

Evaluation and determinants of pre-school effectiveness in Chile

Economics Department

Víctor Giménez
Claudio Thieme
Diego Prior
Emili Tortosa-Ausina

2020 / 02

Evaluation and determinants of pre-school effectiveness in Chile

Víctor Giménez

Universitat Autònoma de Barcelona
Department of Business
victor.gimenez@uab.es

Claudio Thieme

Universidad Diego Portales
Facultad de Economía y Empresa
Claudio.thieme@udp.cl

Diego Prior

Universitat Autònoma de Barcelona
Department of Business
diego.prior@uab.cat

Emili Tortosa-Ausina

IVIE & IIDL & Universitat Jaume I
Department of Economics
tortosa@uji.es

2020 / 02

Abstract

Early intervention in quality education is a way to equalize opportunities, a premise assumed by several countries that has increased interest in preschool education, which is now a relevant area of public policy. However, the performance of preschool education centers has received relatively scarce research attention. Consistent with the above, there are no studies quantifying the level of effectiveness with which these centers operate. This study examines the problem and aims to evaluate the performance of preschool education centers that serve children from lower socioeconomic families in Chile. For this, a centralized DEA Assurance Region model is proposed, which, in addition to avoiding the main criticisms received by the traditional Benefit of Doubt models, shows important differences in the results of efficacy and weights of the composite indicators. In addition, a second stage analysis is carried out using decision trees to identify variables that determine the composition of homogeneous preschool groups according to their effectiveness. The target population evaluated is kindergartens serving families with low socioeconomic status in urban and rural areas in Chile, and results reveal the importance of three structural factors: family income, kindergarten size, and location.

Keywords: assurance region, centralized data envelopment analysis, composite indicators, effectiveness, preschool

JEL classification: C61, H52, I21

Evaluation and determinants of pre-school effectiveness in Chile*

Víctor Giménez

Universitat Autònoma de Barcelona

Claudio Thieme

Universidad Diego Portales

Diego Prior

Universitat Autònoma de Barcelona

Emili Tortosa-Ausina

Universitat Jaume I, IIDL and Ivié

December 3, 2019

Abstract

Early intervention in quality education is a way to equalize opportunities, a premise assumed by several countries that has increased interest in preschool education, which is now a relevant area of public policy. However, the performance of preschool education centers has received relatively scarce research attention. Consistent with the above, there are no studies quantifying the level of effectiveness with which these centers operate. This study examines the problem and aims to evaluate the performance of preschool education centers that serve children from lower socioeconomic families in Chile. For this, a centralized DEA Assurance Region model is proposed, which, in addition to avoiding the main criticisms received by the traditional Benefit of Doubt models, shows important differences in the results of efficacy and weights of the composite indicators. In addition, a second stage analysis is carried out using decision trees to identify variables that determine the composition of homogeneous preschool groups according to their effectiveness. The target population evaluated is kindergartens serving families with low socioeconomic status in urban and rural areas in Chile, and results reveal the importance of three structural factors: family income, kindergarten size, and location.

Key words and phrases: assurance region, centralized data envelopment analysis, composite indicators, effectiveness, preschool.

JEL Classifications: C61, H52, I21.

Communications to: Emili Tortosa-Ausina, Departament d'Economia, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain. Tel.: +34 964387168, e-mail: tortosa@uji.es.

*All four authors are grateful to the INTEGRA Foundation for kindly providing data and facilitating interaction with school managers to gain further insights on the available information. Claudio Thieme and Emili Tortosa-Ausina thank FONDECYT (National Fund of Scientific and Technological Development, grant #1151313) for generous financial support. Víctor Giménez, Diego Prior and Emili Tortosa-Ausina acknowledge the financial support of the Spanish Ministerio de Economía y Competitividad (ECO2017-88241-R and ECO2017-85746-P). Emili Tortosa-Ausina also acknowledges the financial support of Generalitat Valenciana (PROMETEO/2018/102) and Universitat Jaume I (UJI-B2017-33). All authors are grateful for comments from Sergio Perelman, Dovilé Stumbriené, Kristof De Witte and other participants at the 7th International Workshop on Efficiency, September 2019, Barcelona. The usual disclaimer applies.

1. Introduction

There is a consensus in political and academic circles that Chile must advance in the provision and quality of preschool education in order to correct—at least partly—the problems of quality and equality in its education system and thereby contribute to reducing the country's significant levels of inequality (Contreras and Puentes, 2017; Alarcón et al., 2015; OECD, 2015).¹ There are two reasons why education is a high priority. First, the nation's human capital is considered an important explanatory component of wealth and growth rate discrepancies (Bank, 2011); and second, it is an essential component in ensuring equal conditions for all members of a society (Hanushek, 1998, 1986).

The specific case of childhood education is crucial to children's lives because it is the stage in which the foundations of cognitive, linguistic and physical development are laid. Early life experiences can have a critical impact on an individual's life course (Mutindi et al., 2016; McCartney et al., 2007). There is consensus in the academic literature that early intervention in quality education is a source of equal opportunities, among other advantages (Taguma et al., 2012). Its importance has been stressed in numerous studies, including recent relevant contributions such as, for instance, Heckman et al. (2013) and García et al. (2019), who focus on the benefits of early childhood programs (García et al., 2018, extend their analysis to include gender differences), or List et al. (2019), who evaluate early childhood effects on social preferences.² Due to the obvious benefits of early childhood education from a range of viewpoints, efforts should therefore be devoted to improving learning levels and enabling the whole population to access such education. This will undoubtedly affect not only countries' wealth but also their levels of equality.

At the political level, various public policies related to the care and integral protection of children have been implemented in Chile over the last decade (Consejo Asesor Presidencial para la Calidad de la Educación, 2006). Similarly, Michelle Bachelet's government program proposed reforms to preschool education that aimed to increase coverage, improve quality and modernize preschool education institutions, through the creation of the Subsecretary's Office of Preschool Education in the Ministry of Education.

¹An illustration of these problems is reflected by the unrest in the country during October 2019. See, for instance, <https://www.theguardian.com/world/2019/oct/21/chile-braces-after-worst-unrest-in-three-decades-claims-11-lives>.

²In this regard, as García et al. (2019) point out, there is a large body of evidence documenting how early childhood programs can affect the skills of disadvantaged children (Almond and Currie, 2011; Duncan and Magnuson, 2013), as well as the long-term benefits in terms of completed education, adult health, labor income, and crime (Heckman et al., 2010; Havnes and Mogstad, 2011; Campbell et al., 2014).

The rationale provided above, based on solid evidence of low levels of quality in Chilean kindergartens, with no noticeable progress made among children living in urban areas, focused mainly on cognitive skills (Strasser et al., 2009; Sun et al., 2011), but also attempted to redress deficiencies in learning environments and materials, educational environments and socio-emotional development (Lien Foundation, 2012; CEDEP, 2011).

Regrettably, compared to the large body of microeconomic research carried out in Chile and around the world evaluating the internal efficiency of other educational institutions (Johnson and Ruggiero, 2014; Johnes, 2015; De Witte and López-Torres, 2019), performance of preschool education centers has received little academic attention, with a few exceptions (e.g., Moyer-Packenham et al., 2015). Consistent with the above, there is an absence of studies quantifying the level of effectiveness with which these centers operate, or the degree to which different educational contexts affect their performance. In order to meet the challenge of developing higher quality and more effective preschool education systems, the magnitude of the challenge must be accurately quantified.

Our study considers these issues, with the specific aim of evaluating the performance of preschool education institutions in Chile catering for students of relatively lower socioeconomic status, from the perspective of their stakeholders: children, families and school staff. We will also analyze the way in which different educational contexts can affect performance. Specifically, we propose a composite effectiveness indicator based on the merger of three previous methodologies to evaluate performance. After this first stage, in the second stage we use decision trees (Emrouznejad and Anouze, 2010; Breiman et al., 1984) to classify the determinants of the differences in effectiveness found.

Preschool education institutions make up the first of the three sequential levels in which the Chilean school system is organized: (i) preschool (kindergarten) education (up to 6 years); (ii) basic education (8 years, typically children aged 6 to 13 years old); and (iii) secondary education (4 years, typically children aged 14 to 17 years old). Preschool education is organized into six levels according to age. It is free for children over the age of two (lower average level). The second one at the second transition level (children between 5 and 6 years old) was also made compulsory in 2014, but its implementation is still underway.

Responsibility for early childhood education policies lies with the Ministry of Education (Ministerio de Educación), the National Board of Kindergartens (JUNJI, Junta Nacional de Jardines de Infancia) and the INTEGRA Foundation (Fundación INTEGRA).³ Figures for the year

³The JUNJI is an autonomous public corporation funded by the state and supervised by the Ministry of Ed-

2015 show a total of 633,731 children enrolled in early childhood education centers, broken down as follows: 28% served by the JUNJI, 12% by the INTEGRA Foundation, 20% by municipalities, 33% by private subsidized centers, and 7% by private centers (Ministerio de Educación, 2005). Although the total number of students enrolled has increased significantly during the last few years, it is still below the OECD average, especially at earlier levels of education. Today, Chile has embarked on an ambitious plan of investment in infrastructure for early childhood education that will provide greater coverage at this level.

Efficiency and effectiveness are traditional indicators of organizational performance (Golany et al., 1993). According to Farrell (1957), efficiency refers to the ability of a decision making unit to produce the largest attainable output from a given set of inputs, or of using the lowest possible input to obtain a certain output. In turn, effectiveness is related only to the results of the process evaluated. As Achabal et al. (1984) point out, the effectiveness question is concerned with determining which strategy (process characteristics or structural factors) maximizes organizational goals. In both cases, evaluating the performance of educational organizations is a methodologically complex task that can be undertaken from different approaches (Frankel, 2008). From an academic perspective, particularly in operations research, parametric and non-parametric frontier models have gained considerable importance in recent years (Bădin et al., 2014; Emrouznejad and Yang, 2018). On the other hand, composite indicators have become popular and widely used metrics for assessing complex phenomena which cannot be evaluated with a single variable (Maricic et al., 2019). A significant number of these evaluations correspond to measures of effectiveness of organizations, countries or regions, which are usually transformed into rankings with high degrees of publicity (see, for instance Benito and Romera, 2011; Castillo-Giménez et al., 2019; Ciommi et al., 2017).

Because we are interested in evaluating the performance of a sample of kindergartens belonging to the same organization (INTEGRA Foundation), which has centrally established resource endowment standards for all its kindergarteners (and, therefore, resource provision is not the responsibility of their directors), it might be appropriate to use an efficiency measure as a performance indicator. To do this, we propose a composite indicator that combines three methodological developments from the two above mentioned approaches, namely, benefit of the doubt, assurance region and centralized DEA. Its application to this type of organization is

education, whereas the INTEGRA Foundation is a private non-profit foundation, also funded by the state. Both institutions deal with socioeconomically disadvantaged families—usually from the first two quintiles of the per capita income distribution. In addition to these institutions, preschool education is also provided by a series of other agents: municipalities, private kindergartens, and private subsidized kindergartens.

innovative *per se* since, despite the proliferation of empirical studies in the field of education in recent years, to our knowledge this is the first to be carried out at the preschool level. Finally, we aim to identify the determinants of performance differences in structural aspects that have received less scholarly attention, namely: (i) the kindergarten's human capital; (ii) organization of the service; and (iii) the kindergarten's environmental variables.

The results of the successive stages of refinement of the model reveal significant differences for the benefit of the doubt and assurance region models, but not between the assurance region and unique weights models (centralized DEA). For the unique weights model, the average efficiency index is 70.54%, with significant heterogeneities across regions. Likewise, the centralized (unique) weights obtained are 45% for learning, and 32% for user satisfaction and job satisfaction. A second stage analysis using decision trees reveals the importance of three structural factors that mark the effectiveness of kindergartens: (i) size or specialization of the kindergarten, measured as the number of levels it serves; (ii) family income; and (iii) urban or rural location.

The paper proceeds as follows. After this introduction, Section 2 presents the methodology and Section 3 describes the data. Section 4 then reports the results, and finally, Section 5 presents the conclusions of the study.

2. The proposed methodology: measuring education effectiveness

The evaluation of organizational effectiveness is a daily task in areas as diverse as education, health, public transport, energy and production, and for producing performance rankings, for organizational control, or for academic purposes. The literature reports numerous methodologies for creating composite indicators (CI) to measure effectiveness, although they can also be used for other purposes such as providing a big picture of multidimensional processes, which can be used to guide public policies, for instance (Saltelli, 2007). CI add different subindicators into a single summary measure that gives an idea of the status of a system (Stumbriene et al., 2019). The Joint Research Centre-European Commission (2008) handbook covers most of the methods for constructing CI; some of the most noteworthy are the AHP (analytic hierarchy process) (Milutinović et al., 2014), the PCA (principal component analysis) method (Gómez-Limón and Riesgo, 2009), gray relation analysis (Lee and Lin, 2011), the Delphi method (Galo et al., 2014) and BoD (benefit of the doubt) (Melyn and Moesen, 1991; Cherchye et al., 2007b). The first three methods require some prior information on the subindicators that will determine the

weight assigned to each of them. In contrast, BoD is based on the use of a data envelopment analysis model (Charnes et al., 1978) without inputs for obtaining endogenous weights, and requires no previous information (Wang, 2015).

BoD has been applied to construct CI in numerous fields in the literature. Some examples of applications are the Digital Access Indicator (Gaaloul and Khalfallah, 2014), the Human Development Index (Despotis, 2005), the Quality of Life Indicator (Morais and Camanho, 2011), the Internal Market Index (Cherchye et al., 2007a,b), the Competitiveness Index (Bowen and Moesen, 2011), education system performance (Stumbriene et al., 2019), the Technology Achievement Index (Cherchye et al., 2008), the Students' Evaluation of Teaching indicator (De Witte and Rogge, 2011), the Health System Performance Index (Lauer et al., 2004), the Sustainable Energy Index (Wang, 2015), the Citizen Satisfaction with Local Police Effectiveness (Verschelde and Rogge, 2012), the Life Satisfaction Index (Guardiola and Picazo-Tadeo, 2014), the Well-Being Indicator (Peiró-Palomino and Picazo-Tadeo, 2018) and the Environmental Performance Index proposed by Zanella et al. (2013).

Karagiannis (2017) highlights the discussion in the literature about the level of flexibility for weights in the construction of CI. In this regard, an additional advantage of BoD is that it allows three levels of flexibility for determining weights: a) differentiated weights for each unit analyzed without any restriction (Zhou et al., 2007, 2010); differentiated weights but subject to compliance with certain restrictions through the use of assurance region models (Stumbriene et al., 2019); and c) single weights for all units (Emerson et al., 2012; Ertugrul Karsak and Sebnem Ahiska, 2008; Hatefi and Torabi, 2010; Kao, 2010). The common practice is to define variables that in one way or another represent performance dimensions, and assign weights to each of the evaluated dimensions. The definition of these weights is set discretionally by the person or institution making the evaluation, in accordance with their own judgment and/or the judgment of experts. In this way, unique weights are obtained with which to evaluate all decision making units (DMUs). Although it is discretionary and does not optimize the system as a whole (as we will see later), this definition of weights is considered "fair" by the agents as it is based on the credibility of the person setting the weights, who is assumed to have acted without privileging or harming any DMU in particular. The acceptance then comes from that "the judge's rules, good or bad, are like that; at least they all evaluate us in the same way." This is a similar approach to the one used by Cherchye et al. (2007b).

In the academic literature, operations research has prioritized the BoD frontier approach (Charnes et al., 1978). In this approach, unlike the previous case, the optimal weights for

each organization are determined such that performance is maximized. In other words, the methodology assigns high weights to well evaluated dimensions and low or zero weights in the opposite case. Unfortunately, in many cases this leads to undesired results, either organizationally or socially. Take, for example, our case, which has three dimensions (pupils' learning, parental satisfaction and job satisfaction); it would not be acceptable to classify as effective a center where both job and parents' satisfaction were very high but pupils' learning was very low; a more extreme case would be a "Toby club" kindergarten in which only job satisfaction was very high but families' satisfaction and pupils' learning were very low. If we followed this methodology, weights equal to zero would lead to high effectiveness levels.

The "assurance region" methodology (Stumbriene et al., 2019) offers a solution to this problem. Here, a restriction range is defined in which the weights of each outcome variable can be moved. A practical way to determine this range of feasibility of weights is through expert judgment, which is unlikely to coincide in the weights of each variable, thus creating a range of possibilities to use in the "assurance region".

However, if the study is not purely academic, the agents involved might want this definition of weights to be applied to everyone (all units). It will therefore be essential to consider a unique set of weights with which all organizations are evaluated. The usual solution to this problem is to use the average of the experts' evaluations. Under different guises, this is what one stream of the literature proposes to define a 'common set of weights' (see, for instance, Roll et al., 1991). However, this will not eliminate the discretion and lack of optimization of the system as a whole and, therefore, some agents will always be penalized.

To solve this problem, we adapt the method proposed by Mar-Molinero et al. (2014), who extended Lozano and Villa's (2004) notion to provide, in the optimum scenario, a set of unique weights such that the efficiency of the system as a whole is maximized. Our proposal adds to the above the restriction that the unique weights must belong to the range previously defined by the experts. In this paper, therefore, we will propose an innovation that we label assurance region centralized data envelopment analysis. In this way we ensure a method with more properties, which is the best for the system in that it eliminates arbitrary weights and unrealistic results.

We now introduce some notation: K observations (i.e., the total number of education institutions) produce M outputs with the consumption of N inputs. Production technology is defined by the set of feasible input and output vectors: $S = \{(x, y) \mid x \text{ can produce } y\}$. It is also useful to consider the output and input sets associated with technology. For a given input

vector x , the output set denotes all output vectors y that can be produced from the given input vector x : ($P(x) = \{y \mid (x, y) \in S\}$). Also, for a given output vector y , the input set denotes all input vectors x that are capable of producing this output vector: ($L(y) = \{x \mid (x, y) \in S\}$). With these variables, Charnes et al.'s (1978) seminal work proposes the following optimization program (the so-called DEA multiplier model) to estimate the level of efficiency for a given observation o :

$$\begin{aligned}
\left[Z^{DEA}(x_o, y_o) \right] &= \max_{u_r, v_i} \sum_{r=1}^M u_r y_{ro} \\
\text{subject to } &\sum_{r=1}^M u_r y_{rj} - \sum_{i=1}^N v_i x_{ij} \leq 0 \quad j = 1, \dots, K, \\
&\sum_{i=1}^N v_i x_{io} = 1, \\
&u_r, v_i \geq 0.
\end{aligned} \tag{1}$$

When all the units have the same amounts of inputs, this original proposal was modified by Thompson et al. (1986) to generate the so-called “DEA-benefit of the doubt” model (*DEA-BoD*). Some developments of the *DEA-BoD* model can be found in Cherchye et al. (2007b), Rogge (2011) and Färe and Karagiannis (2014). From another perspective, the *DEA-BoD* model has also been proposed as a rational way to determine composite indicators (Joint Research Centre-European Commission, 2008).

The computation of the *DEA-BoD* model is a development of the original *DEA efficiency model*, defined by the following optimization model:

$$\begin{aligned}
\left[Z^{DEA-BoD}(y_o) \right] &= \max_{u_r} \sum_{r=1}^M u_r y_{ro} \\
\text{subject to } &\sum_{r=1}^M u_r y_{rj} \leq 1 \quad j = 1, \dots, K, \\
&u_r \geq 0.
\end{aligned} \tag{2}$$

Regarding the weights to be assigned to the output variables, previous models are extremely flexible as the only restriction to be met in the optimum is that they must be non-negative. However, when there is additional information reflecting the different importance of the outputs, these value judgments can be introduced into the models by including additional constraints in the optimization model. This generates the family of assurance region DEA models (*AR-DEA*). Following Thompson et al. (1986), the adaptation of the *DEA-BoD* to introduce value judgments by following the so-called assurance region of type I (i.e. controlling the relative value of the

weights) gives rise to the *AR-DEA-BoD model* (Stumbriene et al., 2019) and can be estimated by solving the following optimization model:

$$\begin{aligned}
\left[Z^{AR-DEA-BoD}(y_o) \right] &= \max_{u_r} \sum_{r=1}^M u_r y_{ro} \\
\text{subject to } \sum_{r=1}^M u_r y_{rj} &\leq 1 \quad j = 1, \dots, K, \\
L_r &\leq u_r / u_1 \leq U_r \quad r = 2, \dots, M, \\
u_r &\geq 0.
\end{aligned} \tag{3}$$

One criticism of previous proposals is that the weights are flexible and different for each DMU (Wang, 2015). This could be considered as a weakness in itself, as most of the managerial control literature reports the advantage of establishing a common set of weights for all the units, which should be coherent with the organization's strategic goals. As mentioned previously, this can be achieved by considering the centralized DEA, *CDEA* (Lozano and Villa, 2004; Mar-Molinero et al., 2014), and incorporating the value judgment restrictions. This gives rise to the *AR-CDEA-BoD*:

$$\begin{aligned}
\left[Z^{AR-CDEA-BoD}(y_j) \right] &= \max_{u_r} \sum_{j=1}^K \sum_{r=1}^M u_r y_{rj} \\
\text{subject to } \sum_{r=1}^M u_r y_{rj} &\leq 1 \quad j = 1, \dots, K, \\
L_r &\leq u_r / u_1 \leq U_r \quad r = 2, \dots, M, \\
u_r &\geq 0.
\end{aligned} \tag{4}$$

where L_r and U_r represent the lower and upper bound, respectively, of the u_r / u_1 ratio. This helps to reflect any value judgements that we may wish to include in the assessment.

3. Sample, variables and value judgments (weight constraints)

The data was obtained from a variety of databases provided by the INTEGRA Foundation. The definition of a “good” kindergarten, akin to the schooling system, is a multidimensional concept that depends not only on pupils' learning but also on other outputs. Specifically, the definition of a “good” kindergarten is tied to the identity that defines the institution to which they belong. In the case of the INTEGRA Foundation, we identify three key players:

- Children, specifically the learning they achieve.

- User families, specifically their satisfaction.
- Internal users, specifically the job satisfaction of kindergartens' staff.

From the children's perspective: an average value-added measure of learning outcomes per kindergarten was used as the output. The learning data corresponds to the results of the standardized "Learning Appraisal Instrument" test (corresponding to the Spanish initials IEA, *Instrumento de Enseñanza y Aprendizaje*), which measures three dimensions of learning: (i) personal and social training; (ii) communication; and (iii) relationship with the natural and cultural environment. An examiner measures the level of compliance with the behaviors/learning expected for children of the same age in these three dimensions, on a scale of three levels. This test is applied on a census basis to all children in the Foundation kindergartens three times a year (at the beginning, mid-year, and end of the year). In this way, each child's progress can be evaluated with respect to the expected behaviors/learning. In our case, we used the measure of "average value added of kindergarten learning", which is obtained by subtracting the average score obtained by each child at the end of the year minus the score obtained at the beginning of the school year.

From the user families' perspective: our analysis was based on the results of the national study to measure satisfaction systematically carried out by the INTEGRA Foundation. This study measures various aspects of user satisfaction, including perceived quality of the kindergarten, continued use and recommendation by the families. This paper used the Full Satisfaction Index, which corresponds to the percentage of user families with a high probability of being highly satisfied, and of continuing to use and to recommend the kindergarten. That is, the likelihood that in each kindergarten, families fulfilled all of the following: (i) were highly satisfied; (ii) would recommended the kindergarten to relatives or acquaintances; and (iii) intended to continue sending their child to the same kindergarten.

From the staff's perspective: we used the job satisfaction scale from a national survey measuring climate and job satisfaction, conducted by the INTEGRA Foundation. This output was labeled "Job satisfaction in the kindergarten".

We report descriptive statistics for the effectiveness variables in Table 1. In addition, as mentioned earlier in the paper, we consider 18 variables associated with three dimensions to identify structural variables that determine the composition of homogeneous preschool groups

according to their effectiveness. These three dimensions are (i) human capital, (ii) service organization, and (iii) environment. A description of these variables is reported in Table 2.

To define the model weight constraints, we used the judgments of nine experts, of whom five were researchers in preschool education, and the rest were members of the INTEGRA Foundation's management team. Each of these nine experts responded to the following questions: (i) which of the following three characteristics should be found in a kindergarten?; and (ii) if the three characteristics are included in the definition, how important (in percentage terms) should each one of them be? The three variables are those listed at the beginning of this section.

In Table 3 we report the data corresponding to the expert responses for the three variables. The last two rows in the table report the ratios corresponding to the weights (u_2/u_3 and u_1/u_3), which enter the proposed model. The last two columns (and two rows) of the table report the minimum and maximum of the weights' ratios. The sample consisted of 712 kindergartens, of which 586 were urban and 126 were located in rural areas.

4. Results

4.1. Effectiveness

The models described in Section 2 were applied to evaluate levels of effectiveness in the kindergartens. Prior to the calculations, the original variables were standardized based on the maximum value of 100. In this way, the interpretation of the weights is not conditional on the unit of measurement for each variable, and can be directly assimilated to the importance that each model assigns to each dimension of the composite indicator of effectiveness.

4.1.1. *DEA-Benefit of the Doubt (DEA-BoD)*

First, the effectiveness of each kindergarten was calculated using Model (3), that is, the conventional "benefit of the doubt" DEA model (Färe and Karagiannis, 2014), which considers just three outputs without imposing any restrictions on the weights assigned to each output variable. These weights are endogenously and freely determined by the model. Table 4 shows the average results obtained in this first estimation by region and rural or urban location type.

The average effectiveness levels obtained by kindergartens located in rural areas were somewhat higher than in urban areas (87.08% vs. 83.91%), with an average of 84.47%. The average weights assigned by the model to each of the dimensions in the calculation of the effective-

ness composite index (those that most benefit each kindergarten) was 0.07 for the value-added dimension (V1), 0.32 for the family satisfaction dimension (V2), and 0.62 for the job satisfaction dimension (V3). The weights assigned to the first dimension hardly differ according to whether the environment is rural or urban. However, in the case of family satisfaction (V2), the model assigned a lower weight in rural kindergartens (0.29 vs. 0.33) while the opposite occurred for the job satisfaction dimension (0.64 vs. 0.61). The results suggest that there is a higher level of demand in the groups of families and workers in urban areas, and that the greater weight assigned by the model to these dimensions contributed to their outperforming rural areas. The regions with the most effective kindergartens were Aysen (90.90%) and Valparaíso (87.49%). In contrast, the regions of Atacama and Región Metropolitana had the lowest levels of effectiveness, both with 77%.

Interestingly, one variable that *a priori* would seem important in the construction of a kindergarten effectiveness index, namely, the measure of the children's progress (added value), is precisely the one that the model deems less important. This is because when dealing with a dimension like this with a worse performance in relative terms, the model assigns it a significantly lower importance due to the assumption of freedom of weights. On the other hand, job satisfaction has a greater weight than the rest, showing that in general workers' satisfaction is the most successful dimension in the management of the kindergartens. Undoubtedly, the flexibility of the DEA weights is an advantage, but in turn it can become a weakness if we obtain unrealistic weights that may even completely ignore some of the dimensions. In fact, the minimum values of the weights are zero in all cases, which means that a value of zero was assigned to some of them for the effectiveness index of some kindergartens as a consequence of bad behavior. Therefore, it is interesting to introduce experts' and/or stakeholders' value judgments by restricting the possible values associated with the weights of each dimension. Proceeding in this way will allow a certain flexibility in determining the DEA weights while simultaneously adapting to the possible strategies of each evaluated unit.

4.1.2. Results from the Assurance Region model (AR-DEA-BoD)

After consulting nine experts, the range in which the weights quotients should be located was determined (known in the literature as the assurance region). Specifically, the weight assigned to family satisfaction with respect to the weight of the added value should range between 0.4 and 0.7, while the weight of job satisfaction with respect to the weight assigned to added value should range between 0.2 and 0.7. Considering these restrictions and after solving the model

(2), the new results are shown in Table 5. The first consequence of considering the experts' criteria is the change in the average weights assigned by the DEA to each of the dimensions, which indicates that the restrictions introduced were certainly effective from a mathematical point of view. The added value dimension is now allocated the highest weight (0.48), followed by job satisfaction (0.38) and family satisfaction (0.29). Consequently, the absolute freedom of weights in the DEA can lead to unrealistic results, because they are not aligned with what the experts really consider desirable. Hence the importance of introducing some control over them, although without losing the flexibility offered by the DEA to adapt to the different strategies of the units analyzed. In this case, the discrepancies in the importance the experts assign to each criterion means that different weights can be assigned to each dimension for each kindergarten in order to better adapt to their own strategies and results, but within the limits that the experts consider plausible and recommendable. With this new set of weights, the average level of effectiveness decreases to 71.03%, with no significant differences between rural and urban kindergartens. Similarly, the difference in the weights assigned to the different dimensions between the two areas is also very similar, giving greater consistency to the results obtained. By regions, Araucania, Los Ríos and Valparaíso are the most highly rated, while the lowest rated are Coquimbo, Región Metropolitana and Atacama, confirming the potential for improvement of these two last regions, regardless of the model applied.

4.1.3. Results from the DEA “benefit of the doubt” with assurance region and unique weights (AR-CDEA-BoD)

Despite the advantages of evaluating each unit with a personalized set of weights, in practice researchers often prefer to calculate composite indices using the same set of weights for all the units in order to facilitate interpretation. The proposed methodology of the DEA “benefit of the doubt” with assurance region also allows us to calculate a unique set of weights. As discussed in section 2, the centralized DEA offers an excellent alternative to determine the set of unique weights that maximizes the overall average efficiency of the system being analyzed, the set of kindergartens in this case. To do this, we simply have to introduce a new fictitious kindergarten representing the average kindergarten in the system. After calculating the model, the weights assigned to the new unit were 0.45, 0.32 and 0.32, respectively, with an average effectiveness level for the system of 70.54%. These weights do not differ greatly from the average weights calculated previously, but they have a very interesting property inherent to the centralized DEA model: the average of the effectiveness index calculated with these weights for the entire

sample coincides with the level of effectiveness of the fictitious average unit of the system. This allows us to determine how each kindergarten contributes to the overall effectiveness of the system.

Table 6 shows the results obtained after calculating the effectiveness index for each kindergarten using the unique set of single weights of the centralized DEA. As expected due to the reduced flexibility in the determination of weights, the average effectiveness decreased again; it now stands at 70.54%. As before, there are no large differences between the rural and urban environments (69.14% vs. 70.84%). However, the kindergarten with the worst performance is below 40%, confirming that there is considerable general potential for improvement. Once more, the regions of Auracanía and Los Ríos are those with the best performance, and Coquimbo, Atacama and the Metropolitan Region emerge as those with the greatest potential for improvement. The fact that the Metropolitan Region repeatedly appears among the worst performing regions seems to indicate that proximity to a developed, competitive environment such as the country's capital coincides with more demanding families and workers, who rate their satisfaction with kindergarten performance more negatively.

4.2. Evaluating the determinants of education effectiveness via decision trees

The vast majority of studies analyzing the determinants of effectiveness in preschool education focus mainly on factors associated with instructional practices, teacher personality and motivation, leadership and family involvement. On this question, Ma et al. (2016) found that the role of parents (family involvement) was more important than the role of schools and communities. Park et al. (2019) showed that teachers' instructional actions make the greatest difference to students' learning outcomes, although it is not entirely clear which of these actions matter most. In contrast, effective leadership does influence the quality of early childhood education (Hujala et al., 2013). However, the studies examining the relationship between the structural characteristics of kindergartens and their effectiveness are fewer, and report mixed evidence. Within these structural aspects there are three major dimensions: human capital, service organization and environment.

The most commonly studied human capital variables are teachers' educational qualifications and work experience, although results are mixed (Slot et al., 2018). For example, Kuger et al. (2016) and Slot et al. (2017) found a positive relationship between teachers' work experience and process quality; in contrast, Connor et al. (2005) and Wilcox-Herzog (2004) revealed a negative relationship, and Justice et al. (2008) found neither of these associations. Similarly,

findings on teachers' educational qualifications are also inconclusive. Cave and Mulloy (2010) emphasized the importance of continuing professional development for preschool teachers in effective early childhood development and education implementation emphasized the importance of teacher preparedness in terms of professional records preparation, academic and professional training levels of the preschool teachers for effective early childhood development and education implementation. However, Lin and Magnuson's (2018) results from hierarchical linear models indicate that teachers' education does not predict children's early academic skills. On the question of hours spent in preschool education, for a demographically comparable group of children Leow and Wen (2017) found no significant differences between children who attended the Head Start half-day program, and those enrolled on the more intensive full-day program in any of the five academic and social outcome measures examined. We are unaware of any studies that associate characteristics of staff age or type of contract with effectiveness in preschool education.

A similar picture emerges for service organization. Several studies showed that lower child-to-teacher ratios and smaller group sizes were related to higher overall process quality (Barros and Aguiar, 2010; Mashburn, 2008; Phillips et al., 2000). However, other studies found no such associations (Pessanha et al., 2007; Slot et al., 2015).

The only conclusive results are found in environmental characteristics. Sheridan et al. (2014) concluded that rural children experienced greater difficulties with parent-reported externalizing behaviors. Li et al. (2019) obtained similar results; their findings are used to argue that high quality preschool education serving rural children should be prioritized to narrow the achievement gap between rural and urban children. On the question of different ethnicities in the classroom, most research has shown that process quality is lower in classes with higher proportions of ethnic minority children; Kuger et al. (2016) and Slot et al. (2015) analyzed this question in Germany, for instance.

In sum, despite the limited evidence, there is initial support for the indirect relationship of structural teacher and classroom characteristics with children's outcomes through process quality, and strong evidence of environmental characteristics as factors of successful preschool education. Therefore, after measuring efficiency and effectiveness in the first stage, we conducted a second stage analysis to evaluate their determinants. Several methodologies are available for this process, some of which have been reviewed by Léopold Simar and Paul Wilson in a series of papers (Simar and Wilson, 2007, 2011) and, more recently, by Bădin et al. (2014). Other contributions have reviewed the specific literature dealing with two-stage approaches in

contexts such as local government (Narbón-Perpiñá and De Witte, 2018; Aiello and Bonanno, 2018a), or banking (Aiello and Bonanno, 2016, 2018b).

One of the main factors motivating these studies, particularly those by Simar and Wilson (2007, 2011) and Bădin et al. (2014), is related to the problems derived from combining linear programming problems in the first stage of the analysis with parametric methods such as logistic regression or OLS in the second one—given that the scores obtained in the first stage are dependent in the statistical sense. Some proposals attempting to avoid these problems have been developed by Simar and Wilson (2007) and Banker and Natarajan (2008), among others.

In this case we will consider a different non-parametric option, namely, a decision tree mathematical model. The decision tree algorithm helps to classify a sample according to binary (the so-called classification trees) or continuous variables (the so-called decision trees). The decision tree is one of the best known data mining algorithms and serves to create knowledge by analyzing the characteristics of the sample under study. As they mimic human thinking, they are easy to understand, which helps interpretation of the results. In non-parametric efficiency evaluation, decision tree methods have been used in Sohn and Moon (2004), Wu (2009) and Emrouznejad and Anouze (2010).

The main objective in this section is to identify the structural variables of kindergartens that determine the composition of homogeneous preschool groups according to their effectiveness. In our case, the dependent variable to be explained in the decision tree algorithm corresponds to the effectiveness indicator of model (3) (i.e., model $AR - CDEA - BoD$). In this model we use the AID (automatic interaction detection) algorithm since it allows the dependent variable to be continuous (Breiman et al., 1984). The independent variables correspond to the second stage variables indicated above and are described in Table 2.

As shown in Figure 1, kindergartens can be grouped into four segments (nodes 2, 4, 5 and 6) according to their level of effectiveness. From this segmentation we can conclude that the first determining factor in the observed differences in effectiveness corresponds to the number of levels (variable of service organization), and that accounts for the size of the kindergarten. According to this result, kindergartens with more than six levels (large) obtain a lower effectiveness indicator, which on average for this segment corresponds to 65.42%, lower than the average of the total sample (70.50%). This conclusion supports the thesis in the literature on effectiveness in preschool education that size matters, in this case, the size of the kindergarten. This segment represents 10.3% of the total kindergartens in the sample.

From this first classification, node 1 (those with six or fewer education levels) is divided

among kindergartens serving populations of children from families with a monthly income of less than and greater than CH\$335,402 (approximately US\$500) (environmental variable). In contrast to findings in the literature, the kindergartens with lower average family incomes are more effective than those serving higher income families. This may be due to two factors. First, the mission of the Institution that oversees these kindergartens is to work with low-income families; as a result, they are specialized in working with low-income populations and are therefore more effective in caring for them. Second, the effectiveness index includes an indicator of family satisfaction, which reflects the higher demands from families with higher incomes who give lower scores to the kindergartens their children attend. Based on the above, we find a second segment with a lower degree of efficiency corresponding to small kindergartens (with six or fewer levels) with an above-average socioeconomic level served by the INTEGRA Foundation and comprising 17.7% of all the kindergartens.

Similarly, starting with node 3 (representing 72.1% of kindergartens) with an average efficiency level higher than the national average that serves families with lower income, the last two segments of kindergartens are extracted, according to the rural or urban location of the kindergarten (environmental variable). One segment, which performs better, with an average efficiency of 72.53% corresponding to those small kindergartens, low-income families and located in urban areas, and corresponding to 57.2% of kindergartens in the sample. And a last segment, of deficient performance, that represents 14.9% of the sample and that corresponds to small kindergartens, of low socioeconomic level, but located in rural areas. This result is also in line with the literature and points to the need for greater attention to kindergartens located in rural areas.

5. Conclusions and discussion

The importance of early childhood education has been extensively documented from several perspectives and in various knowledge fields, including—apart from education—psychology, economics and less closely related areas such as law (in general) and criminology (in particular), medicine or biology. This area of research has produced some important contributions published in leading general science journals (e.g., Campbell et al., 2014).

In the field of economics, important recent contributions are those by Havnes and Mogstad (2011), Duncan and Magnuson (2013), Heckman et al. (2013), García et al. (2019) and Rosales-Rueda and Triyana (2019), among others. However, some particular issues within this field

remain unexplored. Specifically, whereas a substantial body of the education literature has evaluated either the efficiency or effectiveness with which educational centers operate in primary, secondary or higher education levels, there are no similar initiatives for early childhood education, a gap that can be found regardless of the context examined.

This article aimed to bridge that gap. The study contributes on two broad levels: methodological and empirical. From a methodological point of view, it follows the line of work proposing a composite indicator using data envelopment analysis (DEA), but adds the concepts of assurance region and centralized DEA to develop a methodological innovation that we call composite indicators with assurance region centralized data envelopment analysis. This methodological proposal addresses and overcomes the main criticisms of composite indicators relating to: (i) the normalization of data; (ii) the subjectivity and arbitrariness of weights; (iii) the existence of weights equal to zero; and (iv) the distance of the solution to a desired theoretical framework. It is particularly interesting to note that when the units to be evaluated belong to the same organization, there is a higher interest in maximizing the efficiency of the system, or simply finding a set of weights that fit within a social or corporate ideal, and that shows the best facet of each organizational unit in this pre-established framework. Other advantages of the proposed model are its simplicity and the ease with which it is accepted by the agents or organizations under evaluation.

The article contributes empirically by evaluating the effectiveness of kindergartens, a level of schooling that has not been commonly considered in this type of study, unlike schools and universities, and that represents a very high social and economic value for any society. For this purpose, we used a sample of 712 preschool establishments belonging to the same institution in Chile. Additionally, unlike the vast majority of previous studies in the educational area, we considered not only the results of academic achievement as antecedents for the conceptualization of a good educational establishment, but also the perceptions of the user families and the workers.

The analysis shows, in successive stages, the refinement of the model, and allows the comparison of results. A first “benefit of the doubt” model shows an average efficiency of 84.47%, with approximately 3% difference between rural and urban kindergartens. In the same way, and as we predicted when proposing this new methodology, the weights of the three dimensions are very significantly different from any criterion of acceptable suitability—with average values of 7% for pupils’ learning, 33% for user family satisfaction, and 61% for job satisfaction. Likewise, a large amount of weights equal to zero was obtained, which also moves away from

any social or organizational ideal.

A second assurance region model incorporates restrictions from a panel of experts setting limits for these weights according to professional and academic criteria. The results of this model show a significant fall in the efficiency of the system to 71.03% but provide weights of the different dimensions according to a desired or normative framework. In this way, the weights of each dimension change significantly, with an average weight of 48% for pupils' learning, 29% for user satisfaction and 32% for job satisfaction. Despite this, the results suffer from not being unique weights, which is why a certain degree of unsought arbitrariness persists.

The third model incorporates the additional restriction of looking for a set of unique weights with which to evaluate all units, but restricted to maximizing the effectiveness of the system as a whole. In this case, the efficiency index decreases slightly to 70.54%, with significant heterogeneities between regions of the country. Likewise, the centralized (unique) weights obtained are 45% for learning, and 32% for user satisfaction and job satisfaction.

A second stage analysis using hierarchical classification establishes the existence of four types or segments of kindergartens according to their effectiveness (centralized model). A first point to emphasize is that this classification reveals the importance of three structural factors that mark the effectiveness of kindergartens: (i) size of the kindergarten, measured as the number of levels it serves; (ii) family income; and (iii) urban or rural location of preschool education establishments.

From this segmentation, we should be particularly concerned with the largest kindergartens, and prioritize the implementation of preschool programs (kindergartens) with fewer levels, in order to facilitate better management. Secondly, the differences among kindergartens with children from families of both higher and lower socioeconomic levels also calls for further analysis; this would enable the implementation of differentiated pedagogical and managerial practices beyond their specialization in very low income levels. Finally, it would also be useful to focus on rural kindergartens, in order to bridge the gap between the performance of urban and rural centers.

References

- Achabal, D., Heineke, J. M., and McIntyre, S. H. (1984). Issues and perspectives on retail productivity. *Journal of Retailing*, 60(3):107–129.
- Aiello, F. and Bonanno, G. (2016). Efficiency in banking: a meta-regression analysis. *International Review of Applied Economics*, 30(1):112–149.
- Aiello, F. and Bonanno, G. (2018a). Explaining differences in efficiency: A meta-study on local government literature. *Journal of Economic Surveys*.
- Aiello, F. and Bonanno, G. (2018b). On the sources of heterogeneity in banking efficiency literature. *Journal of Economic Surveys*, 32(1):194–225.
- Alarcón, J., Castro, M., Frites, C., and Gajardo, C. (2015). Desafíos de la educación preescolar en Chile: Ampliar la cobertura, mejorar la calidad y evitar el acoplamiento. *Estudios Pedagógicos*, 41(2):287–303.
- Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72.
- Bădin, L., Daraio, C., and Simar, L. (2014). Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, 214(1):5–30.
- Bank, T. W. (2011). The changing wealth of nations: measuring sustainable development in the new millennium. Technical report, The World Bank, Washington, DC.
- Banker, R. D. and Natarajan, R. (2008). Evaluating contextual variables affecting productivity using Data Envelopment Analysis. *Operations Research*, 56(1):48–58.
- Barros, S. and Aguiar, C. (2010). Assessing the quality of Portuguese child care programs for toddlers. *Early Childhood Research Quarterly*, 25:527–535.
- Benito, M. and Romera, R. (2011). Improving quality assessment of composite indicators in university rankings: a case study of French and German universities of excellence. *Scientometrics*, 89(1):153–176.
- Bowen, H. P. and Moesen, W. (2011). Composite competitiveness indicators with endogenous versus predetermined weights: An application to the World Economic Forum’s global competitiveness index. *Competitiveness Review: An International Business Journal*, 21(2):129–151.
- Breiman, I., Friedman, J., Olsen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth Statistics/Probability Series. International Group, Wadsworth, Belmont, CA, USA.
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., and Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, 343(6178):1478–1485.

- Castillo-Giménez, J., Montañés, A., and Picazo-Tadeo, A. J. (2019). Performance in the treatment of municipal waste: Are European Union member states so different? *Science of the Total Environment*, 687:1305–1314.
- Cave, A. and Mulloy, M. (2010). A Qualitative Examination of Teacher Perceptive. *National Forum of Education Administration & Supervision Journal*, 27(4).
- CEDEP (2011). Efectividad de la sala cuna de la Junta Nacional de Jardines Infantiles. Estudio longitudinal 2007–2010 Informe final, JUNJI, Santiago de Chile.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.
- Cherchye, L., Lovell, C. K., Moesen, W., and Van Puyenbroeck, T. (2007a). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, 51(3):749–779.
- Cherchye, L., Moesen, W., Rogge, N., and Van Puyenbroeck, T. (2007b). An introduction to “benefit of the doubt” composite indicators. *Social Indicators Research*, 82(1):111–145.
- Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., and Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59(2):239–251.
- Ciommi, M., Gigliarano, C., Emili, A., Taralli, S., and Chelli, F. M. (2017). A new class of composite indicators for measuring well-being at the local level: An application to the Equitable and Sustainable Well-being (BES) of the Italian Provinces. *Ecological Indicators*, 76:281–296.
- Connor, C. M., Son, S.-H., Hindman, A. H., and Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders’ vocabulary and early reading outcomes. *Journal of School Psychology*, 43(4):343–375.
- Consejo Asesor Presidencial para la Calidad de la Educación (2006). Final Report of the Presidential Advisory Council for Early Childhood. Technical report, Santiago de Chile. Retrieved August 27, 2019, <https://www.unicef.cl/web/wp-content/uploads/doc.wp/Informefinal.pdf>.
- Contreras, D. and Puentes, E. (2017). Inequality of opportunities at early ages: Evidence from Chile. *The Journal of Development Studies*, 53(10):1748–1764.
- De Witte, K. and López-Torres, L. (2019). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, forthcoming.
- De Witte, K. and Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*, 30(4):641–653.

- Despotis, D. K. (2005). A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society*, 56(8):969–980.
- Duncan, G. J. and Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, 27(2):109–32.
- Emerson, J. W., Hsu, A., Levy, M. A., de Sherbinin, A., Mara, V., Esty, D. C., and Jaiteh, M. (2012). Environmental performance index and pilot trend environmental performance index. *New Haven: Yale Center for Environmental Law and Policy*, pages 1–98.
- Emrouznejad, A. and Anouze, A. L. (2010). Data envelopment analysis with classification and regression tree—a case of banking efficiency. *Expert Systems*, 27(4):231–246.
- Emrouznejad, A. and Yang, G.-l. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, 61:4–8.
- Ertugrul Karsak, E. and Sebnem Ahiska, S. (2008). Improved common weight MCDM model for technology selection. *International Journal of Production Research*, 46(24):6933–6944.
- Färe, R. and Karagiannis, G. (2014). Benefit-of-the-doubt aggregation and the diet problem. *Omega*, 47:33–35.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Ser.A*, 120:253–281.
- Frankel, E. G. (2008). Organizational Effectiveness and Performance. In *Quality Decision Management - The Heart of Effective Futures-Oriented Management*, volume 14 of *Topics in Safety, Risk, Reliability and Quality*. Springer, Dordrecht.
- Gaaloul, H. and Khalfallah, S. (2014). Application of the “benefit-of-the-doubt” approach for the construction of a digital access indicator: A revaluation of the “digital access index”. *Social Indicators Research*, 118(1):45–56.
- Galo, J. J., Macedo, M. N., Almeida, L. A., and Lima, A. C. (2014). Criteria for smart grid deployment in Brazil by applying the Delphi method. *Energy*, 70:605–611.
- García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2019). Quantifying the life-cycle benefits of an influential early childhood program. *Journal of Political Economy*, forthcoming.
- García, J. L., Heckman, J. J., and Ziff, A. L. (2018). Gender differences in the benefits of an influential early childhood program. *European Economic Review*, 109:9–22.
- Golany, B., Phillips, F. Y., and Rousseau, J. J. (1993). Models for improved effectiveness based on DEA efficiency results. *IIE transactions*, 25(6):2–10.

- Gómez-Limón, J. A. and Riesgo, L. (2009). Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. *Journal of Environmental Management*, 90(11):3345–3362.
- Guardiola, J. and Picazo-Tadeo, A. J. (2014). Building weighted-domain composite indices of life satisfaction with data envelopment analysis. *Social Indicators Research*, 117(1):257–274.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.
- Hanushek, E. A. (1998). Conclusions and controversies about the effectiveness of school resources. *Economic Policy Review*, 4(1):11–27.
- Hatefi, S. and Torabi, S. (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics*, 70(1):114–120.
- Havnes, T. and Mogstad, M. (2011). No child left behind: Subsidized child care and children’s long-run outcomes. *American Economic Journal: Economic Policy*, 3(2):97–129.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, 1(1):1–46.
- Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.
- Hujala, E., Waniganayake, M., and Rodd, J. (2013). Cross-national contexts of early childhood leadership. In Hujala, E., Waniganayake, M., and Rodd, J., editors, *Researching Leadership in Early Childhood Education*, pages 13–30. Tampere University Press, Tampere.
- Johnes, J. (2015). Operational research in education. *European Journal of Operational Research*, 243(3):683–696.
- Johnson, A. L. and Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, 221(1):197–210.
- Joint Research Centre-European Commission (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD publishing, Paris.
- Justice, L. M., Mashburn, A. J., Hamre, B. K., and Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 23(1):51–68.

- Kao, C. (2010). Malmquist productivity index based on common-weights DEA: The case of Taiwan forests after reorganization. *Omega*, 38(6):484–491.
- Karagiannis, G. (2017). On aggregate composite indicators. *Journal of the Operational Research Society*, 68(7):741–746.
- Kuger, S., Kluczniok, K., Kaplan, D., and Rossbach, H.-G. (2016). Stability and patterns of classroom quality in German early childhood education and care. *School Effectiveness and School Improvement*, 27(3):418–440.
- Lauer, J. A., Lovell, C. K., Murray, C. J., and Evans, D. B. (2004). World health system performance revisited: the impact of varying the relative importance of health system goals. *BMC health services research*, 4(1):19.
- Lee, W.-S. and Lin, Y.-C. (2011). Evaluating and ranking energy performance of office buildings using Grey relational analysis. *Energy*, 36(5):2551–2556.
- Leow, C. and Wen, X. (2017). Is Full Day Better Than Half Day? A Propensity Score Analysis of the Association Between Head Start Program Intensity and Children’s School Performance in Kindergarten. *Early Education and Development*, 28(2):224–239.
- Li, K., Zhang, P., Hu, B. Y., Burchinal, M. R., Fan, X., and Qin, J. (2019). Testing the “thresholds” of preschool education quality on child outcomes in China. *Early Childhood Research Quarterly*, 47:445–456.
- Lien Foundation (2012). Benchmarking early education across the world. Technical report, Lien Foundation/Economist Intelligence Unit, Singapore. Retrieved August 27, 2019 from http://www.lienfoundation.org/sites/default/files/sw_report_2.pdf.
- Lin, Y.-C. and Magnuson, K. A. (2018). Classroom quality and children’s academic skills in child care centers: Understanding the role of teacher qualifications. *Early Childhood Research Quarterly*, 42:215–227.
- List, J. A., Tungodden, B., Cappelen, A. W., and Samek, A. (2019). The Effect of Early Childhood Education on Social Preferences. *Journal of Political Economy*, forthcoming.
- Lozano, S. and Villa, G. (2004). Centralized resource allocation using data envelopment analysis. *Journal of Productivity Analysis*, 22(1-2):143–161.
- Ma, X., Shen, J., Krenn, H. Y., Hu, S., and Yuan, J. (2016). A meta-analysis of the relationship between learning outcomes and parental involvement during early childhood education and early elementary education. *Educational Psychology Review*, 28(4):771–801.

- Mar-Molinero, C., Prior, D., Segovia, M.-M., and Portillo, F. (2014). On centralized resource utilization and its reallocation by using DEA. *Annals of Operations Research*, 221(1):273–283.
- Maricic, M., Egea, J. A., and Jeremic, V. (2019). A hybrid enhanced Scatter Search—Composite I-Distance Indicator (eSS-CIDI) optimization approach for determining weights within composite indicators. *Social Indicators Research*, 144(2):497–537.
- Mashburn, A. (2008). Quality of social and physical environments in preschools and children’s development of academic, language, and literacy skills. *Applied Developmental Science*, 12:113–127.
- McCartney, K., Dearing, E., Taylor, B., and Bub, K. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, 28(5–6):411–426.
- Melyn, W. and Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public economics research papers*, pages 1–24.
- Milutinović, B., Stefanović, G., Dassisti, M., Marković, D., and Vučković, G. (2014). Multi-criteria analysis as a tool for sustainability assessment of a waste management model. *Energy*, 74:190–201.
- Ministerio de Educación (2005). Estadísticas de la Educación 2015. Technical report, Ministerio de Educación, Centro de Estudios, Santiago de Chile.
- Morais, P. and Camanho, A. S. (2011). Evaluation of performance of European cities with the aim to promote quality of life improvements. *Omega*, 39(4):398–409.
- Moyer-Packenham, P. S., Shumway, J. F., Bullock, E., Tucker, S. I., Anderson-Pence, K. L., Westenskow, A., Boyer-Thurgood, J., Maahs-Fladung, C., Symanzik, J., Mahamane, S., MacDonald, B., and Jordan, K. (2015). Young Children’s Learning Performance and Efficiency when Using Virtual Manipulative Mathematics iPad Apps. *Journal of Computers in Mathematics and Science Teaching*, 34(1):41–69.
- Mutindi, K., Chephgeno, K., and Jeruto, B. (2016). Teacher Factors Affecting the Implementation of Early Childhood Development Education in Kericho Municipality, Kericho County. *Journal of Education and Practice*, 7(15):155–161.
- Narbón-Perpiñá, I. and De Witte, K. (2018). Local governments’ efficiency: a systematic literature review—part ii. *International Transactions in Operational Research*, 25(4):1107–1136.
- OECD (2015). OECD Economic Surveys: Chile 2015. Technical report, OECD Publishing, Paris.
- Park, Y., Brownell, M. T., Bettini, E. F., and Benedict, A. E. (2019). Multiple dimensions of instructional effectiveness in reading: A review of classroom observation studies and implications for special education classrooms. *Exceptionality*, 27(1):1–17.

- Peiró-Palomino, J. and Picazo-Tadeo, A. J. (2018). OECD: One or many? Ranking countries with a composite well-being indicator. *Social Indicators Research*, 139(3):847–869.
- Pessanha, M., Aguiar, C., and Bairrao, J. (2007). Influence of structural features on Portuguese toddler child care quality. *Early Childhood Research Quarterly*, 22:204–214.
- Phillips, D., Mekos, D., Scarr, S., McCartney, K., and Abbott-Shim, M. (2000). Within and beyond the classroom door: assessing quality in child care centers. *Early Childhood Research Quarterly*, 15:475–496.
- Rogge, N. (2011). Granting teachers the “benefit of the doubt” in performance evaluations. *International Journal of Educational Management*, 25(6):590–614.
- Roll, Y., Cook, W. D., and Golany, B. (1991). Controlling factor weights in data envelopment analysis. *IIE Transactions*, 23(1):2–9.
- Rosales-Rueda, M. and Triyana, M. (2019). The persistent effects of early-life exposure to air pollution: Evidence from the Indonesian forest fires. *Journal of Human Resources*, 54:1037–1080.
- Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1):65–77.
- Sheridan, S. M., Koziol, N. A., Clarke, B. L., Rispoli, K. M., and Coutts, M. J. (2014). The influence of rurality and parental affect on kindergarten children’s social and behavioral functioning. *Early Education and Development*, 25(7):1057–1082.
- Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of productive processes. *Journal of Econometrics*, 136(1):31–64.
- Simar, L. and Wilson, P. W. (2011). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, 36(2):205–218.
- Slot, P. L., Bleses, D., Justice, L. M., Markussen-Brown, J., and Højen, A. (2018). Structural and process quality of Danish preschools: Direct and indirect associations with children’s growth in language and preliteracy skills. *Early Education and Development*, 29(4):581–602.
- Slot, P. L., Boom, J., Verhagen, J., and Leseman, P. P. (2017). Measurement properties of the CLASS Toddler in ECEC in The Netherlands. *Journal of Applied Developmental Psychology*, 48:79–91.
- Slot, P. L., Leseman, P. P., Verhagen, J., and Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, 33:64–76.
- Sohn, S. Y. and Moon, T. H. (2004). Decision tree based on data envelopment analysis for effective technology commercialization. *Expert Systems with Applications*, 26(2):279–284.

- Strasser, K., Lissi, M. R., and Silva, M. (2009). Gestión del tiempo en 12 salas chilenas de kindergarten: Recreo, colación y algo de instrucción. *Psyche*, 18(1):85–96.
- Stumbriene, D., Camanho, A. S., and Jakaitiene, A. (2019). The performance of education systems in the light of Europe 2020 strategy. *Annals of Operations Research*, pages 1–32.
- Sun, Y., Correa, M., Zapata, A., and Carrasco, D. (2011). Resultados: Qué dice la evaluación docente de la enseñanza en Chile. In Manzi, J. and González, R., editors, *La evaluación docente en Chile*, pages 91–136. MIDE UC, Santiago de Chile.
- Taguma, M., Litjens, I., and Makowiecki, K. (2012). Quality Matters in Early Childhood Education and Care: Finland 2012. Quality matters in early childhood education and care, OECD, Paris.
- Thompson, R. G., Singleton Jr, F., Thrall, R. M., and Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in Texas. *interfaces*, 16(6):35–49.
- Verschelde, M. and 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.
- Wang, H. (2015). A generalized MCDA–DEA (multi-criterion decision analysis–data envelopment analysis) approach to construct slacks-based composite indicator. *Energy*, 80:114–122.
- Wilcox-Herzog, A. (2004). Actions speak louder than words: How experience and education relate to teachers’ behaviors. *Journal of Early Childhood Teacher Education*, 25(1):11–18.
- Wu, D. (2009). Supplier selection: A hybrid model using dea, decision tree and neural network. *Expert Systems with Applications*, 36(5):9105–9112.
- Zanella, A., Camanho, A., and Dias, T. G. (2013). Benchmarking countries’ environmental performance. *Journal of the Operational Research Society*, 64(3):426–438.
- Zhou, P., Ang, B., and Poh, K. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics*, 62(2):291–297.
- Zhou, P., Ang, B., and Zhou, D. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research*, 96(1):169–181.

Table 1: Variables for the effectiveness model, descriptive statistics

Stakeholder	Variable	Indicator	Mean	Std.dev.	Min.	25% percentile	Median	75% percentile
Pupils Families Staff	Pupils' learning	Learning value added	51.17	14.74	11.46	41.07	49.62	60.53
	Family satisfaction	Full satisfaction index	69.49	17.96	4.70	58.29	69.74	83.40
	Job satisfaction	Job satisfaction index	78.92	13.92	2.03	71.87	80.54	88.65

Table 2: Determinants of effectiveness, descriptive statistics

Variable	Variable description	Categories	Average	Min.	25% per-centile	Median	75% per-centile	Max.	Total
<i>Manager_age</i>	Managers' age (average)		41.5	26.0	35.0	41.0	47.0	69.0	
<i>Teachers_age</i>	Teacher' age (average)	34.1	24.0	30.0	33.0	37.0	60.0		
<i>Technical_staff_age</i>	Technical staff's age (average)		38.1	23.0	34.0	38.0	42.0	51.0	
<i>TP_dir</i>	Head with university degree	0 = NO 1 = YES							153.0 548.0
<i>%_TP_educ</i>	% of teachers with university degree		80.5	0.0	60.0	100.0	100.0	100.0	
<i>%_CF_educ</i>	% of teachers with a permanent contract		78.2	0.0	50.0	100.0	100.0	100.0	
<i>%_JC_educ</i>	% of full time teachers		95.1	50.0	100.0	100.0	100.0	100.0	
<i>%_CF_tecn</i>	% of technical staff with a permanent contract		97.5	0.0	100.0	100.0	100.0	100.0	
<i>%_JC_tecn</i>	% of full time technical staff		82.3	40.0	60.0	100.0	100.0	100.0	
<i>Tipo_JI</i>	Type of kindergarten	Kindergarten and crèche Kindergarten only Crèche only							486.0 212.0 14.0
<i>Extra_timetable</i>	Kindergarten with extra timetable	0= NO 1= YES	0.0						177.0 535.0
<i>Start_year</i>	Year when kindergarten started		1991	1970	1983	1991	1998	2010	
<i>#levels</i>	# of kindergarten levels		1.0	2.0	3.0	5.0	12.0		
<i>N_educ_100n</i>	# of teachers as a percentage of pupils	3.7	0.3	0.0	0.0	0.0	0.0	4.0	
<i>N_tecn_100n</i>	# of technical staff as a percentage of pupils		12.4	4.0	8.0	10.0	18.0	25.0	
<i>N_serv_100n</i>	# of administration (service) staff as a percentage of pupils		3.5	2.0	2.0	4.0	4.0	5.0	
<i>Decile_revenues Urb_rural</i>	National family decile income Kindergarten location	Rural Urban	2.5	1.0	2.0	2.0	3.0	4.0	126.0 586.0
<i>%_malnutrition_risk</i>	Percentage of children at risk of malnutrition		1.9	0.0	0.0	0.0	2.0	18.0	

Table 3: Experts' constraints, scores and descriptive statistics

Variable	Expert									Min.	Max.	Average
	1	2	3	4	5	6	7	8	9			
Pupils' learning	50	45	40	35	45	50	60	50	33.3	33.3	60	45.4
Families' satisfaction	30	20	30	35	35	30	30	25	33.3	20	35	29.8
Job satisfaction	20	35	30	30	20	20	10	25	33.3	10	35	24.8
$u2/u1$	0.6	0.4	0.6	0.7	0.7	0.6	0.6	0.5	0.67	0.4	0.7	0.6
$u3/u1$	0.4	0.7	0.6	0.6	0.4	0.4	0.2	0.5	0.67	0.2	0.7	0.5

To define the model weights' constraints, we used judgments from nine experts, of whom five were researchers in preschool education, and the rest were members of the INTEGRA Foundation's management team. Each of these nine experts responded to the following questions: (i) "which of the following three characteristics should be found in a kindergarten?"; and (ii) "if the three characteristics are included in the definition, how important (in percentage terms) should each one of them be?". The three variables were human capital, service organization and environment.

Table 4: Effectiveness results. Model DEA “benefit of the doubt”

Region	Effectiveness (%)				Weights added value (V_1)			Weights family satisfaction (V_2)			Weights staff satisfaction (V_3)		
	Rural	Urban	Average	Rural	Urban	Average	Rural	Urban	Average	Rural	Urban	Average	Rural
Antofagasta	82.04	82.58	82.49	0.00	0.05	0.04	0.50	0.38	0.40	0.50	0.58	0.57	0.50
Araucanía	81.22	86.05	85.62	0.24	0.11	0.12	0.05	0.34	0.32	0.73	0.57	0.59	0.32
Arica y Parinacota	78.82	86.64	85.66	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00
Atacama	83.50	77.24	77.86	0.00	0.09	0.08	0.50	0.65	0.64	0.50	0.27	0.29	0.64
Aysén	93.68	88.12	90.90	0.00	0.03	0.01	0.00	0.49	0.24	1.00	0.51	0.75	0.24
Bio-Bio	90.31	85.67	86.48	0.18	0.13	0.14	0.23	0.24	0.24	0.61	0.65	0.65	0.24
Coquimbo	90.08	81.99	83.96	0.01	0.03	0.03	0.26	0.28	0.28	0.74	0.69	0.70	0.28
O'Higgins	86.14	85.15	85.45	0.02	0.11	0.08	0.39	0.39	0.39	0.60	0.52	0.54	0.39
Los Lagos	85.05	87.00	86.60	0.06	0.03	0.04	0.12	0.36	0.31	0.84	0.62	0.67	0.31
Los Ríos	—	86.48	86.48	—	0.09	0.09	—	0.32	0.32	—	0.60	0.60	0.32
Magallanes y Antártica	—	86.63	86.63	—	0.00	0.00	—	0.28	0.28	—	0.72	0.72	0.28
Maule	88.84	84.79	86.25	0.01	0.06	0.04	0.44	0.35	0.38	0.56	0.60	0.59	0.38
Región Metropolitana	76.24	78.07	77.94	0.01	0.06	0.05	0.58	0.35	0.37	0.42	0.61	0.59	0.37
Tarapaca	83.12	86.08	85.24	0.38	0.06	0.16	0.04	0.09	0.08	0.60	0.85	0.78	0.08
Valparaíso	90.86	86.65	87.49	0.11	0.04	0.06	0.29	0.35	0.34	0.62	0.62	0.62	0.34
Average	87.08	83.91	84.47	0.08	0.07	0.07	0.29	0.33	0.32	0.64	0.61	0.62	0.32
Std. Dev.	12.01	10.03	10.47	0.21	0.20	0.20	0.41	0.41	0.41	0.44	0.43	0.43	0.41
Max	100.00	100.00	100.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Min	49.25	52.07	49.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: the authors’.

Table 5: Effectiveness results. Model DEA “benefit of the doubt” with assurance region

Region	Effectiveness (%)			Weights added value (V_1)			Weights family satisfaction (V_2)			Weights staff satisfaction (V_3)		
	Rural	Urban	Mean	Rural	Urban	Mean	Rural	Urban	Mean	Rural	Urban	Mean
Antofagasta	61.06	67.43	66.27	0.48	0.47	0.48	0.26	0.31	0.30	0.34	0.31	0.31
Araucanía	72.90	76.65	76.32	0.50	0.48	0.48	0.28	0.30	0.30	0.31	0.31	0.31
Arica y Parinacota	67.34	67.47	67.45	0.51	0.49	0.49	0.21	0.25	0.25	0.36	0.34	0.34
Atacama	56.49	66.42	65.43	0.53	0.48	0.48	0.30	0.32	0.31	0.24	0.30	0.29
Aysen	65.04	78.56	71.80	0.48	0.47	0.48	0.26	0.30	0.28	0.34	0.32	0.33
Bio-Bio	74.28	73.86	73.94	0.50	0.48	0.49	0.29	0.29	0.29	0.31	0.32	0.31
Coquimbo	63.82	63.91	63.89	0.48	0.47	0.48	0.27	0.31	0.30	0.34	0.30	0.31
O'Higgins	74.76	71.31	72.35	0.48	0.48	0.48	0.28	0.29	0.29	0.33	0.32	0.32
Los Lagos	65.63	73.68	72.02	0.49	0.47	0.48	0.26	0.30	0.29	0.34	0.32	0.32
Los Ríos	—	74.62	74.62	—	0.48	0.48	—	0.30	0.30	—	0.31	0.31
Magallanes y Antártica	—	69.44	69.44	—	0.46	0.46	—	0.31	0.31	—	0.32	0.32
Maule	72.59	72.39	72.46	0.48	0.47	0.47	0.29	0.30	0.30	0.31	0.32	0.32
Región Metropolitana	64.18	65.24	65.17	0.48	0.48	0.48	0.32	0.29	0.29	0.29	0.32	0.32
Tarapaca	58.79	70.44	67.12	0.54	0.48	0.49	0.25	0.29	0.28	0.29	0.33	0.32
Valparaíso	76.42	73.95	74.44	0.48	0.47	0.47	0.29	0.30	0.30	0.31	0.31	0.31
Mean	70.10	71.23	71.03	0.49	0.48	0.48	0.28	0.30	0.29	0.32	0.32	0.32
Std. Dev.	12.02	9.67	10.13	0.04	0.04	0.04	0.06	0.05	0.05	0.06	0.04	0.04
Max	100.00	100.00	100.00	0.67	0.67	0.67	0.39	0.39	0.39	0.36	0.36	0.36
Min	41.48	40.57	40.57	0.45	0.45	0.45	0.21	0.21	0.21	0.11	0.11	0.11

Source: the authors’.

Table 6: Effectiveness results. Model DEA “benefit of the doubt” with assurance region and unique weights (centralized DEA)

Region	Effectiveness (%)		
	Rural	Urban	Average
Antofagasta	59.66	67.24	65.87
Araucania	72.36	76.27	75.93
Arica y Parinacota	66.11	66.55	66.49
Atacama	53.08	66.06	64.76
Aysen	63.92	78.47	71.19
Bio-Bio	73.37	73.34	73.35
Coquimbo	63.41	63.42	63.41
O’Higgins	74.39	70.83	71.90
Los Lagos	64.39	73.38	71.52
Los Ríos	—	74.38	74.38
Magallanes y Antártica	—	69.39	69.39
Maule	71.66	72.05	71.91
Región Metropolitana	63.84	64.83	64.76
Tarapaca	56.46	69.78	65.97
Valparaíso	75.04	73.65	73.93
Average	69.14	70.84	70.54
Std. Dev.	12.66	9.74	10.33
Max.	100.00	99.73	100.00
Min.	39.56	40.57	39.56

Source: the authors’.

Table 7: Centralized weights. Model DEA “benefit of the doubt” with assurance region and unique weights (centralized DEA)

Dimension	Rural	Urban	Average
Pupils’ learning	0.45	0.45	0.45
Families’ satisfaction	0.32	0.32	0.32
Job satisfaction	0.32	0.33	0.32

Source: the authors’.

Figure 1: Results for decision trees

