma$) | | | | | | | attract pupils with a lower initial test scores ($-1\sigma$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ideally | | $\Delta s_j$ should be zero for all schools | | | | | | | $\Delta s_j$ should be zero for all schools | | | | | | |
| statistic | | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 |
| $RA$, with $\widetilde{\beta}$ | low | 0.00 | 0.01 | 0.03 | 0.05 | 0.07 | 5.3% | 94.7% | -0.07 | -0.05 | -0.03 | -0.01 | -0.00 | 94.7% | 5.3% |
| | mid | -0.03 | -0.02 | 0.00 | 0.02 | 0.04 | 50.7% | 49.3% | -0.04 | -0.02 | 0.00 | 0.02 | 0.03 | 49.3% | 50.7% |
| | high | -0.07 | -0.06 | -0.04 | -0.02 | 0.00 | 95.4% | 4.6% | 0.00 | 0.02 | 0.04 | 0.06 | 0.07 | 4.6% | 95.4% |
| $VA$ | | -0.03 | -0.02 | 0.00 | 0.02 | 0.04 | 47.6% | 52.4% | -0.03 | -0.02 | 0.00 | 0.02 | 0.04 | 52.4% | 47.6% |
| $RB$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $UO$ | | 0.10 | 0.11 | 0.14 | 0.15 | 0.18 | 0% | 100% | -0.18 | -0.15 | -0.14 | -0.11 | -0.10 | 100% | 0% |

**Table 8b.** incentives for pupil selection when changing the parental education distribution

| change | | attract pupils with highly educated parents ($+1\sigma$) | | | | | | | attract pupils with lowly educated parents ($-1\sigma$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ideally | | $\Delta s_j$ should be zero for all schools | | | | | | | $\Delta s_j$ should be zero for all schools | | | | | | |
| statistic | | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 | p05 | p25 | p50 | p75 | p95 | %<0 | %>0 |
| $RA$, with $\widetilde{\beta}$ | low | 0.00 | 0.04 | 0.05 | 0.07 | 0.08 | 5.1% | 94.9% | -0.08 | -0.07 | -0.05 | -0.04 | 0.00 | 94.9% | 5.1% |
| | mid | -0.05 | -0.01 | 0.00 | 0.01 | 0.03 | 55.8% | 44.2% | -0.03 | -0.01 | 0.00 | 0.01 | 0.05 | 44.2% | 55.8% |
| | high | -0.11 | -0.07 | -0.06 | -0.05 | -0.03 | 100% | 0.0% | 0.03 | 0.05 | 0.06 | 0.07 | 0.11 | 0.0% | 100% |
| $VA$ | | -0.05 | -0.02 | -0.01 | 0.01 | 0.02 | 61.3% | 38.7% | -0.02 | -0.01 | 0.01 | 0.02 | 0.05 | 38.7% | 61.3% |
| $RB$ | | zero for all schools | | | | | zero % | | zero for all schools | | | | | zero % | |
| $UO$ | | -0.04 | 0.00 | 0.01 | 0.02 | 0.04 | 25.2% | 74.8% | -0.04 | -0.02 | -0.01 | 0.00 | 0.04 | 74.8% | 25.2% |