Studies in Educational Evaluation 44 (2015) 50–59

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests

Monica Bravo Sanzana [a], Sonia Salvo Garrido [b,*], Carlos Muñoz Poblete [c]

[a] Estudiante de Doctorado en Ciencias de la Educación, Universidad de La Frontera, Chile
[b] Departamento de Matemática y Estadística, Universidad de La Frontera, Temuco, Chile
[c] Departamento de Ingeniería Eléctrica, Universidad de La Frontera, Temuco, Chile

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form 9 December 2014
Accepted 14 January 2015
Available online 14 February 2015

Keywords:
Classification and regression tree (CART)
Random forest (RF)
Performance in mathematics
Student profiles
Educational expectations

ABSTRACT

Attempting to deepen the understanding of factors that explain student performance, this study seeks to identify and characterize profiles of Chilean students based on academic performance in mathematics. As analytical method, statistical techniques known as random forest (RF) and classification and regression tree (CART) were used to identify groups of eighth-grade elementary students according to their performance in 2011 test, using features related to individual and family behavior. The analysis was performed with a database provided by the Education Quality Measurement System of Chile. Results show that "parents' educational expectations" (42.7%) is key factor to obtain the best children's performances. Additionally, the analysis showed that the "type of school" (26.6%) and the "index of mathematical abilities" (26.1%) influence good performance.

Introduction

Improving the quality of education is an issue that various governments are seriously addressing, Chile is among these governments. The governments' expectations are that the economic resources invested in education will have a direct influence on student performance (Gershberg, González, & Meade, 2012; Gleesona & Donnabháina, 2009; Mizala & Torche, 2012); however, in recent years, it has been observed that despite economic investments, academic results have not steadily improved over time (Carrasco & San Martin, 2012; Elacqua, Santos, Urbina, & Martínez, 2011; Mizala & Romaguera, 2000; Ostoic, Mizala, & Romaguera, 2004; Vera, 2013).

The current Chilean school system owes its origin to educational reform in the eighties, reform that was radical and pioneering in the world with theoretical foundations in the social economy market (Friedman, 1955). Its target was to decentralize education, carried out as a first action that all the educational units administered by the State will stay under the administration of the municipalities (Cox & María José, 1999). Secondly, it created a system of school subsidy per student to finance expenditure on education in public and private establishments (Aedo & Sapelli, 2001).

Thus, three types of institutions were established in the Chilean educational system:

(a) Municipal, with funding from the State (via subsidy per student) and municipal administration which currently represents 37.5%.
(b) Privately funded, with financing from the State (via subsidy per student) and private management, with sustained increase in time, representing a 53.9%.
(c) Private paid, with funding and private management, representing 7.1%.

The remaining 1.5% is associated with nonprofit corporations associated to education (Paredes & Pinto, 2009; Torche, 2005).

This way, the Chilean educational reform decentralized the administration of educational units of the public sector, thus promoting competition among educational units and remaining with an explicit participation in the financing and provision of educational services. However, the Ministry of Education held a

* Corresponding author at: Departamento de Matemática y Estadística, Universidad de La Frontera, Av. Francisco Salazar 01145, Casilla 54-D, Temuco, Chile. Tel.: +56 045 2592864; fax: +56 045 2592 822.
E-mail addresses: m.bravo.sanzana@gmail.com (M. Bravo Sanzana), sonia.salvo@ufrontera.cl (S. Salvo Garrido), carlos.munoz@ufrontera.cl (C. Muñoz Poblete).

http://dx.doi.org/10.1016/j.stueduc.2015.01.002
0191-491X/

centralizing role in the design of curriculum, school schedule, evaluation and promotion system, etc. (Carmichael & Taylor, 2005; Cheng & Lam, 2013; Ostoic et al., 2004).

In this context, the measurement of the quality of the education system (SIMCE) was created in 1988. SIMCE tests measures the performance of Chilean students at various levels. These tests are applied every year, at the same time, in the schools of the whole country, and since 1995 the results are published as public domain. Thus, according to the logic of the reform, parents can evaluate the results of the educational process and choose the educational institution for their children, promoting competition among educational institutions (Lynch, 2010; Yazici, Seyis, & Altun, 2011). From this perspective, the underlying idea is to consider parents as critical consumers of education, however, there is evidence that there are significant differences in the understanding and use of the results of SIMCE tests (Elacqua & Martínez, 2011; Taut, Cortes, Sebastian, & Preiss, 2009). In this focus, this research can be considered as a contribution aimed to identify the key factors that impact the student performance in SIMCE tests.

Regarding the performance of Chilean students, the results obtained from international evaluations indicate that after a decade, this country is still falling below the mean of the Organisation for Economic Cooperation and Development (OCDE) countries (Cifuentes & Oliva, 2009; MINEDUC, 2010; Schwab, 2012). However, when performing a comparative analysis with the countries of South America, the results show that Chile is one of the countries with the best results in the region (Cariola, Cares, & Lagos, 2009). Moreover, Chilean students' performance in national evaluations through the Education Quality Measurement System (SIMCE) indicate that the results in the Language and Communication, Mathematics and Natural Sciences tests do not exhibit any significant upward trend (Valenzuela, Bellei, & De los Ríos, 2008). In the case of mathematics, 65% of the students are at the initial level, which means they are unable to solve simple mathematical problems (MINEDUC, 2011).

So, as the focus of the educational policy in Chile lies in the relationship of the quality education and academic performance obtained by educational establishments, the foregoing represents a theme of equity and of providing opportunities for quality (Elacqua & Martínez, 2011; Volante, Cumsille, Denardin, & Muller, 2008), which demand an informed decision-making and with empirical support on the side of the government and the own educational establishments.

In this context, the question that arises is the following: what measures could be taken to improve the school performance? The evidence has shown that to improve student performance, it is necessary to identify the intervening variables (Bellei, Muñoz, Pérez, & Raczynski, 2004; García & Paredes, 2010; Murillo, 2003, 2007; Scheerens & Bosker, 1997; Scheerens, 2000; Slavin, 1996; Volante et al., 2008) and implement measures that include those that can reasonably be taken through education policy in the medium term (Salvo, Moraga, Miranda, Ramírez, & Vera, 2012).

The international literature has highlighted the most relevant variables at school level to explain the performance of the students are the infrastructure (Murillo & Román, 2011), the quality of education (measured on cognitive skills) has powerful economic effects (Hanushek & Woessmann, 2010), the effect of education policies (Benavota, 2012; Gershberg, Meade, & Andersson, 2009; Paletta, 2012), professional teacher training (Oancea & Orchard, 2012), teachers' perceptions (Georgiou, 2008), family and community relations with the school (Sheldon & Epstein, 2005), preschool education(Bulotsky-Shearera, Bella, & Domínguez, 2012) the type of institution of the educational establishment(Coertjens, Boeve-de Pauw, De Maeyer, & Van Petegem, 2010; Coll, Dahsah, & Faikhamta, 2010; Cordero, Crespo, & Santín, 2010; De Jorge & Santín, 2010; Demir & Kiliç, 2009; Gilleece, Cosgrove, & Sofroniou, 2010; Mancebón-Torrubia, Calero, Choi, & Ximénez-de-Embún, 2010; Sousa & Armor, 2010; Willms, 2010), the selection of the students that make some schools prior to admission (Bellei, 2009; Salvo et al., 2012), the significance of the peer effect, i.e. when an individual student performance is affected by his classmates (Carmichael & Taylor, 2005) and multigrade teaching (Kucita, Kivunja, Maxwell, & Kuyini, 2013).

There is a study that summarizes 15 years of research and covers over 800 meta-analysis. This study discusses the influence factors of the school that contributes with the learning process and show the size of this effect. Among their findings, the "teacher" factor showed the strongest effect. The three strongest single factors were self-reported grades, Piagetian programs, and providing formative evaluation (Hattie, 2008). The work of Hattie (visible learning) has been criticized in several respects. For example, as established by Snook, O'neill, Clark, O'neill, and Openshaw (2009), Hattie does not provide the reader with accurate information on the quality standards used when deciding if a research study is integrated into their meta-analysis; in addition. Most successful outcomes obtained by Hattie are in situations that teachers come from innovations and those may not be the same that the outcomes with teachers in regular classrooms.

Others attempts research the interaction between the performance and variables associated to the student considering the family background, as the age of the student (Ma, 2005), the gender of the student (Carr, Steiner, Kyser, & Biddlecomb, 2008; Chen, Chen, Lee, Chen, & Keith, 2013; You & Sharkey, 2012), socio-economic status of the families (Calero, Choi, & Waisgrais, 2010), income, profession and education of the parents (Anderson, Chui, & Yore, 2010; Kalender & Berberoglu, 2009; McConney & Perry, 2010; Milford, Shelley, & Anderson, 2010; MINEDUC, 2009; Mizala, Romaguera, & Urquiola, 2007; OECD, 2009; Tucker-Drob, 2013), perception of self-achievement and motivation toward objectives (Carroll et al., 2012; Cheng & Lam, 2013; Vera, 2013), motivation and task-oriented (Hirvonen, Tolvanen, Aunola, & Nurmi, 2012),educational perspectives of parents and teachers (Cornejo & Redondo, 2007; World Economic Forum, 2008; Johnson, McGue, & Iacono, 2007; Townsend, 2007).

A study was recently conducted in primary education in Latin America. Using a multilevel methodology, the authors analyzed the performance reached by students and schools in 9 countries. Performance was measure using four variables: self-concept, academic behavior, social coexistence and school satisfaction. The results coincided with research focused on cognitive products, but in the results also appear school and classroom factors associated with the socio-affective development of students. The finding of more integral factors is relevant, since it relates aspects of well-being (comfort) of the educational environment of the student, instead of aspects purely cognitive (Murillo & Hernández-Castilla, 2011).

The classification of these factors allows defining profiles that characterize the existing relationship between the students and its school context. This makes possible to observe how variables interact among them and thus facilitate the design of contextualized actions aimed to improve the educational system.

The use of the techniques of Multinomial multilevel modeling (Lam, Chow-Yeung, Wong, Lau, & Tse, 2013) and Classification and Regression Tree, CART (Breiman, Friedman, Olshen, & Stone, 1984) has been used to create profiles that characterize behaviors of groups of people. CART has been used with much success in medical research (Schiattino & Silva, 2008) since it delivers relationships between factors associated with individuals. In this regard, data analysis of the Longitudinal Study of American Youth (LSAY, in the United States), Grades 7–12, classified students in groups with differential rates of growth in achievement in mathematics. Through individual and family variables, the study