5-2014

Assessing Efficiency of Schools Participating in Startsmart K3+

Yamil Vargas Hedeman
Utah State University

Recommended Citation

Hedeman, Yamil Vargas, "Assessing Efficiency of Schools Participating in Startsmart K3+" (2014). *All Graduate Plan B and other Reports*. 371.
https://digitalcommons.usu.edu/gradreports/371

ASSESSING EFFICIENCY OF SCHOOLS PARTICIPATING IN STARTSMART K3+

by

Yamil Vargas Hedeman

A research paper submitted in the partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Economics

Approved:

Man-Keun Kim Major Professor	Reza Oladi Committee Member
Ryan Bosworth Committee Member	Damon Cann Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

Assessing Efficiency of Schools Participating in StartSmart K3+

by

Yamil Vargas Hedeman, Master of Science

Utah State University, 2014

Major Professor: Dr. Man-Keun Kim
Department: Applied Economics

New Mexico has administered and funded State K3+ program to reduce the achievement gap between students in kindergarten through third grade since 2007. StartSmart K3+ project is an experimental research to examine the cost-effectiveness of State K3+. This research attempts to measure the efficiency of the schools participating in StartSmart using valuable information and data collected by StartSmart K3+. A Data Envelopment Analysis (DEA) originally developed to study production efficiency of micro-level organizations, and a regression model are used to analyze the efficiency of the schools participating in the first year of the project in 2011.

The DEA is used to measure each school's inputs and outputs ratio, such as teachers' qualification and students' performance, compares them and calculate the efficiency score. Efficiency scores generated by the DEA are biased by construction since the DEA constructs a lower bound on the true efficient frontier. Efficiency scores from the DEA are corrected using the bootstrap procedure as suggested in Simar and Wilson (1998, 2000). After generating DEA scores and correcting the bias, a regression model is used to identify the environmental factors that school may not control and affect schools' performance. Two-limit Tobit with limits at zero and unity is used to estimate equations.

Three performance measurements are identified as outputs: 1) average scores in reading, writing, math and vocabulary (each score is considered as one output and thus there are four outputs in total), 2) minimum scores in four subjects (four outputs), and 3) percentage of students with scores above 90 points in each subject tests based on the Woodcock-Johnson III classification (four outputs). Results suggest that between 50% and 58% of the schools were efficient in 2011, depending on the students' performance measurements considered. Three Tobit regression models for three different types of outputs are estimated. Dependent variables are bias-corrected DEA scores and explanatory variables are education level of the closest caregiver, poverty rate in the school district and other variables. Results from the regression model tell us that education level of the closest caregiver is an important factor in explaining school efficiency. The time students spend watching television and playing non-education video games has a high impact in changes in school efficiency too.

Schools in areas with high-risk populations will require a greater share of resources to provide the same quality of education enjoyed in more affluent areas. The goal pursue by the Government of New Mexico of reducing the existing achievement gap between students will be limited by these inefficiencies. An efficiency evaluation could be carried out at the end of each summer session to identify inefficient schools and to better allocate resources. Short and long run policies should be implemented to increase schools' efficiency.

PUBLIC ABSTRACT

Assessing Efficiency of Schools Participating in StartSmart K3+

Yamil Vargas Hedeman

New Mexico has administered and funded State K3+ program to reduce the achievement gap between students in kindergarten through third grade since 2007. StartSmart K3+ project is an experimental research to examine the cost-effectiveness of State K3+. This research attempts to measure the efficiency of the schools participating in StartSmart K3+ using data collected by StartSmart K3+. A Data Envelopment Analysis, which measures the efficiency of decision-making units such as school, is used to measure the efficiency of the schools participating in the first year of the project in 2011. A regression model is used to investigate the effect of environmental variables such as education level of the closes caregivers and poverty rates in the school districts. Results suggest that between 50% and 58% of the schools were efficient in 2011, depending on the outcome measurement considered. Results from the regression model tell us that education level of the closest caregiver is an important factor in explaining school efficiency. The time students spend watching television and playing non-education video games has a high impact in changes in school efficiency too.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background

The State of New Mexico has administered and funded State K3+ program since 2007. State K3+ provides under-achieving students and high-need students in kindergarten through third grade a minimum of 25-additional days of instructions during summer. A primary goal of State K3+ is to reduce the existing achievement gap between students. Nevertheless, not all districts and schools in New Mexico participate in the program. One of the reasons could be that State K3+ is voluntary for districts, schools and teachers. Likewise, the program is not mandatory for students. Information about the effectiveness of the program has been of interest of New Mexico’s decision makers. To investigate the effectiveness of the State K3+, researchers in Utah State University proposed a project called StartSmart, which is an experimental research started in 2011.

In the first year of implementation, in 2011, StartSmart randomly selected a control and intervention group of students distributed in 26 schools over the school districts in New Mexico. A sample of 396 kindergarten students received 25 additional days of instruction during that summer. The additional instruction given to students in StartSmart is identical to the one received by students in State K3+. Data regarding students, educational resources, information about teachers, and performance measurements are collected before, during, and after the summer by StartSmart researchers.

The goal of a cost-effectiveness analysis, like the one implemented by StartSmart, is to assess a program as whole. Nevertheless, the effect size of the program in terms of difference in performance between control and intervention group will be affected by the efficiency of the schools participating in the program. For this reason, answering the question of which schools are relatively efficient to achieve the goal of the program is crucial to improve the effect size of the experiment. In addition, it is important to identify what factors cause these differences and variations among schools.

1.2. Research Goals

In general, the cost-effectiveness of State K3+ program depends on the efficiency of each school in using its resources in a way that can get the best possible outcome. It is important to provide policy makers relevant information regarding the efficiency of schools to redesign State K3+ program in the future. Therefore, the purpose of this research is measuring efficiency of schools participating in StartSmart program. A set of inputs that can be controlled by the school will be used to generate efficiency scores based on test scores that is the measurement of students' performance. In addition it is important to investigate which factors that cannot be controlled by schools are affecting schools' efficiency scores. Results would provide relevant information for decision makers to improve State K3+ program.

To achieve these research goals, a Data Envelopment Analysis (DEA) is used to measure the efficiency of each school. The DEA is a non-parametric approach to measure efficiency of decision making units (DMU) such as schools, hospitals and firms. Basically, the DEA measures each school's input-outputs ratio, compares them, and calculates the efficiency score. Tobit regression model is used to investigate the effects of environmental factors on the efficiency scores identified by the DEA. DEA efficiency scores are the dependent variable in the regression model

1.3. Organization of Research

Chapter 1 introduces research questions. Chapter 2 reviews previous studies regarding effects of summer school programs and efficiency analyses. The DEA and Tobit models are discussed in Chapter 3. Chapter 4 discusses data to use and chapter 5 presents the empirical results. Chapter 6 concludes the study.

2. LITERATURE REVIEW

2.1. Summer Learning and Students Performance

Summer learning programs have emerged as a promising way to address the growing achievement gap between children of the poor families and those of the affluent (Augustine et al., 2013). Several studies have found significant effects of participation in summer school in students' performance (Matsudaira, 2008; Zvoch and Stevens, 2013; Kim and Quinn, 2013). Nevertheless, these studies focus on the effectiveness of a summer program as a whole, ignoring the individual units, i.e., schools that carry out the program.

Matsudaira (2008) uses data from a large school district to measure the effect of summer school attendance on students' achievement. He finds that summer school has a positive impact on scores in math and reading. Also, he suggests that that summer school may be an exceptionally cost-effective way to raise student achievement.

Zvoch and Stevens (2013) measure the effect of participation in summer school for students struggling in reading. They find that kindergarten students who participated in summer school outperformed students who didn't. The performance gap that emerged in literacy scores at the start of the new academic year is an indicator of the potential that summer instruction holds for those who participate in a school-based supplemental support program. This suggests that targeted summer instruction can be a useful strategy to support student learning over the summer months.

Kim and Quinn (2013) conduct a meta-analysis and review various studies on summer reading programs in the U.S. and Canada from 1998 to 2011. They indicate that income-based disparities in measurable aspects of children's home literacy environments may contribute to disparities in reading achievement. They find that summer reading programs had significant benefits for children from low-income family. Therefore, in the absence of an effective summer reading intervention, low-income children may have limited opportunities to practice reading connected text with speed and accuracy and to acquire conceptual and background knowledge.

Similarly, there is accumulating research evidence that teachers' credentials, experience, and years of education may make a difference in children's achievement (Buddin and Zamarro, 2009; Boonena et al. 2013). Buddin and Zamarro (2009) argue that teacher quality is a key element of student academic success, but few specific teacher characteristics influence classroom outcomes. Based on longitudinal student-level data from Los Angeles, they find that students' achievement is unaffected by whether classroom teachers have advanced degrees. However, their results show that student achievement slightly increases with teacher's experience.

Similar findings are found in Boonen et al. (2013). They investigate the effects of teachers' background qualifications, attitudes and beliefs, and instructional practices on students' achievement in math, reading, and spelling in first grade. They find that students with more experienced teachers tended to perform better, whereas students with teachers doing in-service training tended to perform worse. Overall, their results suggest that teachers had a modest to strong effect on student achievement in first grade.

Nevertheless, as noted by Cohen et al. (2003), providing resources, such as highly qualified teachers, is important but will not necessarily assure effective use of these resources. This is one of the reasons why efficiency of public schools has been an important research topic.

2.2. DEA and School Efficiency

The DEA model has been applied to measure the relative efficiency of public schools. Bessent et al. (1982) uses the DEA to measure the efficiency of the Houston Independent School District. More recently, studies applying the DEA model to assess schools efficiency can be found in the literature such as the one developed by Adkins and Moomaw (2007), Rassouli (2011) and Raposo et al. (2011).

In a study based on Oklahoma school districts, Adkins and Moomaw (2007) conclude that additional instructional and non-instruction expenditures improve student performance, but only by a small amount. In addition, they found that school district size, teacher education and experience, and teacher salary affect the technical efficiency of school districts. In the same state of Oklahoma, Rassouli (2011) conduct a study in which the decision units were schools. Given that the efficiency estimates from the first stage are between 0 and 1, data is censored, and so Tobit regression, rather than OLS, is the appropriate method of estimation. They suggest that inefficiencies in schools could be due to exogenous factors such as poverty or increased immigration. For this reason, efficiencies generated by the DEA were then used as dependent variables in a second stage with Tobit regression to assess the effects of variables not included in the first stage on technical efficiency.

Similarly Raposo et al. (2011) follow a two-stage approach with data from public schools from the Northeast Region of Brazil. The educational efficiency was determined only by the variables directly controlled by the school. They found that efficiency scores become more homogeneous as compared to the rank produced from a simple one-stage DEA , after isolating from the effect of environmental variables, such as student's socioeconomic status, that might influence efficiency as well. Raposo et al. (2011) run the regression model in the first stage to control for the effects of the environmental variables. The error terms generated in the first stage are used as the output variable in the DEA model.

More complex application of the DEA can be found in the literature. This is the case of a semi-parametric analysis with non-discretionary inputs used by Afonso and Aubyn (2006) to measure cross-country efficiency. As well as other studies like the one developed by Simar and Wilson (2007) in which they introduce useful tools when using a DEA model to measure schools efficiency. In their work, "Estimation and inference in two-stage semi-parametric models of production processes", Simar and Wilson (2007) propose a modification to isolate the environment factors that can affect the outcomes. They demonstrate that while conventional inference methods are inconsistent in the second-stage regression, consistent inference is both possible and feasible.

3. METHODOLOGIES

The DEA model was developed by Charnes, Cooper and Rhodes (1978), and in homage to them it is also known as the CCR model. Since then, the DEA method has been applied to various areas, such as production engineering, management and economics. In economics alone the applications include themes such as the efficiency of agriculture production, public spending, health services, energy sector and education (Raposo et al., 2011).

According to Cooper, Seiford and Zhu (2004) the empirical orientation and the absence of a need for the numerous a priori assumptions that accompany other approaches (such as regression analysis) have resulted in the use of the DEA in a variety of studies. They also state that because it requires very few assumptions, the DEA has also opened up possibilities for use in cases which have been resistant to other approaches. Basically, this is because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in Decision Making Units (DMU).

3.1. Data Envelopment Analysis (DEA)

The DEA has been developed in the management science tradition with a focus on computing the relative efficiency of different DMUs, for example, firms, schools, hospitals or counties. To define DEA efficiency estimates the following notation is established; let $\mathbf{x}_{\mathrm{j}} \in \mathrm{R}_{+}^{\mathrm{p}}$ denote a vector of p inputs and $\mathbf{y}_{\mathrm{j}} \in \mathrm{R}_{+}^{\mathrm{q}}$ denote a vector of q outputs for DMU j, where j = 1, … , n. The production possibility set is defined by **P** = {(x, y)| outputs **y** can be produced from inputs **x**}. The boundary of **P** is referred to as the production frontier.

Technically inefficient DMUs operate at points that are inferior to the production frontier, while technically efficient DMUs operate somewhere along the frontier. Define an efficiency measure θ for DMU j, $(\mathrm{x}_{\mathrm{j}}, \mathrm{y}_{\mathrm{j}}) \in \mathrm{R}_{+}^{\mathrm{p+q}}$ such that

$$(1) \qquad \theta_j \equiv \sup\{\boldsymbol{\theta} | \left(\mathbf{x}_j, \boldsymbol{\theta}\mathbf{y}_j\right) \in P, \boldsymbol{\theta} > 0\}$$

This is the Farrell (1957) measure of output technical efficiency, which is the reciprocal of the Shephard (1970) output distance function. The DEA estimator θ defined in equation (1) at a specific point (DMU j) can be written in terms of the linear programming (LP) model which is initially proposed by Charnes, Cooper and Rhodes (1978, 1981) and extended by Banker, Charnes, and Cooper (1984),

$$(2) \qquad \hat{\theta}_j = \max\{\boldsymbol{\theta} > \mathbf{0} \mid \boldsymbol{\theta}\mathbf{y}_j \leq \mathbf{Y}\boldsymbol{\lambda}, \mathbf{x}_j \geq \mathbf{X}\boldsymbol{\lambda}, \boldsymbol{\lambda} \in \mathrm{R}_+^n\}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ and $\boldsymbol{\lambda}$ is $n \times 1$ intensity variables. It is noteworthy that the DEA formulation differs slightly along with the assumption of returns to scale.

Under the constant returns to scale (CRS), the LP formulation is given by equation (2) which is called the CCR model (Charnes, Cooper, and Rhodes, 1978). The DEA estimator under the assumption of variable returns to scale (VRS) is found by solving the same LP problem in (2) with additional constraint, $\mathbf{i}'\boldsymbol{\lambda} = 1$, where $\mathbf{i}$ denotes an $n \times 1$ vector of ones. This model is called BCC model (Banker, Charnes, and Cooper, 1984) after authors' names. The additional constraint imposes a convexity condition on allowable ways in which the observations for the n DMUs may be combined (Cooper, Seiford and Tone, 2007). When the above constraint is replaced by $\mathbf{i}'\boldsymbol{\lambda} \leq 1$, the production set exhibits the non-increasing returns to scale (NIRS). Various RTS assumptions are explained in Figure 1 that measures five DMUs' efficiencies. DMUs B and C are efficient and DMUs A, D, and E are inefficient under the CRS assumption. DMU E becomes efficient under the assumption of NIRS and DMUs A and E are efficient under VRS assumption.

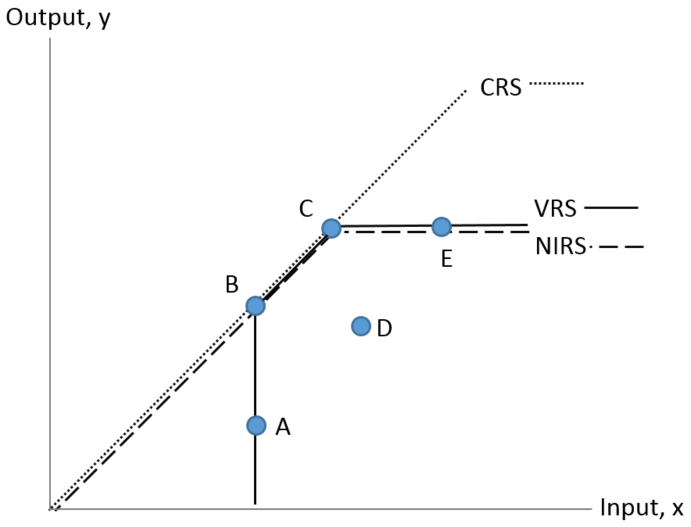

Figure 1. Efficiency and Return to Scale Assumption

The LP models in equation (2) along with additional constraint are run n times to identify the relative efficiencies of all the DMUs. The DEA efficiency estimates are less than equal to 1 by construction. The DMU is said to be efficient if it obtains the DEA estimate of 1. The DEA estimate of less than 1 implies that it is inefficient. Also, $\theta_j^{crs} \leq \theta_j^{nirs} \leq \theta_j^{vrs}$ by construction (See Figure 1).

The existence of increasing or decreasing returns to scale is of importance to many policy decisions as shown in Figure 1. Banker (1996) provides the test statistics of return to scale assuming that the DEA efficiency estimator follows specific distributions such as the exponential distribution or the half-normally distribution (chi-square distribution with degrees of freedom 1). Simar and Wilson (2002) point out that there is no reason to assume a specific distribution for the test and propose a bootstrap procedure avoiding the ad hoc assumptions of Banker (1996).

The statistical test for the returns to scale begins with CRS (Simar and Wilson, 2002). The null hypothesis is the production set exhibits CRS and the alternative hypothesis is that it shows VRS. Various test statistics are possible; however, the mean of ratios $\hat{\theta}_j^{crs} / \hat{\theta}_j^{vrs}$, that is $t_{crs} = n^{-1} \sum_{j=1}^{n} \hat{\theta}_j^{crs} / \hat{\theta}_j^{vrs}$ will be used as in Simar and Wilson (2002). By construction $t_{crs} \leq 1$, the null hypothesis is rejected when t_{crs} is significantly less than 1. The critical value for deciding if the test statistic is significantly less than 1 can be derived from bootstrapping (Simar and Wilson, 2002). For more information about bootstrapping refer to Simar and Wilson (1998, 2000). When the null hypothesis of CRS is rejected, another test is performed with a less restrictive, NIRS versus VRS. The test statistic is similar and decision is made based on the critical value from the bootstrapping.

Related to further statistical analysis with DEA efficiency estimates, for example regression or causal relationship investigation, Simar and Wilson (2007) insist that the statistical analyses may not be consistent unless the DEA estimates are corrected. They show the inconsistency using Monde Carlo experiment, especially the second-stage regression. According to Simar and Wilson (2007) this inconsistency existed because the DEA estimates are biased downward by construction since the DEA constructs a lower bound on the true efficient frontier. In addition, $\hat{\theta}_j$ are serially correlated since a DMU is either efficient or it is related to at least another two DMUs placed on the efficient frontier.

Simar and Wilson (2007) propose bootstrap procedures to improve statistical properties of DEA estimates such that $\widehat{\widehat{\theta}}_j = \hat{\theta}_j - \text{bias}\ (\hat{\theta}_j)$. The bias term is constructed using the bootstrap. The empirical DEA estimates and bias corrected DEA estimates are reported in the following section.

3.2. Second Stage Regression Model

Most of the researches about efficiency have used a two-stage approach. Efficiency is estimated in the first stage and then the estimated efficiencies are regressed on covariates, typically different from those used in the first stage, that are viewed as representing environmental variables. In short, we have the regression model as follows

$$(3) \qquad \hat{\hat{\theta}}_j = \mathbf{z}_j\boldsymbol{\beta} + \varepsilon_j$$

where $\hat{\hat{\theta}}_j$ is the bias corrected DEA score from section 3.1. and $\mathbf{z}_j$ is the vector of the environmental variables.

Most of empirical researches have estimated above equation by assuming a censored Tobit specification. Tobit model is appropriate model because the DEA efficiency scores, the dependent variable, above 1 and below zero is not observed. Especially, values above 1 are all transformed to or reported as a single value of 1. Mathematically, if $\hat{\hat{\theta}}_j^* \leq 0$, the efficiency score for the jth DMU, $\hat{\hat{\theta}}_j = 0$, if if $\hat{\hat{\theta}}_j^* \geq 1$, $\hat{\hat{\theta}}_j = 1$, and if $0 < \hat{\hat{\theta}}_j^* < 1$, $\hat{\hat{\theta}}_j = \hat{\hat{\theta}}_j^*$, where $\hat{\hat{\theta}}_j^*$ is the real efficiency score.

Given equation (3), the likelihood for a sample can be written:

$$(4)\qquad \mathrm{L} = \prod_{\hat{\hat{\theta}}=0} \mathrm{prob}\left(\hat{\hat{\theta}} = 0\right) \prod_{\hat{\hat{\theta}}=1} \mathrm{prob}\left(\hat{\hat{\theta}} = 1\right) \prod_{0<\hat{\hat{\theta}}<1} \mathrm{f}\left(\hat{\hat{\theta}}^{*}\right)$$

where $\mathrm{f}\left(\hat{\hat{\theta}}^{*}\right)$ is the density function of $\hat{\hat{\theta}}^{*}$, i.e., in this case, the normal density function. If there are no observations = 0 or 1 like bias-corrected DEA scores, then the first two terms in equation (4) will not appear in the likelihood function and maximum likelihood estimation (MLE) are obtained by maximizing the third term alone which is the OLS estimator. Marginal effects in the censored regression model is given by (Greene, 2000, p. 909):

$$(5)\qquad \frac{\partial \mathrm{E}[\hat{\hat{\theta}}|\mathbf{z}_{\mathrm{j}}]}{\partial \mathbf{z}_{\mathrm{j}}} = \boldsymbol{\beta} \times \mathrm{prob}(0 < \hat{\hat{\theta}}^{*} < 1)$$

According to Simar and Wilson (2007), Tobit regression is not appropriate in the second stage analysis, especially for corrected DEA scores, because “no coherent account of how the censoring arises has been offered”. In addition, Simar and Wilson (2007) show that $\hat{\theta}_{\mathrm{j}}$ is serially correlated by construction since a DMU is either efficient or it is related to at least another two DMUs placed on the efficient frontier. In other words, the correlation arises “…in finite samples from the fact that perturbation of observations lying on the estimate frontier will… cause changes in efficiencies estimated for other observations” (Simar and Wilson, 2007, p33). Simar and Wilson (2007) suggest a truncated regression model instead.

However, some empirical studies that eliminate inconsistency bias in the efficiency scores estimate the second-stage regression model using ordinary least squares (OLS). A Tobit model is used in this analysis for consistency with previous literature but no censored data is generated and similar results are obtained when using OLS.

4. DATA

The data used in this research are compiled from the StartSmart database, specifically information provided by the twenty-six schools participating in the program in summer 2011. School inputs are i) days attended (maximum 25 days), ii) number of students per teacher (student teacher ratio), iii) teachers' experiences, and iv) teachers' qualification. Teachers' qualification is measured as the ratio of teachers with a master degree. Education expenditure is not considered as an input because it is assumed to be constant, $800 per student, over schools. Basic descriptive statistics of inputs are listed in Table 1.

The output is measured by the test scores in reading, math, writing, and vocabulary that students took in fall 2011 after participating in the StartSmart program in the summer. The tests scores are taken from various subtests in the Woodcock-Johnson test: broad reading, broad math, basic writing and picture vocabulary. Tests scores in four subjects, which are the types of scholastic skills people are generally interested in developing, are appropriate measurement of students' performance for kindergarteners and lower graders. The sample includes the results of the 396 students enrolled in the program. Table 2 contains basic statistics of output measurements.

Table 1. Descriptive statistics of input variables

Variable	Observations	Mean	Std. Dev.	Min	Max
Days attended[1] (days)	26	21.42	3.39	11.89	24.65
Students per teacher[2] (students)	26	12.81	3.53	6.00	21.00
Teacher experience[3] (years)	26	4.71	3.18	1.00	11.00
Teacher qualification[4] (zero-or-one)	26	0.38	0.50	0.00	1.00

Source: StartSmart database.

[1] average number of day students attended school during the 25 instructional days

[2] average number of students per teacher

[3] average years of experience

[4] ratio of teachers with a master degree

Table 2. Descriptive statistics of output variables

Variable	Observations	Mean	Std. Dev.	Min	Max

Average scores	Reading	26	86.55	6.64	74.06	99.09
	Math	26	87.43	9.88	66.48	101.00
	Writing	26	91.23	6.38	75.43	102.82
	Vocabulary	26	95.45	5.38	85.41	105.88
Minimum scores	Reading	26	64.62	16.01	26	90
	Math	26	59.69	18.06	18	91
	Writing	26	65.46	20.19	5	93
	Vocabulary	26	67.92	18.44	7	91
Percentage of students above average	Reading	26	47.28	20.16	0	100
	Math	26	57.60	19.62	14.28	100
	Writing	26	62.73	19.45	14.28	100
	Vocabulary	26	70.50	16.13	39.39	100

Source: StartSmart database.

The percentage of students above standard score average is calculated based on the Woodcock-Johnson III classification presented in Table 3. The classification of standard score and percentile rank ranges provides a guideline for describing a student's relative standing among age- or grade- peers (Mather and Woodcock, 2001). Percentile ranks describe student's relative standing in a comparison group on a scale of 1 to 99. The student's percentile rank indicates the percentage of students from the comparison group who had scores the same or lower than the student's scores. The third column in Table 3 provides a set of verbal labels for the score ranges.

Table 3. Woodcock-Johnson III classification of Standard Score

Standard Score Range	Percentile Rank Range	Classification
131 and above	98 to 99.9	Very Superior
121 to 130	92 to 97	Superior
111 to 120	76 to 91	High Average
90 to 110	25 to 75	Average
80 to 89	9 to 24	Low Average
70 to 79	3 to 8	Low
69 and below	0.1 to 2	Very Low

Source: Woodcock-Johnson III Examiner's Manual.

Part of the data used in the second stage is taken from the household survey collected by StartSmart in spring 2011. Information about the closest caregiver's education level and hours per day children spend watching TV or playing non-educational video or computer games (VGTV) is extracted from this source of data.

The caregiver's education levels are coded from 1 (Kindergarten through Fifths grade) to 9 (one year grad school or more). The average of the education level of the caregiver by school is then used as the education level variable. The mean of the education level of the caregiver is 4.81 meaning that in average parents have at least a high school diploma (Table 4). Table 4 also shows that students spent roughly 2 hours and half per day watching TV or playing non-educational video or computer games.

Poverty rates in the school district are taken from the New Mexico Public Education Department. Household income is the median household income in the city where the school is located (seven cities) taken from the U.S Census Bureau. The location of the school was found using the Common Core of Data (CCD) from the National Center for Education Statistics. Basic descriptive statistics from all four variables are presented in Table 4.

Table 4. Descriptive statistics of variables used in regression model

Variable	Observations	Mean	Std. Dev.	Min	Max
Education[1]	26	4.81	0.71	3.27	5.93
Poverty[2]	26	26.43	10.16	18.9	45.61
Income[3]	26	41,918	7,835	25,990	48,432
VGTV[4]	26	2.57	0.29	1.89	3.00

Source: StartSmart database, NM Public Education Department, U.S Census Bureau

[1] Closest caregiver education level (average by school); Coded as 1 for the lowest level and 9 for the highest

[2] School district poverty rate (percentage)

[3] Median household income in the city where the school is located (dollars)

[4] Time per day a child watch television or play non-educational video games (average by school)

5. EMPIRICAL RESULTS

5.1. School Efficiency Scores

Three different outputs, i.e., three different measurements of students' performance, are generated such as i) average score in each reading, math, writing and vocabulary (Model 1 – each score is considered as one output and thus there are four outputs in total), ii) minimum scores in reading, math, writing and vocabulary (Model 2 – four outputs), and iii) percentage of students with scores above 90 points based on the Woodcock-Johnson III Standard Score average in reading, math, writing and vocabulary (Model 3 – four outputs). First of all, returns to scales are tested using the way proposed by Simar and Wilson (2002). To complete this task, the package FEAR developed by Wilson (2008) is used in the software R. Test results show that, with a 95% level of confidence, average scores (Model 1) and percentage of students with scores above average (Model 3) exhibit non-increasing return to scale (NIRS) (See Figure 1). Minimum scores exhibit variable return to scale (VRS).

All of DEA efficiency scores generated for Models 1, 2 and 3 are represented in Table 5. Thirteen out of 26 schools were relatively efficient in 2011, i.e., the efficiency score is equal to one in Model 1. Efficiency scores from Model 2 suggest that 58% of schools participating in the program in 2011 were efficient. Likewise, efficiency scores from Model 3 report that 58% of the schools were relatively efficient (Table 5). Table 5 also contains bias corrected DEA estimates.

Table 5. School Efficiency Scores

DMU	Model 1	Model 2	Model 3

	NIRS		VRS		NIRS	
	DEA Estimates	Corrected DEA Estimates	DEA Estimates	Corrected DEA Estimates	DEA Estimates	Corrected DEA Estimates
1	0.982	0.963	0.918	0.894	1.000	0.943
2	0.878	0.863	0.958	0.938	0.802	0.781
3	1.000	0.974	1.000	0.951	1.000	0.945
4	0.935	0.919	1.000	0.944	0.909	0.888
5	0.954	0.943	0.913	0.890	0.957	0.933
6	1.000	0.981	1.000	0.973	0.967	0.943
7	0.946	0.928	0.911	0.889	0.982	0.956
8	0.964	0.950	0.993	0.968	0.986	0.961
9	1.000	0.966	1.000	0.939	1.000	0.936
10	1.000	0.972	1.000	0.951	1.000	0.941
11	1.000	0.977	1.000	0.956	1.000	0.955
12	0.976	0.961	0.935	0.913	0.920	0.897
13	1.000	0.981	1.000	0.956	1.000	0.954
14	1.000	0.964	1.000	0.943	1.000	0.939
15	1.000	0.967	1.000	0.953	1.000	0.943
16	0.941	0.929	0.929	0.908	0.899	0.875
17	1.000	0.963	1.000	0.942	1.000	0.946
18	1.000	0.977	0.906	0.882	1.000	0.944
19	0.967	0.955	1.000	0.969	0.962	0.942
20	0.934	0.919	0.925	0.908	0.891	0.867
21	1.000	0.965	1.000	0.944	1.000	0.946
22	0.991	0.976	0.955	0.934	1.000	0.959
23	0.849	0.834	0.677	0.660	0.731	0.711
24	1.000	0.965	1.000	0.943	1.000	0.942
25	1.000	0.964	1.000	0.944	1.000	0.940
26	0.984	0.969	1.000	0.959	1.000	0.955
Efficient schools	**13**	**-**	**15**	**-**	**15**	**-**

The number of efficient schools depends on the output variable being used. As shown in Table 5 some schools are efficient in a model and some are not in the others. This is an important aspect to consider when using the DEA. For instance, when interested in knowing which schools are using resources in a more efficient way to secure a higher average score in the four tests considered here, Model 1 would be the best way to approach that. If the interest is which schools are efficient in achieving a minimum score, then efficiency scores in Model 2 have more relevant quality information. In the case that we want to know what schools are efficient in getting a high percentage of students above the standard score average results in Model 3 are closer to the information needed.

Table 6 shows the eight schools that are not efficient in any of the models. Those schools failed to achieve the higher average or minimum score possible relative to the resources they have and relative to other schools. Among these inefficient schools a total of five are from District 1 (63%). District 3 doesn't have schools listed as inefficient in any of the models.

Table 6. Inefficient schools in all models (not bias corrected DEA scores)

DMU	Model 1	Model 2	Model 3
2	0.878	0.958	0.802
5	0.954	0.913	0.957
7	0.946	0.911	0.982
8	0.964	0.993	0.986
12	0.976	0.935	0.920
16	0.941	0.929	0.899
20	0.934	0.925	0.891
23	0.849	0.677	0.731

Likewise, efficiency scores distributed by districts show that District 2 is the most inefficient district because it has the lower percentage of efficient schools. Only 25% of the schools in District 2 are efficient when average score or percentage of students above standard average is considered as output. While all schools in District 3 are always efficient (Table 7).

Table 7. Percentage of efficient schools by district

District	Model 1	Model 2	Model 3	Number of schools in the district
1	40%	47%	60%	15
2	25%	50%	25%	4
3	100%	100%	100%	4
4	67%	67%	33%	3

District 1 concentrates most of the schools considered in this research. 57% of the schools belong to this district or 15 out of 26 schools. District 2 and District 3 have 4 schools each. The district with less schools participating in StartSmart in summer 2011 is District 4 with only 3 schools. The sample distribution at the district level is consistent with the districts size.

Another application for the DEA model is that a set of efficient levels of inputs and outputs can be found by dividing the observed input or output by the efficiency of each unit. This can be used to set targets for output rather than reduction of inputs (Trick, 1998). In the specific case of an outcome oriented evaluation, like the analysis presented in this study, it is possible to use the results to set targets for desirable outcomes given a certain amount of inputs. For instance, DMU 20 is inefficient because average scores achieved by schools with similar amount of input are higher. In order to be efficient, DMU 20 should achieve scores like the ones in Table 8.

Table 8. Targeted average score for DMU 20 to be efficient (Model 1)

	Reading	Math	Writing	Vocabulary
Current average scores	82.33	87.78	83.22	98.11
Target ($\hat{\theta}_{\mathrm{DMU20}} = 0.934$)	88.15	93.98	89.10	105.04

Same analysis is done using the outputs in Model 2 and 3. For DMU 20 to be efficient, when considering minimum scores as an output, it should have minimum scores as the one shown in Table 9. The target outcome is between 2 and 6 points higher than the current minimum scores. Likewise, the target outcome for DMU 20 estimated using the efficiency score obtained in Model 3 shows that this school should increase the percentage of students that have a score equal or greater than 90. The specific targets under the Model 3 assumptions are presented in Table 10.

Table 9. Targeted minimum score for DMU 20 to be efficient (Model 2)

	Reading	Math	Writing	Vocabulary
Current minimum scores	26	51	54	81
Target ($\hat{\theta}_{\mathrm{DMU20}} = 0.925$)	28	55	58	88

Table 10. Targeted percentage of students for DMU 20 to be efficient (Model 3)

	Reading	Math	Writing	Vocabulary
Current % of students with a score of 90 or higher	44%	44%	33%	77%
Target ($\hat{\theta}_{DMU20}$ = 0.891)	49%	49%	37%	86%

5.2. Tobit Regression Results

Bias-corrected DEA scores are used as the dependent variable in a second stage regression model to assess the effects of variables not included in the first stage. A Tobit regression model identifies the relationship between schools' efficiency and variables listed in Table 4, such as education level of the students' closest caregiver, school district's poverty rate, median household income of the city, and the time that students spend watching non-education television or playing non-educational video games. The parametric model to be estimated takes on the following form:

$$(6) \qquad \hat{\hat{\theta}}_j = \beta_0 + \beta_1 \text{Education}_j + \beta_2 \text{Poverty}_j + \beta_3 \text{VGTV}_j + \varepsilon_j$$

where $\hat{\hat{\theta}}_j$ is bias-corrected DEA scores for school j, Education_j is the average education level of the student's closest caregiver over schools. Poverty_j refers to percentage rate of the district where the school is located. VGTV_j is the time students spend watching non-education television or playing non-educational video games. Income was left out of the model because was highly correlated with poverty rate (-0.91). A correlation table with information about the Pearson's coefficients for each of these variables is in Appendices (Table 13). The estimated coefficients and p-values from all three alternative models are reported in Table 11.

Table 11. Bias-corrected efficiency scores as dependent variable using Tobit

	Model 1		Model 2		Model 3	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Education	0.024^{***}	(0.008)	0.034^{**}	(0.017)	0.035^{**}	(0.013)
Poverty	-0.001^{**}	(0.0006)	-0.0008	(0.001)	-0.002^{**}	(0.0009)
VGTV	-0.034^{*}	(0.018)	-0.045	(0.039)	-0.061^{**}	(0.031)
_cons	0.957^{***}	(0.067)	0.899^{***}	(0.140)	0.972^{***}	(0.112)
Pseudo-R2	0.49		0.22		0.47	
Prob>chi2	0.0006		0.0973		0.0010	
Log likelihood	59.22		39.89		45.81	
Obs.	26		26		26	

Note: dependent variable in Model 1 is bias-corrected DEA score generated using average scores as output. Dependent variable in Model 2 is bias-corrected DEA score using minimum scores. Dependent variable in Model 3 is bias-corrected DEA score using percentage of students above standard score average.
*** 1%, ** 5% and * 10% significance level

Most of explanatory variables are statistically significant and all of variables have the expected sign. The education level of the student’s closest caregiver has a positive effect on schools’ efficiencies; in other words, the caregiver’s education level has a positive impact on students’ performance. The opposite happens with poverty, meaning that the schools have lower efficiency scores the higher is the poverty rate. It indicates that, ceteris paribus, the higher the poverty rate in the school district the lower the efficiency of schools. These results are consistent with previous studies which suggest that school districts heavily populated by students from a less advantage family environment are more likely to be less efficient (Rassouli, 2011).

VGTV has also expected signs in all Models. The negative signs in VGTV means that the more time students spend watching non-education television and playing non-educational video games, the lower performance in school will be, keeping everything else constant.

Another way to interpret these results is by calculating the elasticity at the mean of each of the variables. Elasticities are useful because they are unit-free. They provide a more accessible means of interpreting and explaining the effects of causal variables. This is calculated as the percentage change in Y (the dependent variable) divided by the percentage change in X (the explanatory or independent variable). Elasticities tend to differ when measured at different points on the regression line. Table 12 contains elasticity at the mean for Education, Poverty and VGTV.

Table 12. Margin effect. Elasticity at the mean

	Model 1		Model 2		Model 3	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Education	0.123***	(0.040)	0.177**	(0.088)	0.182***	(0.070)
Poverty	-0.037**	(0.016)	-0.024	(0.035)	-0.070**	(0.028)
VGTV	-0.093*	(0.050)	-0.126	(0.108)	-0.169***	(0.086)

Note: dependent variable in Model 1 is bias-corrected DEA score generated using average scores as output. Dependent variable in Model 2 is bias-corrected DEA score using minimum scores. Dependent variable in Model 3 is bias-corrected DEA score using percentage of students above standard score average.
*** 1%, ** 5% and * 10% significance level

Elasticity's coefficients suggest that Education has the higher effect in changes in the dependent variable in all three models. A 1% increase in education could increase efficiency score in approximately 0.12 to 0.18%. At the same time, VGTV is expected to have significant effects on efficiency score in Model 1 and 3. A 1% decrease in VGTV could increase efficiency score in 0.09 to 0.17%. Poverty has a smaller effect than Education and VGTV in all models and it is not statistically significant in explaining changes in the dependent variable in Model 2.

6. CONCLUDING REMARKS

New Mexico has administered and funded State K3+ program to reduce the achievement gap between students in kindergarten through third grade since 2007. StartSmart K3+ project is an experimental research to examine the cost-effectiveness of State K3+. This research attempts to measure the efficiency of the schools participating in StartSmart using valuable information and data collected by StartSmart K3+. A Data Envelopment Analysis (DEA), originally developed to study production efficiency of micro-level organizations, and a regression models are used to analyze the efficiency of the schools that participated in the first year of the project in 2011.

The DEA is used to measure each school's input-outputs ratio, compares them, and calculate the efficiency score. Efficiency scores generated by the DEA are biased by construction since the DEA constructs a lower bound on the true efficient frontier. Efficiency scores from the DEA are corrected using the bootstrap procedure as suggested in Simar and Wilson (1998, 2000). A regression model is used to identify the environmental factors that affect schools' performance after generating DEA scores. Two-limit Tobit with limits at zero and unity is used to estimate equations.

Three performance measurements are identified as outputs: 1) average scores in reading, writing, math and vocabulary, 2) minimum scores in four subjects, and 3) percentage of students with scores above Woodcock-Johnson Standard Score average (90 points). Results suggest that between 50% and 58% of the schools were efficient in 2011, depending on the students' performance measurements considered. In other words, inefficient schools range from 36 to 50%. This percentage of inefficient schools is a bit high comparing to other studies, e.g., Rassouli (2011), where the percentage of inefficient school is 18%. Nevertheless, any of the previous studies have focused on voluntary summer school program like StartSmart. It would be beneficial to expand the sample students to 2012 when the data becomes available. The inter-temporal changes would be of interest.

Additionally, inefficiencies can't be attributed only to schools. Three Tobit regression models in Table 11 are estimated using bias-corrected DEA scores on education level of the closest caregiver, poverty rate in the school district and other variables. Regression results tell us that education level of the closest caregiver and district's poverty rate are important factors deciding schools' efficiency. This relationship could be critical for any effort focused on trying to increase the efficiencies of individual schools and for policy purposes. These variables are not under control of schools but some of them can be controlled by the family, time watching television, and others can be improved through public policies.

Schools in areas with high-risk populations will require a greater share of resources to provide the same quality of education enjoyed in more affluent areas. The goal pursued by the Government of New Mexico of reducing the existing achievement gap between students will be limited by these inefficiencies. An efficiency evaluation could be carried out at the end of each summer session to identify inefficient schools and to better allocate resources. Following steps after inefficient schools are identified have to be recognizing what are the differences between efficient and inefficient schools in terms of time dedicated to teach subjects related to numeracy and literacy.

Short and long run policies should be implemented to increase schools' efficiency. Policies related to the time students spend watching television and playing non-education video games could be implemented in a short period of time. The state of New Mexico can implement a program for parents to make them aware of the negative impact that the time spent watching non-education television and video games has in students' performance at school.

Strategies aiming to increase caregivers' education level would take more time. Nevertheless, a program to provide courses for parents could be linked to State K3+ and StartSmart. During the summer, parents of children attending State K3+ could be enrolled in an educational program that fits their schedule restrictions. At the same time, the State K3+ itself increase the likelihood of having better educated parents in the future. This will create a cycle that will benefit school efficiencies. Other policies related to decrease districts’ poverty rate and increase household income will take more time, but they are possible too.

REFERENCE

Adkins, L. C., and R. L. Moomaw. 2007. Analyzing the technical efficiency of school districts in Oklahoma. *The Journal of Economics,* 33(2): 41-61.

Afonso, A., and M. Aubyn. 2006. Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Elsevier* 23(3): 476-491.

Banker, R.D., A. Charnes, and W.W. Cooper. 1984. Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science* 30(9): 1078-1092.

Banker, R.D. 1996. Hypothesis tests using data envelopment analysis. *Journal of Productivity Analysis* 7: 139–159.

Bessent, A., W. Bessent, J. Kennington, and B. Reagan. 1982. An application of mathematical programming to assess productivity in the Houston Independent School District. *Management Science* 28(12): 1355-1367.

Boonen, T., J.V. Damme, and P. Onghena. 2013. Teacher effects on student achievement in first grade: which aspects matter most? *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice,* DOI: 10.1080/09243453.2013.778297

Buddin, R., and G. Zamarro. 2009. Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics* 66: 103-115.

Charnes, A., W.W. Cooper, and E. Rhodes. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6): 429-444.

Charnes, A., W.W. Cooper, and E. Rhodes. 1981. Evaluating program and managerial efficiency: An application of Data Envelopment Analysis to program follow through. *Management Science* 27(6): 668-697.

Cohen, D.K., S.W. Raudenbush, and D.L. Ball. 2003. Resources, instruction, and research. *Educational Evaluation and Policy Analysis* 25(2): 119-142.

Cooper, W., L. Seiford, and J. Zhu (Eds.). 2004. Handbook on Data Envelopment Analysis. Norwell, MA: Kluwer Academic Publishers. DOI: 10.1007/978-1-4419-6757-8. New York: Springer Science + Business Media

Cooper, W.W., L.M. Seiford, and K. Tone. 2007. Data Envelopment Analysis: A comprehensive text with model, applications, references and DEA-Solver software. New York: Springer Science + Business Media.

Greene, W. 2000. Econometric Analysis. Upper Saddle River, NJ: Prentice Hall.

Hatamiyan, H., M. Asgari, and S. Heshmati. 2013. Evaluating schools' performance using Data Envelopment Analysis (DEA). *International Research Journal of Applied and Basic Sciences* 5: 53-59.

Kim, J.S., and D.M. Quinn. 2013. The effects of summer reading on low-income children’s literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research* 83(3): 386-431.

Mather, N., and R.W. Woodcock. 2001. Examiner’s Manual. Woodcock-Johnson II Tests of Achievement. Rolling Meadows, IL: Riverside Publishing.

Matsudaira, J.D. 2008. Mandatory summer school and student achievement. *Journal of Econometrics* 142: 829–850.

Raposo, I., T. Menezes, S. Cavalcanti, and L. Maia. 2011. *Public school efficiency using Data Envelopment Analysis: An empirical application for Brazil*. Available at http://www-sre.wu.ac.at/ersa/ersaconfs/ersa11/e110830aFinal01594.pdf (accessed January 28, 2014).

Rassouli, S. 2011. Assessing the efficiency of Oklahoma public schools: A Data Envelopment Analysis. *Southwestern Economic Review*. Available at https://www.cis.wtamu.edu/home/index.php/swer/article/viewFile/59/53 (accessed January 28, 2014).

Shephard, R.W. 1972. Theory of cost and production function. *The Economic Journal* 82(327): 1059-1061.

Simar, L., and P. Wilson. 2007. Estimation and inference in two-stage semi-parametric models of production processes. *Journal of Econometrics* 136: 31–64.

Simar, L., and P. W. Wilson. 2002. Non-parametric tests of returns to scale. *European Journal of Operational Research* 139: 115-132.

Simar, L., and P.W. Wilson. 2000. A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics* 27: 779–802.

Simar, L., and P.W. Wilson. 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* 44(11): 49–61.

Trick, M. 1998. Data Envelopment Analysis. Available at http://mat.gsia.cmu.edu/classes/QUANT/notes/node92.html (accessed January 28, 2014)

Wilson, P. W. 2008. FEAR 2.0: A Software Package for Frontier Efficiency Analysis with R. *Socio-Economic Planning Sciences* 42: 247-254.

Zvoch, K., and J.J. Stevens. 2013. Summer school effects in a randomized field trial. *Early Childhood Research Quarterly* 28: 24– 32.

APPENDICES

Table 13. Correlation Matrix

Variables	Model_1	Model_2	Model_3	Education	Poverty	Income	VGTV
Model_1	1						
Model_2	0.6930	1					
Model_3	0.9265	0.7228	1				
Education	0.5926	0.4118	0.5474	1			
Poverty	-0.5112	-0.2556	-0.5190	-0.4732	1		
Income	0.4643	0.2875	0.4866	0.4197	-0.9125	1	
VGTV	-0.0231	-0.0646	-0.0450	0.2592	-0.3488	0.2544	1

Note: Model 1 is bias-corrected DEA score generated using average scores as output. Model 2 is bias-corrected DEA score using minimum scores. Model 3 is bias-corrected DEA score using percentage of students above standard score average.