

KU LEUVEN

DEPARTMENT OF ECONOMICS

Measuring and explaining organizational effectiveness of school districts:

Evidence from a robust and conditional benefit-of-the-doubt approach

Kristof DE WITTE and Fritz SCHILTZ

FACULTY OF ECONOMICS AND BUSINESS



DISCUSSION PAPER SERIES DPS17.20

DECEMBER 2017

Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional Benefit-of-the-Doubt approach.

Kristof De Witte [‡] δ and Fritz Schiltz δ

(δ): Katholieke Universiteit Leuven (KU Leuven), Leuven Economics of Education Research, Naamsestraat 69, 3000 Leuven (Belgium), kristof.dewitte@kuleuven.be; fritz.schiltz@kuleuven.be.

(\ddagger): Maastricht University, Top Institute for Evidence Based Education Research, Kapoenstraat 2, MD 6200 Maastricht (the Netherlands), k.dewitte@maastrichtuniversity.nl.

Abstract

This paper constructs an innovative approach to measure school district organizational effectiveness. We use the fully nonparametric Benefit-of-the-Doubt approach to construct a composite indicator which reduces subjectivity in choosing weights attached to different functions of school districts. The model is extended to reduce the influence of outliers (robust) and to account for the exogenous environment (conditional). We apply the suggested technique to unique and self-collected data from surveys and interviews taken from school board members and principals. We illustrate why accounting for respondent characteristics is crucial when evaluating school districts. Next, we analyze the obtained aggregate measures to identify which organizational structures, board characteristics and management styles are of importance for the organizational effectiveness of school districts.

Keywords: Education; School district organizational effectiveness; Economics of education; Empirical management.

JEL-classification: L2, I2, I21

We are grateful to seminar participants in Rome and Leuven, and participants of the XXV Meeting of the Economics of Education Association, the 4th Workshop on efficiency in education, the 6th Workshop on efficiency and productivity analysis. In particular, Léopold Simar, Cinzia Daraio, José Manuel Cordero, John Ruggiero, Nicole Adler, Giannis Karagiannis, Giovanna D’Inverno, and Maria Silva. The authors acknowledge financial support from the Horizon 2020 program through the EDEN project (Education Economics Network) [grant number 691676]. Corresponding author: Fritz Schiltz.

1. Introduction

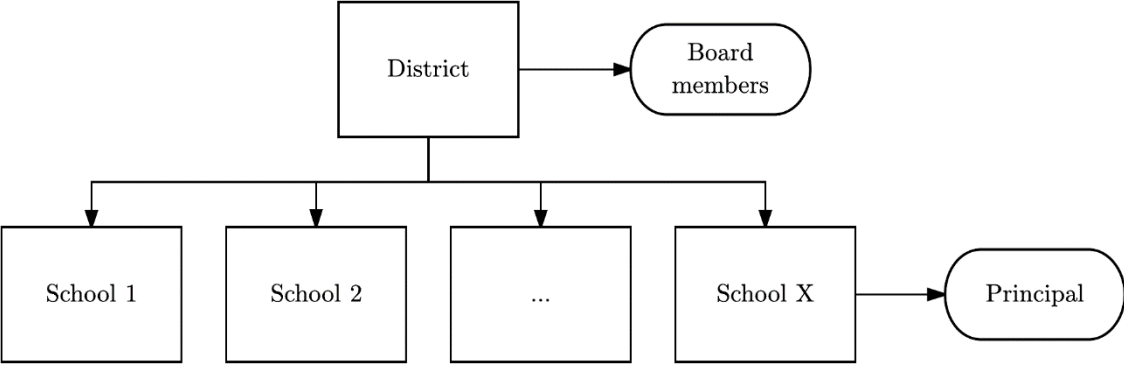
The New Public Management theories have created a growing tendency to improve effectiveness in public agencies and have stressed the importance of professionalism and optimization (Alonso, Clifton, & Díaz-Fuentes, 2015). Since its introduction in the 1980s, hospitals, nursing homes and other government-owned agencies have been transformed into more ‘market-friendly’ organizations. More recently, this observed trend is spreading towards education, both at the school and the school district level (Jarl, Fredriksson, & Persson, 2012). School districts are in charge of supervising schools, ranging from only one school to vast networks of schools. Their organization depends on the educational system. For example, in the US, school districts operate in a dedicated geographical area, while in many European countries school districts are not bound by catchment areas. Figure 1 displays the general hierarchical relationship between schools and the district. Schools are led by a school principal, while school districts are run by the school board.¹

Topics on organizational effectiveness have been extensively discussed at the school level (Ball, 2011; Bessent & Bessent, 1980; Di Liberto, Schivardi, & Sulis, 2015), while the study of organizational effectiveness is largely ignored at the school district level. Nevertheless, the latter is argued to play a major role in the management of schools because school districts determine the schools' mission and goals, and the selection and support of the school leader (Jackson, 2003). Student achievement has also been linked to the degree of organizational effectiveness in a school district (Bidwell & Kasarda, 1975; Ford & Ihrke, 2016). Using a transition probability model, Gritz and Theobald (1996) have found a relationship between organizational decisions (e.g. spending priorities) and length of stay in teaching by teachers. Unnever, Kerckhoff and Robinson (2000) identified a significant link between educational resources at the district level and student outcomes, stressing the importance of an adequate financial policy, one of the functions of a school district. The organizational effectiveness of school districts has also been linked to financial outcomes (Saatcioglu et al., 2011). Using data from 1800 high schools in 8 countries, Bloom et al. (2015) identified principal leadership and having strong accountability to an external governing body as major drivers of school-level variance in management quality. As this governing body is the school board at the district level, their results indicated a decisive role for school districts.

Organizational effectiveness in education is commonly measured as a single-item scale where respondents indicate their overall satisfaction with the school district (Bradshaw, Murray, & Wolpin, 1992). As opposed to a single question trying to capture the complex concept of organizational effectiveness, some studies have constructed a multi-item scale encompassing all district functions (e.g. Bloom et al., 2015). Most commonly, all components of a multi-item scale are averaged in order to obtain a composite indicator of organizational effectiveness. An important limitation of this approach is

¹ Throughout this text, we use the term ‘school board’ and ‘school district’ interchangeably.

Figure 1: Organizational structure in education systems.



that all dimensions have the same weights such that this outcome variable does not grant freedom to school districts to differ in their priorities. Some school districts might find one dimension more important than others such that assuming fixed weights for all districts does not grasp the heterogeneity in education. Moreover, in the absence of a consensus on the relative importance of the components, any choice of fixed weights will be subjective to some extent. Fixed weights may favor school districts who perform well on aspects with high weights, while disfavoring districts who excel on aspects with low assigned weights. As a final limitation, it is unclear for school districts what the scores imply as they are not constructed in a relative perspective to other school districts.

This paper contributes to the literature in several ways. First, we construct an innovative empirical measure of school district organizational effectiveness that summarizes all district functions into one composite indicator. Our approach is robust for outlying observations (or measurement error) and fully nonparametric. As a result, no a priori assumptions on the functional form (i.e. weighting the different functions) are required. In addition, the indicator is a relative measure that is easy to interpret and allows school districts sufficient leeway in setting their priorities. That is, school districts are compared to a frontier composed of the most effective districts. This frontier is constructed in a way that environmental factors, outside the influence of school districts, are taken into account in order to only benchmark school districts relative to districts operating in largely similar environments. This can be argued to be of major importance when operating environments (e.g. socio-economic composition of school districts) cannot be changed by the evaluated unit. Imposing the assumption that the environment can be controlled will result in an unfair comparison of units with frontier points outside their attainable set. In this paper, we use survey data and illustrate the importance of conditioning on respondent characteristics. Our measure of organizational effectiveness can be seen as ‘perceived’ effectiveness. Hence, the perceived frontier might be different depending on the evaluator. Evaluated units cannot change *who* is evaluating their effectiveness and hence, respondent characteristics can be considered exogenous. Applying the separability test of Daraio, Simar, & Wilson (2017), we formally reject the assumption that the evaluation does not depend on respondent characteristics (i.e. separability). Including differences in perceptions can be considered a requirement when using survey data. Therefore, we tailor the BoD

approach to a robust, conditional measure of organizational effectiveness of school districts. We further extend this approach, following Li, Simar, & Zelenyuk (2016), by estimating bandwidths for continuous variables (school district environment) separately for unique combinations of discrete variables (respondent characteristics), in order to obtain conditional measures. However, to the best of our knowledge, this methodology has not been previously applied to measure the organizational effectiveness of school districts. Note that when evaluating organizational effectiveness of other entities, this approach can also be useful, especially when using survey data.²

As a second contribution, we analyze the obtained aggregate measures to identify which organizational structures, board characteristics and management styles are of importance for the organizational effectiveness of school districts. To increase internal consistency, we use the same analytical framework, following Daraio & Simar (2005), avoiding the need to impose separability (Daraio et al., 2017; Simar & Wilson, 2007). In order to illustrate the usefulness of our approach, we apply the method to a detailed and self-collected dataset for school districts in Flanders. Our results indicate that school districts adhering to a participative management style and districts with more experienced board members outperform other districts in terms of organizational effectiveness. Also, private school districts achieve significantly higher levels of organizational effectiveness, along with larger districts and consolidated rather than cooperating districts.

The paper unfolds as follows. In the following section we provide an overview of the available literature and in so doing, motivate the need for an empirical, quantitative approach towards measuring and explaining organizational effectiveness at the school district level. Section 3 introduces the methodology. Next, we present the setting and data. In section 5 we present the results and explain the observed variance in organizational effectiveness. Section 6 discusses our findings and concludes.

2. Literature and background

As outlined in the introduction, many leading papers have pointed towards school districts as determinants of high performing schools. However, school districts as the unit of analysis received little attention. A common approach in existing studies is to estimate education production functions or cost functions and include school district characteristics, such as district size.³ Some papers have studied school district organization from a political economy point of view, by focusing on labor unions and teacher wages (e.g. Rose & Sonstelie, 2010). However, most papers are situated in the educational or pedagogical literature. Hence, evidence from these studies on organizational effectiveness of school districts is almost entirely based on case-studies as a method to define best practices. The most prominent example is known as the ‘Lighthouse Study’, conducted by the Iowa Association of School Boards (Delagardelle, 2008). Organizational effectiveness was measured by the adherence to ‘seven

² To facilitate further applications, the R code is available upon request.

³ School district size has been studied extensively (Andrews, Duncombe, & Yinger, 2002; Leach, Payne, & Chan, 2010).

conditions for productive change', put forward by Gemberling, Smith, & Villani (2000). Following this report, high-achieving school districts are those districts more likely to perform well on the subsequent set of conditions: Vision, Standards, Assessment, Accountability, Climate and Culture, Collaboration and Community Engagement, and Continuous Improvement. A more comprehensive description of these concepts can be found in Gemberling et al. (2000). A measure of the 'degree of adherence' was constructed by interviewing board members and school staff. This approach is similar to a study by Ford and Ihrke (2016), who identified a relationship between district-level outcomes and adherence to a set of best practices. Johnson (2012) proposes another set of 12 key practices, based on a review of the literature. There are only few quantitative studies focusing on the school district level⁴, including Hoffman (1995), who investigates a sample of 133 school boards. He defines effective school boards (districts) as "school boards that involve school team and parents (committee) in their decision making process". (Hoffman, 1995, p. 308). This finding indicates the importance of a bottom-up approach in managing school districts. With respect to the Flemish educational setting, earlier work on the organization of school districts is marked by several limitations (Caenepeel, 1988; Devos, 2008). They are rather dated and are incomplete because of the non-proportional sampling of the different school districts in relation to their education providers (community, official subsidized, grant aided) and their diversity (number of pupils, types of education etc.). Moreover, their analysis has a low power, such that it is difficult to draw strong and general conclusions out of them.

The current literature on the organizational effectiveness of school districts has also been criticized. Usdan et al. (2001) argue that studies proposing sets of best practices should be interpreted with caution since "best practices are case-and country specific" (Usdan et al., 2001). Johnson (2012, p. 89) argues that research on the organizational effectiveness of school districts is "saturated with many opinion-based articles". Similar statements have been made by Alsbury (2008) and Delagardelle (2008). Land (2002) observed that the available literature is "rife with conclusions and recommendations based on personal experience, observations and opinions and a heavy reliance on anecdotal evidence rather than on well-designed research studies" (p. 265).

To summarize, only few studies investigate *what* constitutes a good school district, or *how* a school district should be organized. To the best of our knowledge, there is no well-established evaluation methodology to compare the effectiveness of school districts that have multiple tasks and are operating in a heterogeneous environment. Therefore, we provide a framework to benchmark school districts and obtain a measure of organizational effectiveness.

⁴ This in contrast with the abundant quantitative literature on school management practices (e.g. Ball, 2011; Bloom et al., 2015, Di Liberto et al., 2015). However, this level of organization is not the focus of our paper. Nevertheless, we will include variables at the school level to control for the heterogeneity between schools when we assess the drivers of school district organizational effectiveness in section 5.

3. Methodology

3.1. ‘Benefit-of-the-Doubt’ (BoD)

To estimate the multidimensional measure of school district organizational effectiveness, we use a nonparametric model, rooted in data envelopment analysis (DEA). Our methodology follows the ‘Benefit-of-the-Doubt’ model (BoD), as an extension of DEA, originally developed by Farrell (1957) and put into practice by Charnes et al. (1978).⁵ Formally, all inputs are assumed to be equal to 1 for all evaluated school districts. However, this idea is not new. It was originally proposed by Thompson et al. (1986), and Adolphson, Cornia, & Walters (1991), coined the term ‘Benefit-of-the-Doubt’ by Melyn & Moesen (1991) who applied the method to obtain an indicator of macroeconomic performance, and subsequently formalized under a DEA framework by Lovell and Pastor (1999). In 2008, this approach was suggested by the OECD as a method to construct composite indicators (OECD, 2008), spurring many applications in different fields: among others, in bibliometrics (García-Romero, Santín, & Sicilia, 2016), health care (Shwartz, Burgess, & Zhu, 2016), competitiveness (Li & Zhao, 2015), police effectiveness (Verschelde & Rogge, 2012), and energy (Zanella, Camanho, & Dias, 2015).

In our application, outputs consist of four school district functions. The weights, u , attached to the different functions, y (1, ..., r, ..., s) are endogenously determined by the model such that this set of weights is not subject to ex ante assumptions.⁶ In other words, the weights for the evaluated school district o are chosen in such a way to maximize its relative strengths: $\max \sum_{r=1}^s u_r y_{ro}$. By applying this model, each school district is granted the ‘benefit of the doubt’ in determining the relative importance of each district function. The BoD model described above can be translated into the following notation (Cherchye, Moesen, Rogge, & Puyenbroeck, 2007):

$$\max_u \theta_o(y) = \sum_{r=1}^s u_r y_{ro} \quad (1)$$

s. t.

$$\sum_{r=1}^s u_r y_{rj} \leq 1 \quad (j = 1, \dots, n) \quad (1a)$$

$$u_1, u_2, \dots, u_s \geq 0 \quad (1b)$$

$$\sum_{r=1}^s u_i = 1 \quad (1c)$$

With n the number of evaluated school districts, s the number of outputs y , weighted by u for school district o , generating an indicator of school district organizational effectiveness θ_o . The maximization is subject to three constraints. First, the *normality constraint* (1a) imposes the organizational

⁵ Examples of DEA applications in education are widely available. For example, at the school level (Bessent & Bessent, 1980; Charnes, Cooper, & Rhodes, 1978), at the district level (Färe, Grosskopf, & Weber, 1989; McCarty & Yaisawarng, 1993), and in higher education (Johnes & Johnes, 2009). An in-depth literature review on efficiency techniques in education is available in De Witte & López-Torres (2017).

⁶ We are grateful to an anonymous referee for valuable suggestions on the matter of output transformations (Barnum, Coupet, Gleason, McWilliams, & Parhankangas, 2017). In the evaluation of school districts, we can reasonably assume that a decrease in effectiveness for one district function can be offset by a higher effectiveness for another function. School districts set priorities and divide their time and board members accordingly.

effectiveness of school districts to be equal to at most 1 for all districts in the reference set, when the optimal weights of district j are attached to the corresponding school district functions. Values of $\theta_o=1$ denote an effective unit, while $\theta_o \leq 1$ indicates that unit o can be considered ineffective. This implies that organizational effectiveness scores below unity indicate that the district performs ‘worse’ since there exist other districts in the sample outperforming district j , even when its optimal set of weights are attached. Second, the *non-negativity constraint* (1b) requires weights to be positive (but not strictly). In other words, school district organizational effectiveness is a non-decreasing function of the components of this indicator. For example, a higher evaluation of the financial policy (one of the district functions, see 4.) will not lower the aggregate indicator of organizational effectiveness. Lastly, the third constraint reflects the assumption of *variable returns to scale* (1c) imposed on the data. In BoD model (1), and in general DEA applications, convexity is implicitly imposed. Convexity implies that points on the frontier used to evaluate observations can be constructed based on linear combinations of actual data points. This assumption is particularly strong since some combinations of observations can never be actually realized (Cherchye et al., 2007). Adjusting the BoD model to release the assumption of convexity results in the ‘Free disposal hull’ (FDH) model, developed in Deprins, Simar, & Tulkens (1984). Formally, FDH adds the following condition:

$$u_i \in \{0,1\} \quad (1d)$$

3.2. Robust BoD

Due to its deterministic nature measurement errors or outliers in the data can heavily influence the obtained measures of school district organizational effectiveness. Therefore, we adapt (1) to the robust order- m approach of Cazals, Florens, & Simar (2002). This method limits the influence of outliers and measurement errors by repeatedly drawing (B times) samples (size $m < n$) from the available data set. This re-sampling technique computes BoD estimates for each subsample of size m and averages these B estimates into an indicator which is robust to the limitation of BoD as a deterministic estimator:

$$\hat{\theta}_o^{OM}(y) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_o^b(y) \quad (2)$$

By drawing, with replacement, subsamples of size m , a less extreme benchmark is used, increasing the score attached to an evaluated unit. Note that when $m \rightarrow n$, the ‘robust BoD’ estimates correspond with those obtained using the deterministic model (Cazals et al., 2002).⁷ Also, since $m < n$, it is possible that an evaluated school district does not belong to the set of districts used to obtain its own score of organizational effectiveness. This observation will then be located above the production frontier, resulting in a score above unity (i.e. ‘super-effective’). The order- m indicator of school district organizational effectiveness is the BoD score of a school district relative to the expected maximum score among m randomly drawn school districts. Hence, a super-effective school district can be seen as a

⁷ Remark that as $m=n$, these scores do not necessarily coincide because the resampling technique is based on random sampling samples of size m from the main sample of size n .

district that is doing better than the average m other school districts in its reference sample. Depending on the choice of m , the proportion of super-effective observations varies since the size of the drawn sample (m) relative to the total sample size n influences the probability of observation o not belonging to the production frontier.⁸ In order to compute the organizational effectiveness of school districts in the following section, we have set m at levels in accordance with Daraio & Simar (2007). In our application, we use $m=40$.⁹ The number of samples (B) drawn to calculate our robust measure of organizational effectiveness will affect the accuracy of this measure. Therefore, B is preferably as high as possible, although more repetitions will require more computing power. We have set B equal to 2000 for all composite indicators.

3.3. Robust and conditional BoD

Environmental variables that are not under the control of school districts need to be taken into account when calculating a (robust) BoD score of organizational effectiveness. This point has been developed by Simar & Wilson (2007) and further stressed in Simar & Wilson (2011). Failure to include this set of exogenous variables ($Z=z_1, \dots, z_R$) returns biased evaluations of school district organizational effectiveness when Z affects the attainable set (i.e. the production frontier). If the attainable set is not affected by Z , the ‘separability condition’ is said to hold. In the application at hand, imposing separability is especially stringent since it implicitly assumes that all survey respondents evaluate school districts relative to the same frontier. That is, the underlying, ‘true’ effectiveness levels do not depend on values of Z .¹⁰ Previous studies using survey data have shown that respondent characteristics do significantly affect the attainable set – or at least the perceived attainable set (Cordero, Salinas-Jiménez, & Salinas-Jiménez, 2017; Verschelde & Rogge, 2012). In addition, other environmental variables can be related to the frontier, such as the location of the evaluated unit (Fusco, Vidoli, & Sahoo, 2017) and its political environment (Kristof De Witte & Geys, 2013). In order to circumvent the restrictive separability condition, we choose to develop our model building on the work of Cazals et al. (2002) and Daraio & Simar (2005; 2006; 2007). These authors propose a method to incorporate Z into the analysis to obtain so-called ‘conditional’ measures. Intuitively, school districts will be evaluated relative to a reference group characterized by a similar environment (observations i for which $z_i \approx Z$). We tailor the model in (2) to a robust and conditional Benefit-of-the-Doubt indicator:

$$\hat{\theta}_o^{OM}(y|z) \tag{3}$$

In contrast to (2), where all observations have an equal probability of belonging to the reference sample, the probability of being drawn for each school district depends on its vector Z when estimating

⁸ Therefore, m can be used in a dual meaning: (i) a trimming parameter to choose a specific level of robustness, and (ii) the number of *potential competitors* drawn for every subset b when evaluating unit o .

⁹ The value of m is set in a way to attain a sufficiently small decrease in the proportion of super-effective school districts, see Figure A1 . However, adjusting m does not alter our findings (i.e. we also considered $m=100$).

¹⁰ For a clear graphical illustration of this assumption, see Simar & Wilson, 2007; Figure 1, p. 207).

the conditional order- m BoD score. To do so, we smooth Z by estimating a kernel function around z . The probability of being drawn then depends on the resulting kernel density function, evaluated at Z . For example, school districts evaluated by young respondents are more likely to be drawn for membership in the reference sample (of size m) when the evaluated school district is assessed by a young respondent. Considering possible systematic response patterns across respondents, the concept of benchmarking school districts relative to districts operating in largely similar environments (in a conditional approach) can be argued to be preferable over imposing separability (in an unconditional approach), where Z is assumed to be independent of the attainable set.¹¹ In the specific case where all variables in Z are continuous, the approach of Bădin, Daraio, & Simar (2010), based on Hall, Racine, & Li (2004) can be implemented. In the more general case where Z consists of both continuous and discrete variables, different strategies can be followed. One possibility here is to create subsamples by setting the bandwidth equal to zero for discrete variables. Another possibility is to apply the kernel methods suggested by Li & Racine (2004) in order to smooth all variables in Z , both discrete and continuous. This approach is adopted in De Witte & Kortelainen (2013) and has been applied in other studies estimating robust conditional indicators (e.g. Cordero, Alonso-Morán, Nuño-Solinis, Orueta, & Arce, 2015; Verschelde & Rogge, 2012). More recently, Li, Simar, & Zelenyuk (2016) proposed an alternative procedure that allows different bandwidths for continuous variables in different subgroups and allows smoothing over the discrete variables (Li et al., 2016).¹²

The resulting estimates are robust to outlying observations and measurement error, and account for systematic response patterns and heterogeneity in the educational landscape by including Z as a condition when estimating school district organizational effectiveness. Again, $\hat{\theta}_o^{OM}(y|z)$ can be larger than unity due to subsampling in the order- m approach. Super-effective units can now be seen as school districts outperforming the average m other school districts in its reference sample, operating under largely similar environmental conditions.

3.4. Inference

As an additional advantage of estimating robust conditional BoD indicators, the influence of variables in Z can be interpreted. The effect of Z on school district organizational effectiveness can be evaluated by looking at the ratio of the conditional (3) over the unconditional (2) estimate:

$$Q_o = \frac{\hat{\theta}_o^{OM}(y|z)}{\hat{\theta}_o^{OM}(y)} \quad (4)$$

This ratio can be nonparametrically regressed on a variable in Z of interest. Graphically, the slope of the smoothed regression line offers an interpretation of the marginal effect of this variable on the attainable

¹¹ Daraio, Simar, & Wilson (2017) provide a formal test of separability, building on Kneip, Simar, & Wilson (2016).

¹² We follow this idea by estimating separate bandwidths for continuous variables for unique combinations of discrete variables. For the sake of completeness, we provide in Table A5 the results obtained by using the smoothing method proposed by Li & Racine (2004), which is preferable in settings with many categories and small samples. Our main findings are robust to this alternative smoothing approach.

set (Daraio & Simar, 2005; 2007). In our conditional and robust BoD model – with output orientation and effectiveness scores between 0 and 1 (excluding super effective districts) –, a positive slope indicates that the variable in Z , say z , can be deemed unfavorable. That is, the frontier used to evaluate school districts with a high value of z will be positioned lower than the frontier constructed by the unconditional model. Analogously, a negative slope indicates a positive effect and a straight line indicates the absence of an effect. A combination of effects is also possible since the line is smoothed using a nonparametric regression.¹³ It is possible to test the statistical significance of these effects. In order to obtain significance levels, we follow the nonparametric bootstrap approach by Li & Racine (2007). As noted before, an alternative approach would be to regress unconditional effectiveness scores on explanatory variables, assuming separability holds (i.e. ‘two-stage approach’). In section 5 we formally reject the validity of this assumption to motivate a conditional rather than a two-stage approach. This is particularly relevant here due to possible systematic response patterns when using survey data (see before). In the same vein, we can include explanatory variables of interest in the conditional estimates in order to mimic second-stage regressions, without the need to impose separability. Variables can be gradually added by extending Z when estimating the conditional scores. In section 5 we regress the ratio, Q , on variables in our dataset to identify determinants of organizational effectiveness.

4. Education system and data

4.1. The organization of school districts in Flanders

This paper analyzes the organizational effectiveness of school districts in Flanders, the northern part of Belgium. The region of Flanders makes an interesting application as school districts are not organized based on geographical criteria or catchment areas. Schools organize themselves in school districts, governed by school boards, which are responsible for the financial state, HR-policy, strategy, etc. of their school(s) (see again Figure 1). In practice, school districts are not necessarily in charge of these responsibilities and often leave the decision making to the schools themselves. Some districts act as a financial supervisor while others are actively involved in the pedagogical and HR policies of their schools.

Correspondences between the Flemish educational system and other countries are significant as, in line with many other OECD countries like the Netherlands, Germany, the UK, Spain and Italy, it is marked by a recent trend towards decentralization (e.g. Burns & Köster, 2016). School districts in Flanders have full autonomy over the scale of their operations. Nevertheless, financial incentives are provided by the government to encourage larger scale joint structures. ‘School communities’ are the most common form of cooperation. Communities of schools are formed on a voluntary basis, yet this organizational form is encouraged by the Flemish government.¹⁴ Consolidated school districts have a

¹³ See for example, Daraio & Simar (2006, p.532).

¹⁴ School districts belonging to different types of providers are allowed to cooperate, although most school communities remain provider-specific.

structure where different districts are merged into one, led by a unified school board in charge of different schools. Some school districts choose not to consolidate nor cooperate and hence do not have a joint structure.

While not all schools are organized by the government, all Flemish schools are publicly *subsidized*. The subsidy is granted to the school district, which oversees the distribution of subsidies to its schools. As some schools are organized by the government, we observe both government and non-government districts. The former consists of community, local and central government institutions while private education providers make up the latter. In sum, we observe a heterogeneous educational landscape, ranging from very small one-unit school districts to large, professionally run, school districts. In the following sections we will exploit this variety in organizational structures to obtain the drivers of school district organizational effectiveness.

4.2. Data

The data used in this study consists of a combination of survey data and administrative data on school and school district characteristics. Survey data was obtained by sending out comprehensive surveys to school principals, board members and presidents of all school districts in Flanders. In an earlier stage of this research, a qualitative analysis suggested that respondents indicate four functions as the major responsibilities of a school district: pedagogics, human resources (HR), financial policy, and investments & infrastructure.¹⁵ The survey subsequently sent out to principals and school board members across Flanders was designed in such a way to cover these district functions. The questionnaire comprised 272 items, including questions with respect to respondent characteristics, variables identified as essential in the literature (Devos et al., 1999; Vanhoof et al., 2012), and a set of (five-point) Likert questionnaire items dealing with each district function.¹⁶ The number of respondents equals 150, consisting of 98 school principals and 52 school board members. The response rate reaches almost 30% and is particularly high because all surveys were sent out with the support of education providers in Flanders. The self-collected survey data are matched to administrative data (covering all districts) provided by the Flemish Ministry of Education. All data corresponds to the year 2012.

Summary statistics are provided in Table 1. For some school districts we use evaluations of school board members (at the district level) and for others we use data from school principals. In our dataset, it does not occur that multiple types of respondents can be identified with one school district.¹⁷ In essence, we assume school principals to be able to evaluate the school district supervising their schools. By

¹⁵ In-depth interviews were conducted in an earlier stage of this research with 90 members of school boards, which are covering 955 schools. The interviewees were selected to optimally represent the diverse educational landscape in Flanders.

¹⁶ Throughout this paper, the assumption is made that Likert type items are not problematic to obtain a measure of effectiveness, as in Verschelde & Rogge (2012). As suggested by one referee, extending the conditional BoD to accommodate discrete and bounded data (Chen, Cook, Du, Hu, & Zhu, 2017) can be a promising avenue for future research.

¹⁷ Note that the organizational effectiveness of every school district is evaluated by a single respondent as, to obtain reliable insights, we invited only high-level respondents that are familiar with these matters. These are either school board members or school principals. The former group consists of only 8 people per district, while the latter is at most one if there is one school in a district. Hence, the number of respondents per district is low by construction.

including respondent characteristics in our conditional estimates, we also account for the type of evaluator when comparing school districts. Hence, if systematic response patterns exist between groups, then we account for this using conditional models. As a result, we are able to evaluate school districts using opinions of school board members and school principals.

From Table 1, we can see that most respondents are female, respectively 66 and 71 percent for principals and members of the school board. Board members also tend to be older, especially in the oldest age groups. Almost 14 percent are above the age of 61, compared to 0 percent for school principals. The age variable is aggregated as a dummy indicating whether the respondent has an age below 50 years. More than half of the respondents (54%) holds a Master's degree or above (e.g. PhD). This number is inflated by almost all board members holding a 'high' degree. School board members are mostly retired and/or volunteers, although one in four is also considered a civil servant. This is partly due to board members taking up a political mandate. Remarkably, political mandates appear more frequently among school principals compared to school board members. Some board members (17 percent) are still active, either as employees or self-employed, whereas the majority is considered 'inactive' by official statistics.

In addition to respondent characteristics, we also include variables suggested in the existing literature to be of importance. The set of variables includes the structure of the organization, socio-economic environment, management style, and school board composition (expertise). This allows us to explain differences in school district organizational effectiveness, and determine which variables can be identified as decisive. Table 1 indicates that most of the districts in our sample (80.1%) are subsidized, but not organized, by the government. Other districts (19.9%) can be organized by community, local, or central governments. Apart from the type of organizer, districts can differ in their organizational structure. We disentangle three organizational structures (see section 2.1.): no joint structure (2.65% of the sample), community (88.08%), and consolidation (9.27%). As expected, most school districts belong to a school community as participation is strongly encouraged by the Flemish government.

School districts in our sample consist of almost 6 schools on average, ranging from 1 to 59 schools per district. An average school district employs almost 280 full time equivalents (FTE) and less than half at the median. As illustrated by these two variables, there is a large variation in size in our sample.¹⁸ To mitigate possible bias from this strong variation, we follow Groenez et al. (2015) to group our sample in either small or large districts. The chosen cutoff value is identical to theirs and set at 6 schools per district. This dummy variable indicates that 22.5% and 77.5% of Flemish school districts can be considered as large and small, respectively. The percentage of tenured personnel reflects the share of FTE at the district level registered as tenured, amounting to 72% on average. The variables 'GON resources' and 'Maternal education' are included in our analysis to account for differences in student

¹⁸ Table A1 shows that variation in the full population is also high when size-related variables are considered. This heterogeneity in size and, hence, organizational structure is a specific characteristic of the Flemish educational landscape and supports our choice of Flanders for the study of school districts.

composition between districts. On average, slightly above 1 percent of the students in our sampled districts is eligible for additional funding to support integration of special needs students (GON). One in sixteen mothers (0.0625) of students in our sample attained a high school diploma as their highest degree. Both indicators vary strongly across districts, within a range of [0-0.06] and [0-0.292], respectively.

In line with Hoffman (1995), we include an indicator of management style to capture the type of collaboration between school principals and the school district (see again Figure 1). A participative management style is a type of management in which stakeholders at all levels in Figure 1 are encouraged to get involved in decision making. This style corresponds to a ‘bottom-up approach’. Supportive management by school districts leaves all major decisions to lower levels of the organization (i.e. school principals). In supportive school districts, schools operate as autonomous entities with only limited coordination between them. School districts adopting a centralized management style are the opposite of participative districts, by centralizing the decision-making process. All processes run top-down, and schools retain only limited autonomy. We create a dummy variable by counting the number of district functions that is considered ‘participative’, ‘supportive’ or ‘centralized’ by the respondent and set the dummy equal to one if the majority of district functions follows a participative style. By construction, the other management styles cannot be the majority, and make up the counterfactual. This allows us to check the claim by Hoffman (1995), who defines effective school boards (districts) as “school boards that involve school team and parents in their decision-making process”. Table 1 indicates that around one in five school districts adheres to a participative management style for the majority of their functions. A final variable is the available expertise in the evaluated school districts. It is measured as the number of district functions where expertise is available.

Table A3 in the Appendix compares mean values of variables from the administrative dataset for the entire population of schools and school districts (Table A1) and for the sample of our analysis (Table A2). The last column of Table A3 displays t-statistics and only indicates significant differences (at the 1% level) for both size variables, i.e. schools per district and FTE per district. The sample mean is skewed by the largest district containing 59 schools. This explains the discrepancy between sample mean (5.7) and population mean (3.3). We account for this skewed distribution by recoding the size variable (see above). Maternal education is slightly higher in the population, but this difference is negligible. Other variables do not differ significantly, confirming the representativeness of our sample.

Table 1: Summary statistics: Respondent characteristics and explanatory variables.

Variables		N=150^a				
Respondent Characteristics		Yes	No			
Gender (Male = yes)		32.45%	67.55%			
Age (< 50 years old = yes)		33.11%	66.89%			
Education: Master's or above		54.30%	45.70%			
Political Mandate		5.96%	94.04%			
Board Member		34.44%	65.56%			
Status ^b :	Active	17.00%	83.00%			
	Civil servant	25.00%	75.00%			
	Retiree	46.00%	54.00%			
	Volunteer	12.00%	88.00%			
Explanatory variables		Yes	No			
Non-government board		80.13%	19.87%			
More than 6 locations		22.52%	77.48%			
Participative Management ^c		21.19%	78.81%			
Structure	No joint structure	2.65%	97.35			
	Community	88.08%	11.92%			
	Consolidation	9.27%	90.77%			
		Mean	SD	Median	Min	Max
Participative Management ^d		3.94	2.684	3.00	0.00	10.00
Expertise		3.89	1.376	4.00	1.00	7.00
Schools per district		5.74	9.109	3.00	1.00	59.00
FTE per district		278.40	485.80	123.90	4.50	3213.00
% tenured ^e		0.72	0.056	0.717	0.514	0.85
GON resources ^f		0.011	0.0080	0.010	0.00	0.06
Maternal education ^g		0.063	0.051	0.048	0.00	0.29

a. To maintain an equal number of observations throughout our analysis, we applied the multiple imputation technique.

b. Status of respondents are only listed for board members. All principals are listed as civil servants. Respondents were allowed to select multiple statuses; e.g. most retirees are also volunteers.

c. The first variable 'Participative management' is a dummy variable equal to one if the majority of district functions is organized in a 'participative manner'.

d. This variable is discrete and simply counts the number of functions where the district's management style is considered participative. e. This variable indicates the percentage of full-time equivalents (FTE) registered as tenured personnel.

f. The Flemish funding mechanism provides additional funding for special needs students. This variable captures the share of eligible students in a school district.

g. Maternal education indicates the share of students in each district whose mother only obtained a high school degree or below.

Table 2: Scores per function and organizational effectiveness of school districts.

Functions	N	Mean	SD	Median	Min	Max	<i>u</i>
Human Resources	150	0.805	0.164	0.813	0.303	1.064	0.056
Financial policy	150	0.924	0.113	1	0.401	1.037	0.826
Investment & Infrastructure	150	0.813	0.167	0.800	0.100	1.047	0.110
Pedagogics	150	0.828	0.208	0.903	0.200	1.129	0.039
Organizational effectiveness							
Unconditional	150	0.942	0.087	0.996	0.603	1.066	
Conditional 1	150	0.968	0.060	1.000	0.636	1.002	
Conditional 5	150	0.980	0.051	1.000	0.651	1.000	

Note: Summary statistics are presented in histograms in Figure A2 and A3 in the Online Appendix.

5. Results

Table 2 displays the scores for every function, obtained by aggregating the questionnaire items in Table A4 (i.e. HR (6), financial policy (4), investment & infrastructure (3), and pedagogics (4)).¹⁹ Aggregation is done by applying model (2) for each function separately. Next, we obtain an unconditional composite indicator of organizational effectiveness by considering the previously computed district function scores as output variables in (2). Motivation to follow this aggregation procedure in two steps is twofold. First, it is consistent with our reasoning to abstain from imposing subjective weights, outlined in the introduction and methodology sections. Alternative approaches to reduce dimensions such as taking averages, or to perform PCA (Bădin, Daraio, & Simar, 2012) would not follow this reasoning. Second, including all 17 items in the assessment would decrease the discriminatory power of the model (Dyson et al., 2001), especially when applying (3).²⁰ Also, in the application at hand, it is interesting to consider the importance of each district function with respect to the overall score of organizational effectiveness. This issue is related to the r endogenously attached weights, u . These weights are also called ‘shadow prices’ in the literature and we derive them by regressing the overall score on all four components. Running the regression without an intercept returns coefficients which add up to one. The resulting shadow prices are listed in the final column of Table 2. Clearly, the financial policy of school districts is a major driver of organizational effectiveness (0.83). Investment & infrastructure also contributes to effectiveness (0.11) while both HR and pedagogics are of rather limited importance, with a combined weight equal to 0.09.

The overall (unconditional) measure of organizational effectiveness equals 0.941 on average. This number can be interpreted as follows: if all school districts would perform as well as the districts in their reference set, school districts would be able to improve their organizational effectiveness by almost 6% on average. Moreover, the worst performing district should be able to increase its organizational effectiveness by almost 40 percent. This large discrepancy in organizational effectiveness might be due school districts being compared to a frontier which they cannot reach. In other words, unconditional models impose that the attainable set is not dependent on Z . When survey data is used, this assumption is unlikely to hold since the perceived frontier might be different depending on the evaluator. Evaluated units cannot change *who* is evaluating their effectiveness and hence, respondent characteristics can be considered exogenous. Applying the test of Daraio et al. (2017), we formally check whether these exogenous respondent characteristics meet the separability condition and firmly reject it for all

¹⁹ The internal consistency of these items is discussed in Online Appendix A1.

²⁰ We also asked respondents to evaluate every district function on a scale of 1 to 10. Using these evaluations, we can compare the composite indicators for every function (2) with our single-item scale to evaluate the consistency of the district function scores. Depending on the chosen district function, both Spearman and Pearson correlation coefficients attain values between 0.4 and 0.7, as displayed in Figure A4. As we motivated in section 3, a multi-item evaluation allows school districts autonomy in setting priorities, which is not the case in the single-item scale. This supports our choice to obtain composite indicators for all district functions.

respondent characteristics.²¹ Including differences in perceptions (systematic response patterns related to respondent characteristics) can be considered a necessary requirement when using survey data to benchmark school districts (or other entities). Therefore, we estimate (3) to obtain a robust, and conditional measure of organizational effectiveness of school districts.

Once we allow a fair comparison of school districts by including respondent and school district environment characteristics (Conditional 1, Table 2), the average score increases (0.966) while the standard deviation decreases (0.065). This is intuitive, considering that we now benchmark school districts relative to districts operating in largely similar environments. Nevertheless, the lowest performing school district can still improve its organizational effectiveness by 37%, despite being compared to similar school districts. We gradually extend the conditional model by including variables related to the organization of school districts, identified as essential in the literature. This results in 4 additional conditional estimates.

To assess the impact of the variables in *Z* on the organizational effectiveness of school districts, we regress the ratio of unconditional scores over conditional scores (*Q*) on *Z*. Regression results are displayed in Table 3. We estimated 5 nonparametric regressions, one for each conditional estimate. All models estimated here control for respondent characteristics and school district environment – i.e. these variables are included in *Z*. From Table 3, we can see that board members, male and young respondents are generally more unfavorable in their evaluation of school districts. Other respondent and district environment characteristics are alternatingly or non-significant once organizational variables are included.

In the bottom panel of Table 3, we can see that all variables related to the organization of school districts are considered ‘favorable’. This is consistent with the literature since we selected variables identified in previous studies as being crucial for the organizational effectiveness of school districts. From models 1 to 5, we find that private school districts significantly outperform public school districts in terms of organizational effectiveness. In models 4 and 5, we find that district size (>6 schools per district) has a significantly favorable effect on organizational effectiveness. Also, consolidated districts are evaluated significantly favorable, while the difference for districts without a joint structure is not significant. A participative management style adhered by the school district also results in a favorable evaluation of the districts, as well as more experienced board members, although only the latter relationship is found to be significant – unlike in Table A5 where both variables are significant.

²¹ We follow the approach of splitting the sample into subgroups for discrete variables (respondent characteristics here), outlined in Daraio et al. (2017, Appendix C, p.4) and Kneip et al. (2016, p 437-439). We apply the deterministic rule for sample splitting provided by Daraio et al. (2017, Appendix C, p.5-6). *p*-values for respondent characteristics are all well below 0.001.

Table 3: Statistical inference of school district organizational effectiveness.

	N=150		Model 1		Model 2		Model 3		Model 4		Model 5	
	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value
<i>School district environment</i>												
Maternal education	unfavorable	0.890	unfavorable	0.470	unfavorable	0.399	unfavorable	0.943	unfavorable	0.009		
GON resources	favorable	0.349	favorable	0.107	favorable	0.329	favorable	0.723	favorable	0.903		
% tenured	unfavorable	0.079	favorable	0.463	unfavorable	0.916	unfavorable	0.339	unfavorable	0.675		
<i>Respondent characteristics</i>												
Gender (male=yes)	unfavorable	0.004	unfavorable	0.313	unfavorable	0.504	unfavorable	0.104	unfavorable	0.000		
Age (<50 years old = yes)	unfavorable	0.013	unfavorable	0.132	unfavorable	0.031	unfavorable	0.338	unfavorable	0.000		
Education: Master's or above	unfavorable	0.202	favorable	0.001	favorable	0.002	favorable	0.013	unfavorable	0.000		
Board member	unfavorable	0.118	unfavorable	0.007	unfavorable	0.007	unfavorable	0.015	unfavorable	0.007		
<i>Organization</i>												
More than 6 schools	favorable	0.580	favorable	0.305	favorable	0.432	favorable	0.005	favorable	0.001		
Private school district			favorable	0.290	unfavorable	0.177	favorable	0.051	favorable	0.032		
Participative management					favorable	0.379	favorable	0.556	favorable	0.512		
Consolidation							favorable	0.000	favorable	0.000		
No joint							favorable	0.415	favorable	0.352		
Expertise									favorable	0.052		

Note: p-values in bold indicate significance at least at the 10% level.

6. Discussion and Conclusion

This paper introduced an innovative approach to measure and explain school district organizational effectiveness. To the best of our knowledge, there is no well-established evaluation methodology to compare the effectiveness of school districts that have multiple tasks and are operating in a heterogeneous environment. Therefore, we provide a framework to benchmark school districts and obtain a measure of organizational effectiveness. We tailored the Benefit-of-the-Doubt model to a robust, conditional composite indicator. Our approach is robust for outlying observations (or measurement error) and fully nonparametric. As a result, no weights need to be imposed on district functions, which allows school districts sufficient leeway in setting their priorities. This can be considered to be of major relevance for the education sector where decentralization and autonomy are becoming increasingly important (Burns & Köster, 2016; Woessmann, Lüdemann, Schütz, & West, 2007). Also, school districts are compared to a frontier composed of the most effective districts, constructed in a way that environmental factors, outside the influence of school districts, are taken into account. In this paper, we use survey data and illustrate the importance of conditioning on respondent characteristics. Our measure of organizational effectiveness can be seen as ‘perceived’ effectiveness. Hence, the perceived frontier might be different depending on the evaluator. Evaluated units cannot change *who* is evaluating their effectiveness and hence, respondent characteristics can be considered exogenous. In our application, school districts were evaluated in a significantly unfavorable way by male and young respondents, and board members, after rejecting the separability condition for all

respondent characteristics. Including variables related to the organization of school districts reveals that school districts adhering to a participative management style and districts with more experienced board members outperform other districts in terms of organizational effectiveness. Also, private school districts achieve significantly higher levels of organizational effectiveness, along with larger districts and districts that follow a consolidated rather than a cooperative organizational structure.

It is important to ask to what degree the results in this paper can be generalized to other settings and there are certainly a number of caveats worth noting. First, we do not claim to present causal evidence, but we offer a contribution in terms of an empirical framework to benchmark school districts. Future research might expand the selection of environmental variables, consider multiple time periods or exploit exogenous shocks in school district policies. Second, in our dataset it does not occur that multiple respondents can be identified with one school district. If this would be the case, different respondents should again be weighted to obtain a score for each school district. Internal heterogeneity in assessing the organizational effectiveness of school districts and corresponding weighting procedures will have to receive further attention. Third, despite similarities between the education system in Flanders and other OECD countries, outlined in section 2, results cannot be simply extrapolated to other countries. However, the framework proposed here can serve as a starting point for further research. Considering the ability of conditional models to account for the exogenous environment, the methodology can be easily extended to other educational systems.

Policy implications can be deduced from our findings to improve the organizational effectiveness of school districts. School board members should be recruited based on experience. Also, investments in management training to improve cooperation between principals (at the school level) and board members (at the district level) will likely result in higher organizational effectiveness. There is a tendency to increase professionalism by enlarging the scale of operations in public sector entities (Alonso, Clifton, & Díaz-Fuentes, 2015). Our findings suggest possible benefits from increasing the scale of school districts, although policy makers should consider the organizational structure when pursuing this – i.e. through consolidation rather than cooperation. Finally, considering the persistent differences between public and private school districts, mediocre public districts can boost their organizational effectiveness significantly by implementing ‘best practices’ observed in the private sector. Once again, this illustrates the importance of benchmarking between school districts. Our approach aimed to provide a method to do so and, in so doing, strengthen the support schools, principals and ultimately students receive from their school districts.

References

- Adolphson, D. L., Cornia, G. C., & Walters, L. C. (1991). A Unified Framework for Classifying DEA Models. *Operational Research '90*, 647–657.
- Alonso, J. M., Clifton, J., & Díaz-Fuentes, D. (2015). Did New Public Management Matter? An empirical analysis of the outsourcing and decentralization effects on public sector size. *Public Management Review*, 17(5), 643–660.

- Alsbury, T. L. (2008). School Board Member and Superintendent Turnover and the Influence on Student Achievement: An Application of the Dissatisfaction Theory. *Leadership and Policy in Schools*, 7(2), 202–229.
- Andrews, M., Duncombe, W., & Yinger, J. (2002). Revisiting economies of size in American education: are we any closer to a consensus? *Economics of Education Review*, 21(3), 245–262.
- Bădin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3), 818–833.
- Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2), 633–640.
- Ball, S. (2011). Evaluating Educational Programs. *ETS Scientific and Policy Contributions Series, ETS SPC–11*(April).
- Barnum, D., Coupet, J., Gleason, J., McWilliams, A., & Parhankangas, A. (2017). Impact of input substitution and output transformation on data envelopment analysis decisions. *Applied Economics*, 49(15), 1543–1556.
- Bartusek, L. (2000). IASB's Lighthouse Study: School boards and student achievement. *Iowa School Board Compass*, 5(2), 1–12.
- Bessent, a. M., & Bessent, E. W. (1980). Determining the Comparative Efficiency of Schools through Data Envelopment Analysis. *Educational Administration Quarterly*, 16(2), 57–75.
- Bidwell, C. E., & Kasarda, J. D. (1975). School District Organization and Student Achievement. *American Sociological Review*, 40(1), 55–70.
- Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does Management Matter in schools? *The Economic Journal*, 125(584), 647–674.
- Bradshaw, P., Murray, V., & Wolpin, J. (1992). Do Nonprofit Boards Make a Difference? An Exploration of the Relationships Among Board Structure, Process, and Effectiveness. *Nonprofit and Voluntary Sector Quarterly*, 21(3), 227–249.
- Burns, T., & Köster, F. (2016). *Governing Education in a Complex World*. OECD Publishing.
- Caenepeel, L. (1988). *Scholen besturen: samenstelling, structuur en functioneren van inrichtende machten van katholieke scholen*. Acco.
- Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1), 1–25.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Chen, Y., Cook, W. D., Du, J., Hu, H., & Zhu, J. (2017). Bounded and discrete data and Likert scales in data envelopment analysis: application to regional energy efficiency in China. *Annals of Operations Research*, 347–366.
- Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. Van. (2007). An introduction to “benefit of the doubt” composite indicators. *Social Indicators Research*, 82(1), 111–145.
- Cordero, J. M., Alonso-Morán, E., Nuño-Solinis, R., Orueta, J. F., & Arce, R. S. (2015). Efficiency assessment of primary care providers: A conditional nonparametric approach. *European Journal of Operational Research*, 240(1), 235–244.
- Cordero, J. M., Salinas-Jiménez, J., & Salinas-Jiménez, M. M. (2017). Exploring factors affecting the level of happiness across countries: A conditional robust nonparametric frontier analysis. *European Journal of Operational Research*, 256(2), 663–672.
- Cornforth, C. (2001). What Makes Boards Effective? An examination of the relationships between board inputs, structures, processes and effectiveness in non-profit organisations. *Corporate Governance: An International Review*, 9(3), 217–227.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24(1), 93–121.
- Daraio, C., & Simar, L. (2006). A robust nonparametric approach to evaluate and explain the performance of mutual funds. *European Journal of Operational Research*, 175(1), 516–542.
- Daraio, C., & Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, 28(1), 13–32.
- Daraio, C., Simar, L., & Wilson, P. W. (2017). Central limit theorems for conditional efficiency measures and tests of the “separability” condition in nonparametric, two-stage models of production. *The Econometrics Journal*, forthcoming.
- De Witte, K., & Geys, B. (2013). Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. *European Journal of Operational Research*, 224(3), 592–602.

- De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17), 2401–2412.
- De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363.
- Delagardelle, M. L. (2008). The lighthouse inquiry: Examining the role of school board leadership in the improvement of student achievement. *The Future of School Board Governance: Relevancy and Revelation*, 191–224.
- Deprins, D., Simar, L., & Tulkens, H. (1984). *Measuring labor inefficiency in post offices. The Performance of Public Enterprises: Concepts and measurements. M. Marchand, P. Pestieau and H. Tulkens (eds.)*. North-Holland, Amsterdam.
- Devos, G. (2008). Professionalisering van schoolbesturen: hefboomen voor een sterke bestuurskracht. In *Personeel en organisatie. Schoolleiding en begeleiding 2* (Vol. 19, pp. 1–22). Plantyn.
- Devos, G., Verhoeven, J. C., Beuselink, I., den Broeck, H., Vandenberghe, R., Van den Broeck, H., & Vandenberghe, R. (1999). *De rol van schoolbesturen in het schoolmanagement*. Garant.
- Di Liberto, A., Schivardi, F., & Sulis, G. (2015). Managerial practices and student performance. *Economic Policy*, 30(84), 683–728.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245–259.
- Färe, R., Grosskopf, S., & Weber, W. L. (1989). Measuring school district performance. *Public Finance Review*, 17(4), 409–428.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.
- Ford, M. R., & Ihrke, D. M. (2016). Do School Board Governance Best Practices Improve District Performance? Testing the Key Work of School Boards in Wisconsin. *International Journal of Public Administration*, 34(2), 87–94.
- Fusco, E., Vidoli, F., & Sahoo, B. K. (2017). Spatial heterogeneity in composite indicator: A methodological proposal. *Omega*, 1–14.
- García-Romero, A., Santín, D., & Sicilia, G. (2016). Another brick in the wall: a new ranking of academic journals in Economics using FDH. *Scientometrics*, 107(1), 91–101.
- Gemberling, K. W., Smith, C. W., & Villani, J. S. (2000). *The key work of school boards guidebook*. National School Boards Assoc.
- Gritz, R. M., & Theobald, N. D. (1996). The Effects of School District Spending Priorities on Length of Stay in Teaching. *Journal of Human Resources*, 31(3), 477–512.
- Groenez, S., Juchtmans, G., Smet, M., & Stevens, C. (2015). Analyse van het nieuwe financieringsmechanisme voor de werkmiddelen van scholen. Evaluatie van het financieringsdecreet van 2008: Eindrapport.
- Hall, P., Racine, J., & Li, Q. (2004). Cross-Validation and the Estimation of Conditional Probability Densities. *Journal of the American Statistical Association*, 99(468), 1015–1026.
- Herman, R. D., & Renz, D. O. (1998). Nonprofit organizational effectiveness: Contrasts between especially effective and less effective organizations. *Nonprofit Management and Leadership*, 9(1), 23–38.
- Hoffman, R. H. (1995). Contextual influences on school effectiveness: The role of school boards. *School Effectiveness and School Improvement*, 6(4), 308–331.
- Jackson, S. (2003). Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications. *Journal of Management*, 29(6), 801–830.
- Jarl, M., Fredriksson, A., & Persson, S. (2012). New Public Management in Public Education: a Catalyst for the Professionalization of Swedish School Principals. *Public Administration*, 90(2), 429–444.
- Johnes, G., & Johnes, J. (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Economics of Education Review*, 28(1), 107–113.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273–288.
- Johnson, P. A. (2012). School Board Governance: The Times They Are A-Changin'. *Journal of Cases in Educational Leadership*, 15(2), 83–102.
- Kaufman, R. A., & Herman, J. J. (1991). *Strategic planning in education: Rethinking, restructuring, revitalizing*. Technomic Pub Co.
- Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing Hypotheses in Nonparametric Models of Production. *Journal of Business & Economic Statistics*, 34(3), 435–456.
- Land, D. (2002). Local School Boards Under Review: Their Role and Effectiveness in Relation to Students' Academic Achievement. *Review of Educational Research*, 72(2), 229–278.
- Leach, J., Payne, A., & Chan, S. (2010). The effects of school board consolidation and financing on student performance. *Economics of Education Review*, 29(6), 1034–1046.

- Lee, V. E., & Burkam, D. T. (2003). Dropping Out of High School: The Role of School Organization and Structure. *American Educational Research Journal*, 40(2), 353–393.
- Li, D., Simar, L., & Zelenyuk, V. (2016). Generalized nonparametric smoothing with mixed discrete and continuous data. *Computational Statistics and Data Analysis*, 100, 424–444.
- Li, Q., & Racine, J. (2004). Cross-Validated Local Linear Nonparametric Regression. *Statistica Sinica*, 16, 485–512.
- Li, Q., & Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Li, S. K., & Zhao, L. (2015). The competitiveness and development strategies of provinces in China: a data envelopment analysis approach. *Journal of Productivity Analysis*, 44(3), 293–307.
- Lovell, C. A. K., & Pastor, J. T. (1999). Radial DEA models without inputs or without outputs. *European Journal of Operational Research*, 118(1), 46–51.
- McCarty, T. A., & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. *The Measurement of Productive Efficiency: Techniques and Applications*, 271–287.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34(1), 100–117.
- Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available. *Center for Economic Studies, KU Leuven*, 17.
- Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory*, 3, 248–292.
- OECD. (2008). *Handbook on constructing composite indicators: Methodology and User guide*. Paris.
- Rose, H., & Sonstelie, J. (2010). School board politics, school district size, and the bargaining power of teachers' unions. *Journal of Urban Economics*, 67(3), 438–450.
- Saatcioglu, A., Moore, S., Sargut, G., & Bajaj, A. (2011). The Role of School Board Social Capital in District Governance: Effects on Financial and Academic Outcomes. *Leadership and Policy in Schools*, 10(1), 1–42.
- Shwartz, M., Burgess, J. F., & Zhu, J. (2016). A DEA based composite measure of quality and its associated data uncertainty interval for health care provider profiling and pay-for-performance. *European Journal of Operational Research*, 253(2), 489–502.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.
- Simar, L., & Wilson, P. W. (2011). Two-stage DEA : caveat emptor. *Journal of Productivity Analysis*, 36(2), 205–218.
- Thompson, R. G., Singleton, F. D., Thrall, R. M., & Smith, B. A. (1986). High-Energy Physics Lab in Texas. *Interfaces*, 16(6), 35–49.
- Unnever, J. D., Kerckhoff, A. C., & Robinson, T. J. (2000). District variations in educational resources and student outcomes. *Economics of Education Review*, 19(3), 245–259.
- Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. *Procedia Economics and Finance*, 20(15), 679–686.
- Usdan, M., McCloud, B., Podmostko, M., & Cuban, L. (2001). Leadership for student learning: Restructuring school district leadership. *Institute for Educational Leadership*.
- Vanhoof, J., Deneire, A., & Van Petegem, P. (2012). *Waar zit beleidsvoerend vermogen in (ver) scholen?* Plantyn.
- Vershelde, M., & Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. *European Journal of Operational Research*, 223(1), 214–225.
- Woessmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003. *OECD Education Working Papers*, No. 13.
- Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research*, 245(2), 517–530.

Online Appendix

A1. Internal consistency of the questionnaire.

The in-depth interviews with different stakeholders (board members, board presidents and school principals) indicated four district functions as most decisive with respect to school district organizational effectiveness: Financial policy, HR, pedagogics, and investment & infrastructure. The questionnaire was designed in a way to cover all these functions extensively. Using factor analysis, subsets of these function-related questions were constructed to make sure every set corresponded to only one underlying factor (= district function). As a result, we obtained a multi-item scale for every district function²². We check the internal consistency of every district function using Cronbach's alpha:

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_{i=1}^N S_{Y_i}^2}{S_X^2} \right)$$

with $S_{Y_i}^2$ and S_X^2 representing the sample variance of the total score X and the individual scores Y_i , respectively. The value of α indicates to what degree the observed variation in responses is item-specific (Cronbach, 1951). The value of α is dependent on the number of items, the relation between items, and the number of underlying concepts (Cortina, 1993). Since we applied factor analysis to create sets of items for every district function, the homogeneity condition (only one underlying concept) is satisfied by construction. The chosen measure for each district function depends on the validity of τ -equivalence. Using the results of our factor-analysis, we can compare the factor loadings of different items with respect to the underlying factor. When all loadings are comparable in size, the assumption of τ -equivalence is said to hold. In this case, α can be seen as an indicator of the reliability of the questionnaire, i.e. the likelihood of an evaluated set of questions to result in reliable data. If the above condition cannot be validated by the observed factor loadings, another measure of reliability should be used, ω (McDonald, 1981). Depending on this assumption, either α or ω are calculated. Overall, the measure of consistency (α and ω) take on values between 0.58 (financial policy) and 0.67 (HR). On average, consistency equals 0.64 if we include all district functions used to obtain the effectiveness indicator. Although this level is below the range of [0.70-0.95] recommended in the literature (Nunnally & Bernstein, 1994), the obtained values are nonetheless close to the proposed lower bound. Also, the importance of the value of α (and ω) should not be overstated because of the sensitivity of these indicators with respect to respondent characteristics (Ursachi, Horodnic & Zait, 2015). The crucial insight of the above analysis is the observation that all sets of questions represent strictly one district function. In other words, homogeneity is satisfied within all district functions, which is a prerequisite to obtain composite indicators for all functions, and eventually an indicator measuring overall school district organizational effectiveness.

²² Table A4 in the Appendix displays these four functions, and all matching items with the method of scoring.

A2. Tables

Table A1: Summary statistics population

Variables	N	Mean	SD	Median	Min	Max
Schools per district	1,060	3.326	5.736	2	1	95
FTE per district	1,060	138.5	277.6	48.35	3.200	4443
% Tenured (district)	1,060	0.712	0.0783	0.719	0.192	0.931
GON resources	1,060	0.00985	0.00855	0.00859	0	0.0923
Maternal education	1,060	0.0695	0.0827	0.0459	0	0.833
Organizational structure*	No joint structure: 4.52%		Community: 88.56%		Consolidation: 6.92%	

Note: * Communities of schools are formed by voluntary cooperation among schools. 'Consolidation' corresponds to a structure where one district is in charge of different schools. 'No joint structure' indicates that districts chose not to cooperate and did not become a community.

Table A2: Summary statistics sample

Variables	N	Mean	SD	Median	Min	Max
Schools per district	151	5.742	9.109	3	1	59
FTE per district	151	278.4	485.8	123.9	4.500	3213
% Tenured (district)	151	0.716	0.0564	0.717	0.514	0.848
GON resources	151	0.0108	0.00802	0.0102	0	0.0604
Maternal education	151	0.0625	0.0511	0.0476	0	0.292
Organizational structure	No joint structure: 2.65%		Community: 88.08%		Consolidation: 9.27%	

Table A3: Comparison of sample and population means using administrative data (2012).

Variable	Sample	Population	t-statistic (p-value)
Schools per district	5.742	3.326	3.2587 (0.0014)***
FTE per district	278.4	138.5	3.5373 (0.0005)***
% Tenured (district)	0.716	0.712	0.9047 (0.3671)
GON resources	0.0108	0.00985	1.3982 (0.1641)
Maternal education	0.0625	0.0695	-1.6741 (0.0962)*
Structure	No joint	4.52%	-1.2210 (0.2240)
	Community	88.08%	-0.1816 (0.8561)
	Consolidation	9.27%	0.9930 (0.3223)

*** p<0.01, ** p<0.05, * p<0.1

Table A4: Questionnaire items per district function and corresponding scoring.

District function	Item	Scoring
Financial Policy	How would you rate your school district's financial policy?	1 (totally insufficient) – 10 (perfect)
	Statutory provisions and duties with respect to the district's financial policy are known by all board members.	1 (fully disagree) – 5 (fully agree)
	The degree of support schools receive from their district with respect to the school's financial policy.	1 (totally insufficient) – 5 (more than sufficient)
	The distribution of resources over schools within the district is clear to me	1 (fully disagree) – 5 (fully agree)
HR	How would you rate your school district's HR policy?	1 (totally insufficient) – 10 (perfect)
	The way the school district selects personnel (management, coordinating staff, policy makers).	1 (totally insufficient) – 5 (more than sufficient)
	The way the school district performs follow-up and performance interviews with school personnel	1 (totally insufficient) – 5 (more than sufficient)
	The school district is concerned with the welfare and wellbeing of personnel when innovations are implemented.	1 (fully disagree) – 5 (fully agree)
	The school district is aware of negative changes within its personnel.	1 (fully disagree) – 5 (fully agree)
	The school district is concerned with mentoring teachers and principals on the payroll of its district.	1 (fully disagree) – 5 (fully agree)
Pedagogics	How would you rate your school district's pedagogic policy?	1 (totally insufficient) – 10 (perfect)
	The degree of support schools receive from their district with respect to students' disciplinary files.	1 (totally insufficient) – 5 (more than sufficient)
	Follow up of education quality standards of schools by the district.	1 (not applicable) – 5 (fully applicable)
	School board members are fully aware of ongoing pedagogic projects in their schools.	1 (fully disagree) – 5 (fully agree)
Investment & Infrastructure	How would you rate your school district's procurement policy?	1 (totally insufficient) – 10 (perfect)
	How would you rate your school district's investment policy?	1 (totally insufficient) – 10 (perfect)
	Other schools are involved in the signing of contracts (management and contractors).	1 (not applicable) – 5 (fully applicable)

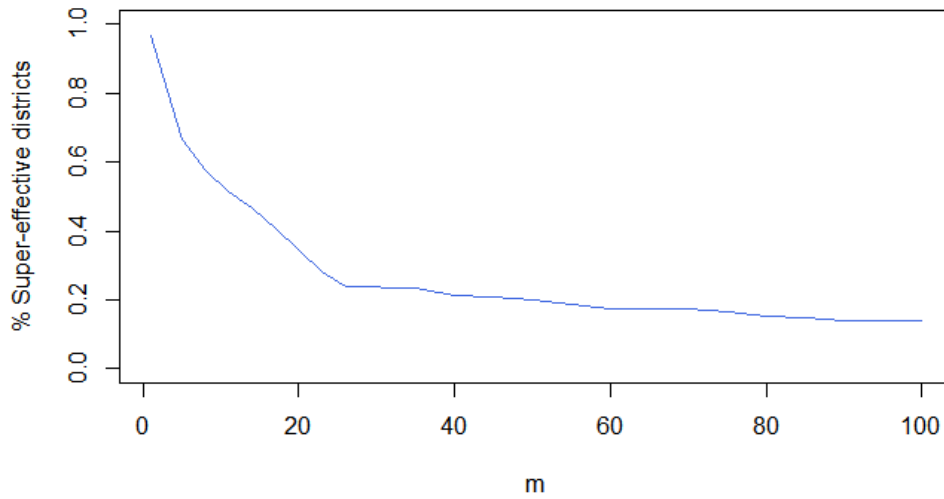
Table A5: Statistical inference of school district organizational effectiveness following the procedure by Li & Racine (2004).

	N=150		Model 1		Model 2		Model 3		Model 4		Model 5	
	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value	Influence	p-value
<i>School district environment</i>												
Maternal education	unfavorable	0.022	unfavorable	0.112	unfavorable	0.102	unfavorable	0.047	unfavorable	0.021		
GON resources	favorable	0.392	unfavorable	0.374	favorable	0.136	favorable	0.517	unfavorable	0.991		
% tenured	unfavorable	0.098	unfavorable	0.047	unfavorable	0.070	unfavorable	0.508	unfavorable	0.399		
<i>Respondent characteristics</i>												
Gender (male=yes)	unfavorable	0.017	unfavorable	0.000	unfavorable	0.012	favorable	0.012	unfavorable	0.000		
Age (<50 years old = yes)	unfavorable	0.035	unfavorable	0.000	unfavorable	0.002	unfavorable	0.191	unfavorable	0.000		
Education: Master's or above	unfavorable	0.108	favorable	0.007	favorable	0.024	favorable	0.003	favorable	0.086		
Board member	unfavorable	0.151	unfavorable	0.094	unfavorable	0.007	unfavorable	0.633	favorable	0.427		
<i>Organization</i>												
More than 6 schools	unfavorable	0.314	unfavorable	0.312	favorable	0.018	favorable	0.004	favorable	0.004		
Private school district			favorable	0.009	favorable	0.060	favorable	0.005	favorable	0.004		
Participative management					favorable	0.607	favorable	0.650	favorable	0.003		
Consolidation							favorable	0.004	favorable	0.167		
No joint							favorable	0.406	favorable	0.570		
Expertise										favorable	0.066	

Note: p-values in bold indicate significance at least at the 10% level.

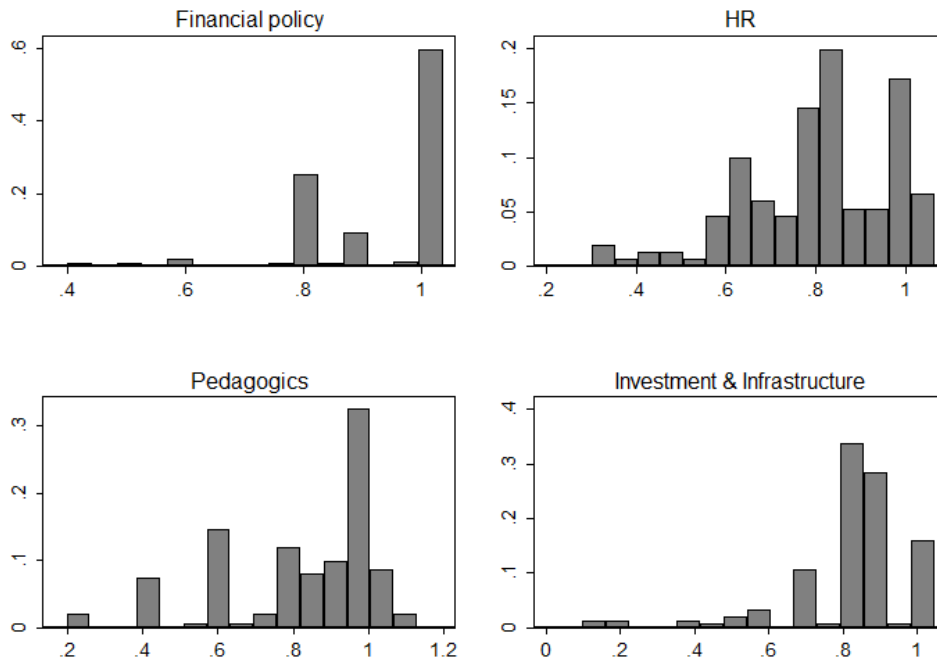
A3: Figures

Figure A1: Marginal decrease in percentage super-effective school districts.



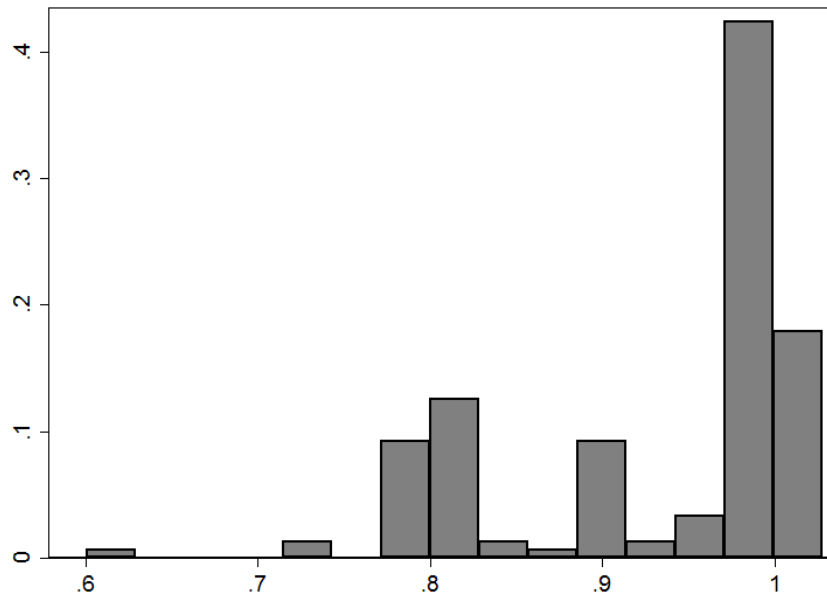
Note: Depending on the choice of m , the proportion of super-effective observations varies since the size of the drawn sample (m) with respect to the total sample size n influences the probability of observation o not belonging to the production frontier. The value of m is set in a way to attain a sufficiently small decrease in the proportion of super-effective school districts (here, $m=40$).

Figure A2. Organizational effectiveness per district function: Unconditional estimates.



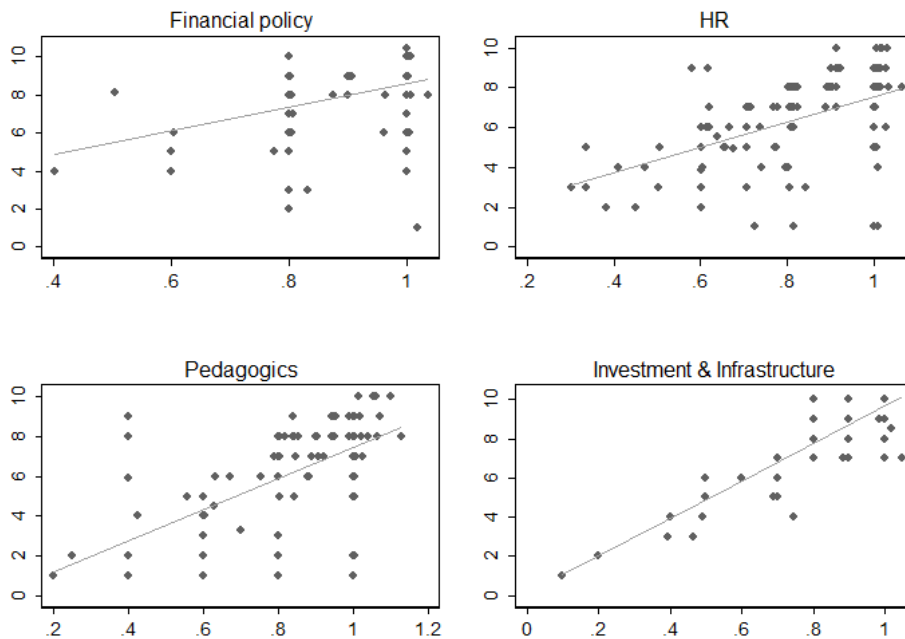
Note: All histograms display the distribution of robust order- m BoD scores. Horizontal axis indicates the composite indicator of scores obtained by aggregating the scores awarded to the items listed in Table A4. Vertical axis indicates the fraction of the observations.

Figure A3. Organizational effectiveness of school districts: Unconditional estimates.



Note: The vertical axis indicates the fraction of observations attaining the effectiveness score on the horizontal axis. Total organizational effectiveness scores are obtained by aggregating composite indicators displayed in Figure A2.

Figure A4. Correlations between single-item scales and Benefit-of-the-Doubt scores.



Note: Spearman and Pearson correlation coefficients attain values between 0.4 and 0.7. Hence, our robust BoD estimates do not perfectly coincide with the single-item scales but there appears to be a strong correlation between both measures. This finding reflects the advantage of our endogenous weighting approach: allowing more variables into our evaluation (without fixing weights ex ante) broadens our approach towards measuring organizational effectiveness, while the obtained scores are still in line with respondents' perceptions.

