

# Optimization modeling for resource allocation in the Chilean public educational system

Jorge Amaya, Dominique Peeters, Paula Uribe, Juan P. Valenzuela

# Optimization modeling for resource allocation in the Chilean public educational system \*

Jorge Amaya<sup>†</sup>      Dominique Peeters<sup>‡</sup>      Paula Uribe<sup>§</sup>      Juan P. Valenzuela<sup>¶</sup>

## Abstract

This paper presents a decision model for planning in public education, without considering quality constraints. Given a set of established schools in a municipality, the aim of the model is to find the optimal supply in terms of which schools should be kept open or closed, which grades (or year levels) should be made available and the number of classes that should be provided within each such grade to meet student demand. The model includes specific information about curriculum requirements in terms of the number of hours per year per subject to be provided for each grade level. The objective function essentially seeks to minimize the total fixed and variable costs to schools, which mainly relate to human resources (principal and direction, teaching, administrative and support staff). We propose an integer optimization model for solving this problem and we apply it to two relatively large municipalities in Chile.

**Key words:** Education planning, Optimization models.

---

\*Supported by CONICYT, under projects BASAL CMM-Univ. de Chile and FONDEF VRI-V10P0021.

<sup>†</sup>Centro de Modelamiento Matemático and Departamento de Ingeniería Matemática, Universidad de Chile

<sup>‡</sup>Center for Operations Research and Econometrics, Université Catholique de Louvain

<sup>§</sup>Centro de Modelamiento Matemático, Universidad de Chile

<sup>¶</sup>Centro de Investigación Avanzada en Educación and Departamento de Economía, Universidad de Chile

# 1 Introduction

In Chile, one of the main challenges for increasing competitiveness and advancing development of the country is to improve the standard and quality of education, particularly in the public education system. The public sector represents around 40% of total primary and secondary education infrastructure and, in general, caters to the poorest segment of the young population.

The Chilean system of elementary and high-school education can be decomposed into three well-identified sectors: the public sector, managed by local authorities (municipalities); the private subsidized system, managed by private sponsors with partial funding from the state; and the private non subsidized sector, managed by private enterprise which derives economic benefit from educational activity. In Chile, there is almost no restriction on the establishment of private schools in any geographic region. The private subsidized system is distinguished by a specific agreement between a private sponsor and the national Ministry of Education. Schools participating in this system are allowed to impose relatively moderate fees on parents and a complementary subsidy which is allocated by the state. All three systems are competitive, seeking to attract the best students, but the public sector is completely free and open, in the sense that all candidates must be admitted and if there is no place available at a given public school in the district, then the authority must look for places at another school in the public system, even outside the municipality. In general, families choose schools in the vicinity of their homes, especially for primary school. In addition, there is no gender segregation, most schools being mixed, and racial balance is not an issue in the country.

The Ministry of Education only provides general regulations for managing the Chilean educational system. These are in the form of a minimal standard for the teaching curriculum which schools are required to meet or exceed. In the particular case of the public sector, the state, throughout federal, regional and local levels, also provides economic resources, but the rest of the responsibility falls on the local authorities. These resources are often limited, which is especially severe, because (according to the law) the public system must satisfy current demand and no financial contribution may be asked of students' families<sup>1</sup>. In this context, the optimal use of public resources becomes a critical concern in each municipality.

During the last ten years, household income in the country has experienced a systematic increase. This has caused a fall in demand for public education as higher incomes have brought previously unaffordable private subsidized sector schooling within the reach of many more families, also demographic changes in the whole country are reducing total population at schooling age, affecting both public and private schooling services. Both sources are producing an annual 2% reduction of public schools market share on national educational system ([3]). Management of public resources is also strictly limited by legal regulations, particularly in terms of human resources; for instance,

---

<sup>1</sup>Public schools cannot charge fees to students' families before high school. Moreover when some public high schools use charges to families, they are very small comparing those fees charged by private subsidized schools ([12]).

teachers having tenure positions must receive higher wages but transfers from central government are not differentiated to afford this larger costs. This represents a real challenge for the public sector: Municipalities must propose an attractive program to students and their families, while at the same time using public funds in an efficient way.

The problem we address in this paper can be described as follows: Given the set of existing public schools in a municipality and knowing the expected demand for the next school year, the aim is to decide how to structure the supply of student places and classes for one-year planning purposes. This means that the model must decide which schools must be kept open, the grade levels to be offered by each school and the number of parallel classes within each grade. The main constraints are essentially the satisfaction of demand, infrastructure availability (number of classrooms in each school) and compliance with minimal teaching curriculum requirements, expressed in terms of the number of teaching hours per week for each subject. This last variable is important because it determines the human resources that need to be allocated, which represents a very important part of the total budget of the municipality.

The location of schools can have a great economic and social impact, particularly in countries having centralized public education systems. The criteria for the location, size or structuring of schools can vary due to factors ranging from spatial or demographic considerations to social integration and administrative or economic constraints. Different authors have proposed various models, depending on particular state regulations and main objectives. For example, a dynamic multi-period model is proposed in [1] for the evolution of school networks, allowing size reduction and/or expansion within certain pre-defined facility size limits. This generates a relatively complex mixed linear programming program, whose variables represent assignment, location and capacity decisions. The objective is to minimize the total discounted cost. In [2], the same authors use the simulated annealing heuristic to tackle this kind of problem and report good behavior of the strategy in real applications.

In 1993, Church and Murray ([4]) proposed a model to address the problem of school consolidation. Their work is about the assignment of students to schools with constraints that take into account racial issues and school balancing. In the situation we study here, we select schools to remain in operation and decide how many classes of a given grade will be made available, looking to meet demand for student places, but we only know the maximum capacity of each grade/school, not the exact number of students assigned to each grade. Another difference is that we also consider the balancing of classes within each ‘cycle’ for each school, but not the balance between schools as they do in their paper (in the Chilean system, primary education is partitioned into two so-called ‘cycles’ which will be described later). Church and Murray also consider a fixed number of schools to operate, while in our case, this decision is made by the model. Also, our objective function considers the cost of operating a school, unlike in their work where the monetary cost of the system is fixed, and they endeavor to use resources in a balanced way.

In [5], the authors proposed a decision support system for the districting problem, which uses

distances between the student population and schools as the main criteria for assignment. Also, reference [6] deals with the problem of school redistricting in which certain, stipulated blocks of the region must be assigned to establishments, combining the use of a binary decision model with a geographic information system, and obtaining encouraging practical results in real-world case studies. A simple binary optimization model is used in [7] to assign students to subjects at university level, but in that case the criteria are essentially the preferences of the students. In [8], the authors deal with the existence of contradictory interests of all parties (administrators, teachers, parents, students, etc.). They propose using multi-criteria models for spatial allocation of educational resources, and give some critical comparison of the different multi-criteria techniques that can be used: Pareto optimality, data envelopment analysis, parametric programming and goal programming methods. In [9], the authors focus on the location problem, for the real case of two counties in the state of Rio de Janeiro, using heuristic procedures to solve the *p-median* problem. The criterion is to minimize the expected sum of overall home-school distances of students. The problem of assigning students to schools, considering both capacity and racial balance, and minimizing the total distance from houses to schools, is addressed in [10]. The resulting integer linear problem is solved by solving a sequence of network instances. By using a variant of the *p-median* model, in [11], the authors propose to maximize the accessibility of students to schools (through minimization of total distances) and they apply the resulting facility location model to real case studies.

This paper is organized as follows: In Section 2 we give the main principles for the planning model and in Section 3 we define the variables, parameters and equations describing the optimization model. Section 4 is devoted to the application of the model to two real case studies and, in Section 5, we give some general conclusions and ideas for further research.

## 2 The conceptual model

We begin by describing the main constraints of the system, most of which stem from legal and structural regulations. First, the total demand for each grade must be met (the public sector cannot deny enrollment to students). Second, the minimum number of hours per week of each subject is essentially imposed by the academic program established by the state authority. Furthermore, the specific academic plan of a school may fine-tune this number. Third, the maximum number of students in each class is limited by law, but a lower limit could be imposed depending on the education level, on the infrastructure available at the school and on preferences of local authorities related to improve average quality for a particular school or the whole local public educational system. Quality of education is not explicitly included as an endogenous factor in our model.

Our purpose is to design optimization models suggesting which schools should remain open in a given area and, for each of these schools, the forecasted number of students at each level, so that we can determine the cycles that are to be set up, the number of classes for each grade

that are to be created, how many teachers should be recruited, etc. The objective is to provide a given level of service at minimal cost. Practically, we have to cope with an existing situation and to consider the possibility of closing/consolidating schools. Theoretically, one could merge any set of schools to optimize resources. Practically, this is not always possible due to geographic restrictions, population distribution and quality heterogeneities, among other factors. In order to implicitly take those restrictions into account, the model considers a partition (or zonification) of the geographic area of interest, each zone containing a certain number of schools that can be merged without restrictions. This must be based on location criteria and satisfy local demand. In our two particular applications those partitions are defined by the decision-maker.

The Chilean education system is made up of four cycles as follows:

- i) Pre-school: 2 years (children being 4 and 5 years of age),
- ii) Primary1: 4 years,
- iii) Primary2: 4 years, and
- iv) High school: 4 years.

The inputs to our model are essentially the resources, school capacities, costs and demands. This means the existing number of classrooms (infrastructure availability), unitary costs of teaching and other personnel (variable), other fixed and variable costs of operating each school, capacities (how many students can be admitted to each class), and the total capacity of the school in each grade. Our first two versions of the model enforce satisfaction of expected demand for each school and grade, but consider neither the random behavior of the demand nor its dynamic evolution over time.

The main assumptions of our model are:

- i) The scope of the proposed model is one year, which means that all data are annual and the output is designed to decide on one year operation.
- ii) The total demand is known for each grade and geographic zone. In the first version, this demand must be satisfied by schools belonging to the same corresponding one, but the two other version relax this condition to enable a more realistic representation of the problem (schools can satisfy demand coming from other zones).
- iii) If the model decides that a given school should be kept, at least one cycle must be made available, but if two non consecutive cycles are to be provided, then any intermediate cycles must also be provided (for example, if the first primary cycle and high school cycle are opened, then the second primary cycle must also be available in order to offer continuity to students completing the first primary cycle).
- iv) We also impose an upper limit (in the form of a specified constant) on the number of classes within every grade and, if a grade is to be made available, then the entire corresponding cycle

must also be available.

- v) On the other hand, it is required that the number of students per classroom be smaller than a specified limit, and this limit being fixed according to pedagogic and infrastructure criteria. The classmate size is usually related to quality of education, but we do not explicitly include quality issues in our formulation.
- vi) The model also requires that the total number of classes in a given school be restricted to the available infrastructure, i.e., the number of classrooms.
- vii) Each school must have a coherent distribution of students and grades; this means that the number of parallel classes in the same grade must be as similar as possible for every grade throughout the same cycle. Note that we impose balancing inside a school and not between schools.
- viii) Finally, concerning educational requirements, it is necessary to comply with the total number of teaching hours per week for every grade and subject. This is dictated by national regulations and each school's specific program.

### 3 The mathematical model

In this section, we describe three embedded versions of the model that we propose for this planning problem. The first one simply seeks to minimize the total cost, satisfying the structural and legal regulations of the Chilean educational system. The second one introduces the notion of student allocation to schools for every grade by including the corresponding costs of those allocations in the objective function. In this case, we introduce a new variable that represents the number of students belonging to a given zone and grade allocated to a specific school. Finally, the third model relaxes the constraint on satisfying demand, which permits to deal with the eventual overestimation in the pre-enrollment process.

#### 3.1 First version: minimal cost of the system

We begin by presenting the mathematical model that has been developed for assigning the demand for education services to the different offer alternatives. Let us define the following variables (that

characterize a decision):

- $x_{ik}$  : Number of classes in grade  $i = 1, \dots, N_G$  at school  $k = 1, \dots, N_S$
- $y_{ik}$  : Binary, 1 if grade  $i$  is available at school  $k$ , 0 if not
- $z_k$  : Binary, 1 if school  $k$  is open, 0 if not
- $u_{jk}$  : Binary, 1 if cycle  $j = 1, \dots, N_C$  is to be provided by school  $k$ , 0 otherwise
- $v_{mik}$  : Number of hours per year assigned to subject  $m = 1, \dots, N_T$ , in grade  $i$  at school  $k$
- $W_{jk}$  : Maximum difference between the number of classes within cycle  $j$  at school  $k$

We also define the parameters (problem data), which can be classified in different categories: Financial, physical, legal regulations data, etc.

- $w_k$  : Fixed annual cost of school  $k$
- $q_{mi}$  : Teaching cost per hour for subject  $m$ , grade  $i$
- $t_{ik}$  : Capacity (maximum number of places per class) of grade  $i$  at school  $k$
- $T_k$  : Number of available classrooms at school  $k$  (capacity of the school)
- $H_{mik}$  : Number of hours of subject  $m$  required per year in grade  $i$  at school  $k$
- $S_{ik}$  : Maximum number of classes in grade  $i$  at school  $k$ , if school  $k$  is open
- $s_{ik}$  : Minimum number of classes in grade  $i$  at school  $k$ , if school  $k$  is open
- $\{C_j\}$  : Partition of grades into cycles (pre-school, primary1, primary2, high school)
- $\{Z_h\}$  : Partition of the geographic area where schools are located into zones.
- $d_{ih}$  : Demand for grade  $i$  in zone  $h = 1, \dots, N_Z$
- $p_{jk}$  : Penalty coefficient for heterogeneity within cycle  $j$  at school  $k$

With these definitions, we can now specify the constraints of the optimization problem to be solved and the corresponding objective function. A linear-integer-binary optimization problem can be written, whose solution yields the optimum configuration for a set of schools, so as to satisfy the annual education demand for a given student population.

The objective function seeks to minimize total fixed and variable school operating costs, mainly related to human resources (teachers, administrative and support staff). The first term of the objective function represents the minimization of the total cost of teaching hours. The second term seeks to minimize the number of schools to be open, while the third term promotes the creation of a similar number of parallel classes in each grade throughout a cycle at each school. The penalty coefficient  $p_{jk}$  can also be interpreted as an indirect cost; in fact, class homogeneity is a strategy for managing the size of the student population between consecutive grades, from one year to the next. Additional costs could arise if the number of teachers to be employed were to vary between consecutive grades, increasing teacher turnover. Furthermore, this issue could be tackled in a more efficient way by considering a dynamic model (medium/long term planning over several years), but this is beyond the scope of this work.

This first version can be written as:



$$(P_0) \quad \min \sum_{mik} q_{mi} v_{mik} + \sum_k w_k z_k + \sum_{jk} p_{jk} W_{jk}$$

$$u_{jk} \leq z_k \leq \sum_{j'} u_{j'k} \quad \forall j, k \quad (1)$$

$$s_{ik} y_{ik} \leq x_{ik} \leq S_{ik} y_{ik} \quad \forall i, k \quad (2)$$

$$u_{jk} = y_{ik} \quad \forall i, j, k \text{ such that } i \in C_j \quad (3)$$

$$\sum_i x_{ik} \leq T_k \quad \forall k \quad (4)$$

$$v_{mik} \geq H_{mik} x_{ik} \quad \forall m, i, k \quad (5)$$

$$u_{jk} + u_{j'k} - 1 \leq u_{lk} \quad \forall k, 1 \leq j \leq j' - 2 \leq N_C - 2, j + 1 \leq l \leq j' - 1 \quad (6)$$

$$-W_{jk} \leq x_{ik} - x_{i'k} \leq W_{jk} \quad \forall j, k \text{ such that } i, i' \in C_j \quad (7)$$

$$\sum_{k \in Z_h} t_{ik} x_{ik} \geq d_{ih} \quad \forall i, h \quad (8)$$

$$x_{ik}, v_{mik} \geq 0, \text{ integer} \quad \forall m, i, k \quad (9)$$

$$u_{jk}, y_{ik}, z_k, \text{ binary} \quad \forall i, j, k \quad (10)$$

$$W_{jk} \geq 0 \quad \forall j, k \quad (11)$$

Constraint (1) says that a school must be open when at least one cycle is on offer at that school. It also means that when no cycles are created at a school, then the school will not be open. In this constraint, the right inequality  $z_k \leq \sum_{j'} u_{j'k}$  could be redundant, because variable  $z_k$  has a positive coefficient in the objective function, so it tends to zero when  $u_{jk} = 0$  for all  $j, k$ . But, in the solving process this constraint can help to avoid branching in those variables.

Constraint (2) says that if a given grade is opened at school  $k$ , then it must satisfy bounds on the number of classes in that grade.

Constraint (3) means that if a cycle is to be available at a school, then all grades belonging to that cycle must be provided and if one grade is to be taught, then the entire cycle must be taught.

Note that (1), (2) and (3) together imply that when a school is closed, all grades within it are closed (the  $x$  variable for each grade is zero), so the constraint is satisfied. Also, there is no incentive to change the variable value just to achieve a more balanced solution, because if the variable changes to a value greater than zero, the cost of the solution increases (because the whole school must be open if at least one grade is to be offered).

Constraint (4) gives an upper bound on the number of classes to be taught at each school.

Constraint (5) says that the number of teaching hours of the school must satisfy the requirements

given by the academic program of the school, for each subject.

Constraint (6) imposes a coherent structure, in the sense that, if two non-consecutive cycles are offered, then the intermediate cycles must also be offered.

Constraint (7) says that the number of classes in a given cycle must be quite smooth. This allows for a coherent school, in the sense that the number of parallel classes in a given grade does not differ with the number of classes in other grades in the same cycle.

Constraint (8) requires that the total capacity of the system in a given zone or area be sufficient to satisfy demand for every grade.

### 3.2 Second version: student allocation at minimal cost

The second version of the model (below) is more specific. We essentially create a new variable to represent the number of students to be allocated to each school/cycle/grade and we express the satisfaction of demand in a different manner. The model considers the cost of assignment of children to schools, but in fact only seeks to give an estimate of the total cost of the system, since student assignment to schools in Chile is not mandatory and parents have the final decision about where to send their children. Under rational assumptions, it is expected that families would tend to choose the nearest school, but in reality this is not always the case. We introduce the variables

$a_{ihk}$  : The number of children in grade  $i$  from zone  $h$  who are allocated to school  $k$ .

To enable the calculation of allocation costs, we must define the corresponding parameters

$\rho_{ihk}$  : A penalty factor associated with a child in grade  $i$  from zone  $h$  who is allocated to school  $k$ .

The main change in the constraints is the introduction of new conditions for the satisfaction of demand (19–20), which replace constraint (8).

The new version is:

$$\begin{aligned}
(P_1) \quad & \min \sum_{mik} q_{mi} v_{mik} + \sum_k w_k z_k + \sum_{jk} p_{jk} W_{jk} + \sum_{ihk} \rho_{ihk} a_{ihk} \\
& u_{jk} \leq z_k \leq \sum_{j'} u_{j'k} \quad \forall j, k \quad (12) \\
& s_{ik} y_{ik} \leq x_{ik} \leq S_{ik} y_{ik} \quad \forall i, k \quad (13) \\
& u_{jk} = y_{ik} \quad \forall i, j, k \text{ such that } i \in C_j \quad (14) \\
& \sum_i x_{ik} \leq T_k \quad \forall k \quad (15) \\
& v_{mik} \geq H_{mik} x_{ik} \quad \forall m, i, k \quad (16) \\
& u_{jk} + u_{j'k} - 1 \leq u_{lk} \quad \forall k, 1 \leq j \leq j' - 2 \leq N_C - 2, j + 1 \leq l \leq j' - 1 \quad (17) \\
& -W_{jk} \leq x_{ik} - x_{i'k} \leq W_{jk} \quad \forall j, k \text{ such that } i, i' \in C_j \quad (18) \\
& \sum_h a_{ihk} \leq x_{ik} t_{ik} \quad \forall i, k \quad (19) \\
& \sum_{k \in Z_h} a_{ihk} \geq d_{ih} \quad \forall i, h \quad (20) \\
& x_{ik}, v_{mik} \geq 0, \text{ integer} \quad \forall m, i, k \quad (21) \\
& u_{jk}, y_{ik}, z_k, \text{ binary} \quad \forall i, j, k \quad (22) \\
& W_{jk}, a_{ihk} \geq 0 \quad \forall i, j, h, k \quad (23)
\end{aligned}$$

**Remark 3.1** The coefficient  $\rho_{ihk}$  accounts for several aspects. Firstly, it may include the private or individual cost (to families), for example, transportation from area  $h$  to school  $k$  (which can be supposed to be zero for  $k \in Z_h$ ), the social perception of this travel by the families living in area  $h$ , the perception of the value of the travel time, the age of children in grade  $i$ , etc. Secondly, it may include the marginal cost to school  $k$  for admitting pupils from other zones. In this context, the objective function of  $(P_1)$  is a weighted additive combination of three criteria: direct cost of education, indirect cost due to heterogeneity within cycles and private costs to the families of students coming from other zones. Evidently, this aspect could be handled by using multi-criteria modeling and appropriate algorithms.

**Remark 3.2** In order to incorporate the competitive private system into this model, we should consider one fictitious school, representing the total effect of the private actors, which belongs to an additional, fictitious zone and which has very large allocation costs.

### 3.3 Third version: compromise between minimum cost and demand satisfaction

Finally, we can ease the requirement that total demand always be met by changing Constraint (20) into a relaxed inequality. To encourage demand to be satisfied, we add to the objective function a

term of the form  $\sum_{ih} \tau_{ih}(d_{ih} - \sum_{k \in Z_h} a_{ihk})$ , where  $\tau_{ih}$  represents the cost of not admitting a pupil in grade  $i$  from zone  $h$  to a school in his own zone.

The rest of the model remains the same as in  $(P_1)$ , by replacing Constraint (20) by

$$\lambda_{ih} d_{ih} \leq \sum_{k \in Z_h} a_{ihk} \quad \forall i, h.$$

this is a relaxation which permits only a certain given proportion  $\lambda_{ih} \in [0, 1]$  of the (local) demand in each zone to be satisfied. The model proposed below allows situations to be handled in which demand is overestimated, for example, as a result of pre-enrollment processes.

$$(P_2) \quad \min \sum_{mik} q_{mi} v_{mik} + \sum_k w_k z_k + \sum_{jk} p_{jk} W_{jk} + \sum_{ih} \tau_{ih} (d_{ih} - \sum_{k \in Z_h} a_{ihk}) + \sum_{ihk} \rho_{ihk} a_{ihk}$$

$$u_{jk} \leq z_k \leq \sum_{j'} u_{j'k} \quad \forall j, k \quad (24)$$

$$s_{ik} y_{ik} \leq x_{ik} \leq S_{ik} y_{ik} \quad \forall i, k \quad (25)$$

$$u_{jk} = y_{ik} \quad \forall i, j, k \text{ such that } i \in C_j \quad (26)$$

$$\sum_i x_{ik} \leq T_k \quad \forall k \quad (27)$$

$$v_{mik} \geq H_{mik} x_{ik} \quad \forall m, i, k \quad (28)$$

$$u_{jk} + u_{j'k} - 1 \leq u_{lk} \quad \forall k, 1 \leq j \leq j' - 2 \leq N_C - 2, j + 1 \leq l \leq j' - 1 \quad (29)$$

$$-W_{jk} \leq x_{ik} - x_{i'k} \leq W_{jk} \quad \forall j, k \text{ such that } i, i' \in C_j \quad (30)$$

$$\sum_h a_{ihk} \leq x_{ik} t_{ik} \quad \forall i, k \quad (31)$$

$$\lambda_{ih} d_{ih} \leq \sum_{k \in Z_h} a_{ihk} \quad \forall i, h \quad (32)$$

$$x_{ik}, v_{mik} \geq 0, \text{ integer} \quad \forall m, i, k \quad (33)$$

$$u_{jk}, y_{ik}, z_k, \text{ binary} \quad \forall i, j, k \quad (34)$$

$$W_{jk}, a_{ihk} \geq 0 \quad \forall i, j, h, k \quad (35)$$

**Remark 3.3** The term  $\sum_{ih} \tau_{ih} d_{ih}$  is a constant, but we decided to keep it in the objective function to highlight the fact that failing to satisfy demand comes with a cost.

## 4 Case studies

In this section, we show two real cases, corresponding to two real municipalities (Municipality 1 and Municipality 2) in Chile. Both have most population of middle and upper-middle class and a large percentage of students enrolled in private subsidized schools. Most of public schools attend poor and low-middle class students. The decision-makers decided to divide both municipalities into eight zones each, having different, known demands per grade. This choice could be delegated to the model, but we defer this approach to future study. The Municipality 1 system has 15 schools and Municipality 2 has 24 schools, with 10,320 and 19,946 students respectively. For these case studies we have imposed some realistic general hypotheses:

- i) At least one establishment must be opened in each zone (demand satisfaction).
- ii) The cost of a teaching hour is supposed to be a constant, but in general this value should depend on the subject or topic and the specific teacher conditions.
- iii) The scope of the study is for one year, that is to say, the aim is only to provide for annual planning. This means that all data (demand, costs, etc.) are annual.
- iv) As we have said previously, we consider four levels of education: Pre-school (children of 4–5 years old), primary 1 (1st to 4th grade), primary 2 (5th to 8th grade) and high school (1st to 4th grade).
- v) Although the demand is given (estimated) for each school, the model only seeks to find the optimal allocation and the satisfaction of demand for each grade in the eight predefined zones. We permit the model to decide on the schools that will effectively be open during the year and which cycles will be provided by each such school.
- vi) For each school, the model decides which cycles are to be on offer and how many parallel classes will be made available in each grade within those cycles.

We use costs and demand for the 2010 academic year and that costs are expressed in US dollars, while the demand corresponds to the number of students applying for enrollment at each school at the end of 2010. That is to say, the demand is the expected student population for 2011.

The annual fixed cost corresponds to the purchase and maintenance of infrastructure (buildings, equipment, etc.) and the total cost includes the fixed cost plus the wages of academic and non-academic staff of the school. This includes the cost of teaching hours. Rational planning based on the application of our model should essentially reduce the human resource cost (variable component) as well as the fixed cost due to establishments that are closed.

For both municipalities, we solved the three mathematical models introduced in the previous section, but for  $(P_2)$  we considered two instances:  $\lambda = 1$  and  $\lambda = 0.5$ , that is, by forcing the satisfaction

of the total demand and by relaxing this constraint to allow as little as 50% of the demand to be satisfied. All instances were run using CPLEX-9.1 on an IBM-iDataplex with 2 processors.

In these simulations we used some realistic values of the parameters:

- $t_{ik}$ : Maximum number of children per class; 35 for pre-school and 45 for the rest<sup>2</sup>.
- $S_{ik}$ : Maximum number of parallel classes in a grade; varies from 0 to 3 in Municipality 2 and 0 to 4 in Municipality 1.
- $s_{ik}$ : Minimum number of classes; this is 0 everywhere.
- $d_{ih}$ : Demand in zone  $h = 1, \dots, N_Z$  for grade  $i$ . The following tables show total demand per school.
- $H_{mik}$ : The number of teaching hours/week of subject  $m$  required for grade  $i$  at school  $k$ ; this varies from 0 to 8 in Municipality 2 and from 0 to 9 in Municipality 1. This is not constant, because it depends on each subject and on the particular pedagogic choices made by each school. In Chile, the law establishes minimum curriculum requirements for each subject, but schools can add supplementary hours at their discretion.
- $q_{mi}$ : The unit cost per hour of subject  $m$  in grade  $i$ ; this is constant at \$10.9.
- $p_{jk}$ : The cost of heterogeneity within cycle  $j$  at school  $k$ . The number of classes for each grade at every school must be similar, so we use a constant factor of \$1 to represent this weight. Another possibility is to limit the number of parallel classes by an upper bound fixed ahead of time, but in that case there is a risk of infeasibility.
- $\rho_{ihk}$ : the annual cost of education per student in grade  $i$ , zone  $h$  at school  $k$ . Here we take a constant value of  $\rho_{ihk} = \$1,163$ .
- $\tau_{ik}$ : The unit cost per annum of non-admittance is fixed at  $\tau_{ik} = \$2,114$ .

In the following subsections we show the main data and results for our two case studies. As we will explain below, the application of these mathematical models may produce a significant reduction in total costs to both municipalities. These savings come from the optimal configuration of classes (satisfying legal and institutional regulations concerning the number of students per classroom) and from the optimal allocation of teaching resources (personnel) to the teaching hours needed.

The instances generated by the four versions of the model are given in Table 1. The number

---

<sup>2</sup>It is a legal constraint but municipalities could define a lower maximum to reach better quality outcomes. Our computational results and the corresponding conclusions could dramatically be different depending on this value, but here we only used the legal bounds. Currently, average class in urban public primary schools is around 31 students ([3]).

of variables and constraints is given by the solver after preprocessing, so that the result shown is an under estimation of the problem size. These figures show that these instances correspond to medium size mixed linear programming problems, which can be easily solved by most existing commercial and academic software.

## 4.1 Main data

Tables 2 and 3 show the main data for each case study. The first column contains an identifier for the school. The second column is the current number of available classrooms in each school. In fact we consider here the classrooms as indistinguishable but it is well known that, in general, some classrooms are adapted for pre-school or primary-school pupils while others are allowed to be used by high-school students. The next column is the zone to which the school belongs (in both cases, we consider eight zones for planning purposes). The next two columns contain the total fix cost (infrastructure, general services, etc.) and the total annual cost, which includes the variable cost, namely teachers' salaries. The last column shows the total expected demand for the next year, but the calculations use the detailed distribution of the demand for every grade and school.

## 4.2 Simulations results

The results are reported in Table 4. The school that is eliminated by  $(P_0)$ ,  $(P_1)$  and  $(P_2)$  in Municipality 1 is school 14. In the case of relaxed  $(P_2)$ , with  $\lambda = 0.5$ , the decision is to close school 12, but both schools belong to zone 7. This relaxed version of the model also proposes to satisfy a large part of the total demand, leaving 1.4% unsatisfied.

The total monthly hours, 81,496 for Municipality 1 and 107,308 for Municipality 2 in the base case, can be decomposed into two components. The first is related to the “classroom hours”, that is, the number of hours dedicated to teaching, and corresponds to 75,336 and 96,972 monthly hours, respectively. The second part represents the hours dedicated to supporting the educational labor (for example, the hours of the principal are included in this figure), and corresponds to 6,160 and 10,336 monthly hours, respectively.

After the optimization, the monthly “classroom hours” decrease by up to 43–47% for Municipality 1 and by up to 28–34% for Municipality 2, which can be interpreted as the current percentage of overstaffing levels. This decrease results in a total annual savings of up to around \$5,000,000–\$5,500,000 for both cases and four versions of the model. In fact, if we calculate the average of current daily teaching hours for both cases, the result is between 9 and 12, which does not reflect reality while, after optimization, this value decreases to 7 in all cases; this is consistent with an ordinary academic curriculum in any school. Regarding the number of classes in the base case, 308 are needed in Municipality 1 and 548 in Municipality 2 (maybe not really needed, because

of eventual overestimation). After the optimization, this number decreases by between 8% and 13% for Municipality 1 and by between 10% and 14% for Municipality 2. Note that, although the optimization efficiency varies from 32 to 47% in terms of teaching hours, the percentage of improvement in terms of number of classes is much smaller. This suggests that for these cases, the most significant inefficiency is found in the allocation of human resources.

The enormous difference between base case and optimized cases can be explained by the overestimation of the required teaching staff and the fact that the total cost in the base case includes the salaries of the director or head of the school and other non-teaching staff.

## 5 Conclusions and further remarks

The model presented here in three versions permits the planner to estimate the infrastructure and human resources needed in order to satisfy demand for the municipal school system and the corresponding costs of education at each school and for each grade. At the same time, the model gives the configuration of each school in terms of the size of the student population and the number of classes in each grade. The most significant extra costs of base case from optimal alternative come from the allocation of human resources, but it is sensitive to definition of maximum classmates size, which could also be simulated with the general model introduced here. Some opened questions for further research arise from this work.

- i) The model provide a general framework for handling real instances of economic planning of the education system and they do not include the decision of how many teachers are necessary to fulfill the teaching demands of each school. The model only gives an indication of the total number of hours that are necessary to satisfy student demand.
- ii) We did not consider the quality outcomes as a constraint in our simulations, but the choice of variables to represent this factor in the mathematical models is an open question.
- iii) A dynamic version of the model should be developed to take into account the future evolution of demand over several years and to propose more realistic planning. This is a crucial point, because the mathematical models presented here might suggest closing/keeping a given school in a particular case, but according to demographic evolution, this could be a bad decision (closing/opening schools is a major decision, having a very high economic and emotional impacts).
- iv) In this paper, we have not considered the random behavior of demand, which is inherent to future demographic and economic evolution, which is specially true of Chile. This needs a new model together with algorithms for model resolution using stochastic or robust optimization techniques.



- v) Concerning the situation in which some schools have to be closed, the model should include the benefits or costs of that decision. Closing a school therefore means that significant public infrastructure is disaffected and the use of this resource is then left as an open question. This aspect should be explicitly included in the model.
- vi) The current model only deals with the public sector, without considering the other two sectors of the Chilean education system, which are very dynamic and competitive. The challenge is to study how these sectors influence the demand for public education, in order to propose more general education models.
- vii) An open question is to study the structure of the optimization models presented here (in particular, problem  $(P_2)$  seems to have a staircase structure). In fact, we have not discuss the algorithmic aspects of solving these models, since the instances we solved were only of moderate size, but in real instances having a large number of schools and variables, it may be necessary to propose adapted algorithmic strategies, for example, decomposition or heuristic methodologies. Larger instances could appear if, for example, these models were extended to the dynamic case, in which the evolution of demand over several years could generate a much larger number of variables and constraints.

**Acknowledgements.** The authors would like to acknowledge the data provided by two Chilean municipalities.

## References

- [1] Antunes, A. and D. Peeters. A dynamic optimization model for school network planning. *Socio-Economic Planning Sciences*, 34 (2000) 101-120.
- [2] Antunes, A. and D. Peeters. On solving complex multi-period location models using simulated annealing. *European Journal of Operational Research* , 130 (2001) 190-201.
- [3] Bellei, C., P. Gonzalez and J.P. Valenzuela. Fortalecer la educación pública: un desafío de interés nacional. In: *Ecos de la Revolución Pingüina* (Bellei, C., D. Contreras and J.P. Valenzuela, eds). UNICEF and CIAE-Universidad de Chile. Pehuén Editores (2010) 225-256.
- [4] Church, R. L. and A. T. Murray. Modeling school utilization and consolidation. *Journal of Urban Planning and Development (ASCE)*, 119 (1993) 23-38.
- [5] Ferland, J. and G. Gunette. Decision support system for the school districting problem. *Operations Research*, 38 (1990) 15-21.
- [6] Caro, F., T. Shirabe, M. Guignard and A. Weintraub. School redistricting: embedding GIS tools with integer programming. *Journal of the Operational Research Society*, 55 (2004) 836-849.

- [7] Geiger, M.J. and W. Wenger. On the assignment students to topics: a Variable Neighborhood Search approach. *Socio-Economic Planning Sciences*, 44 (2010) 25-34.
- [8] Malczewski, J. and M. Jackson. Multicriteria spatial allocation of educational resources: an overview. *Socio-Economic Planning Sciences*, 34 (2000) 219-235.
- [9] Pizzolato, N. D. and H. B. Fraga da Silva. The location of public schools: evaluation of practical experiences. *International Transactions in Operational Research*, 4 (1997) 13-22.
- [10] Schoepfle, O.B. and R.L. Church. A fast, network-based, hybrid heuristic for the assignment of students to schools. *Journal of the Operational Research Society*, 40 (1989) 1029-1040.
- [11] Teixeira, J., A. Antunes and D. Peeters. An optimization-based study on the redeployment of a secondary school network. *Environment and Planning B: Planning and Design*, 34 (2007) 296-315.
- [12] Valenzuela, J. P., C. Bellei and D. De los Ríos. Segregación escolar en Chile. In: Martinic, S. and Elacqua, G. (Eds). *Fin de Ciclo? Cambios en la gobernanza del sistema educativo*. OREALC-UNESCO. Pontificia Universidad Católica de Chile, (2010) 209-229.

## APPENDIX: Tables

<b>Municipality 1</b>	Model $P_0$	Model $P_1$	Model $P_2$	Relaxed Model $P_2$
# integer variables	13,628	14,896	14,896	15,028
# binary variables	258	258	258	357
# constraints	1,145	1,324	1,357	2,629
<b>Municipality 2</b>				
# integer variables	14,597	16,472	16,472	16,574
# binary variables	521	521	521	687
# constraints	2,689	2,940	2,974	4,472

Table 1: Optimization instances.

<b>School Id Municipality 1</b>	Available classrooms $T_k$	Zone $Z_h$	Annual fixed cost $w_k$	Annual total cost	Effective annual demand
1	20	8	\$ 340,158	\$ 948,428	723
2	18	8	\$ 159,858	\$ 717,439	439
3	38	6	\$ 391,602	\$ 1,534,137	992
4	32	2	\$ 365,143	\$ 1,371,324	929
5	6	3	\$ 48,444	\$ 247,652	70
6	29	1	\$ 384,239	\$ 1,276,876	834
7	18	4	\$ 315,956	\$ 897,361	545
8	38	7	\$ 593,062	\$ 1,734,583	1,192
9	55	6	\$ 834,686	\$ 2,421,765	1,885
10	20	8	\$ 195,865	\$ 628,751	429
11	16	7	\$ 201,030	\$ 437,242	331
12	23	7	\$ 319,585	\$ 925,828	641
13	20	5	\$ 243,611	\$ 809,302	507
14	14	7	\$ 71,888	\$ 222,942	89
15	22	7	\$ 323,172	\$ 942,087	714
Total	369		\$ 4,788,296	\$ 15,115,718	10,320

Table 2: Main data for Municipality 1

School Id Municipality 2	Available classrooms $T_k$	Zone $Z_h$	Annual fixed cost $w_k$	Annual total cost	Effective annual demand
1	32	6	\$ 236,366	\$ 938,919	1,054
2	27	6	\$ 218,565	\$ 1,080,281	843
3	33	6	\$ 307,573	\$ 1,176,893	1,300
4	23	2	\$ 182,962	\$ 650,316	793
5	23	2	\$ 200,763	\$ 751,755	860
6	17	2	\$ 147,358	\$ 597,985	608
7	19	2	\$ 154,281	\$ 625,691	603
8	20	2	\$ 163,182	\$ 649,291	614
9	18	2	\$ 154,281	\$ 600,346	584
10	14	6	\$ 120,656	\$ 502,346	467
11	20	4	\$ 165,160	\$ 580,811	622
12	16	2	\$ 170,105	\$ 608,060	707
13	22	2	\$ 194,829	\$ 729,094	905
14	18	3	\$ 180,984	\$ 680,272	777
15	24	1	\$ 205,708	\$ 763,290	929
16	14	6	\$ 124,612	\$ 620,352	492
17	37	3	\$ 308,562	\$ 1,109,958	1,361
18	25	2	\$ 225,488	\$ 854,034	833
19	39	2	\$ 330,320	\$ 1,148,444	1,563
20	28	2	\$ 248,234	\$ 922,908	1,092
21	27	3	\$ 234,388	\$ 820,356	1,004
22	40	5	\$ 356,033	\$ 1,441,796	1,667
23	9	7	\$ 89,008	\$ 297,848	184
24	6	8	\$ 53,405	\$ 220,173	84
Total	551		\$ 4,772,822	\$ 18,371,218	19,946

Table 3: Main data for Municipality 2

<b>Municipality 1</b>	Base case	$P_0$	$P_1$	$P_2$	Relax. $P_2$
Demand	10,320	10,320	10,320	10,320	10,320
Satisfied demand	10,320	10,320	10,320	10,320	10,175
Numb. of schools	15	14	14	14	14
Total cost	\$ 15,115,718	\$ 10,114,303	\$ 10,114,303	\$ 10,114,303	\$ 9,542,702
Annual saving		\$ 5,001,415	\$ 5,001,415	\$ 5,001,415	\$ 5,573,016
% reduction/base case		33%	33%	33%	37%
Tot. teach. hours/month	75,336	42,596	42,596	42,596	40,040
% reduction/base case		43%	43%	43%	47%
Total classes	308	284	284	284	268
% reduction/base case		8%	8%	8%	13%
<b>Municipality 2</b>	Base case	$P_0$	$P_1$	$P_2$	Relax. $P_2$
Demand	19,946	19,946	19,946	19,946	19,946
Satisfied demand	19,946	19,946	19,946	19,946	19,849
Numb. of schools	24	24	24	23	24
Total cost	\$ 18,371,218	\$ 13,148,708	\$ 13,148,708	\$ 12,991,598	\$ 12,835,449
Annual saving		\$ 5,222,510	\$ 5,222,510	\$ 5,379,620	\$ 5,535,769
% reduction/base case		28%	28%	29%	30%
Tot. teach. hours/month	96,972	66,096	66,096	66,156	63,624
% reduction/base case		32%	32%	32%	34%
Total classes	548	490	490	490	474
% reduction/base case		11%	11%	11%	14%

Table 4: Results for Municipality 1 and Municipality 2.