Strategies for Data and Learning Analytics Informed National Education Policies: the Case of Uruguay

Cecilia Aguerrebere
Plan Ceibal
Av. Italia 6201, CP 11500
Montevideo, Uruguay
caguerrebere@ceibal.edu.uy

Cristóbal Cobo
Plan Ceibal
Av. Italia 6201, CP 11500
Montevideo, Uruguay
ccobo@ceibal.edu.uy

Marcela Gomez
Plan Ceibal
Av. Italia 6201, CP 11500
Montevideo, Uruguay
migomez@ceibal.edu.uy

Matías Mateu
Plan Ceibal
Av. Italia 6201, CP 11500
Montevideo, Uruguay
mmateu@ceibal.edu.uy

ABSTRACT

This work provides an overview of an education and technology monitoring system developed at Plan Ceibal, a nationwide initiative created to enable technology enhanced learning in Uruguay. Plan Ceibal currently offers one-to-one access to technology and connectivity to every student and teacher (from primary and secondary education) as well as a comprehensive set of educational software platforms. All these resources generate massive amounts of data about the progress and style of students learning. This work introduces the conceptual framework, design and preliminary results of the Big Data Center for learning analytics currently being developed at Plan Ceibal. This initiative is focused on exploiting these datasets and conducting advanced analytics to support the educational system. To this aim, a 360 degrees profile will be built including information characterizing the user's online behavior as well as a set of technology enhanced learning factors. These profiles will be studied both at user (e.g., student or teacher) and larger scale levels (e.g., per school or school system), addressing both the need of understanding how technology is being used for learning as well as to provide accurate feedback to support evidence based educational policies.

CCS Concepts

•**Social and professional topics** → *Computing education; Student assessment;*

Keywords

big data; Plan Ceibal; education policies; technology enhanced learning

LAK '17, March 13 - 17, 2017, Vancouver, BC, Canada

ISBN 978-1-4503-4870-6/17/03…$15.00

DOI: http://dx.doi.org/10.1145/3027385.3027444

1. INTRODUCTION

Plan Ceibal[1] is a national policy programme that has been implemented over the last nine years in Uruguay. At its very beginning, the government agency created for such purposes, promoted equal opportunities by providing a laptop and Internet access to every child and teacher. After distributing over 700.000 laptops and deploying Internet connections for schools and communities, reaching 85% of the students in the country, Plan Ceibal has made considerable progress towards enabling higher levels of social inclusion and equity by reducing the digital divide in all socio-economic contexts.

With a reduced digital divide, Plan Ceibal is now focused on providing the public educational system with an integrated set of tools, technologies, methodologies and digital strategies to enhance the learning and teaching process. Plan Ceibal currently offers a set of educational software platforms for teaching, learning, training, hosting, exchanging and creating information. Virtual learning environments (VLE) at Plan Ceibal allow real-time interaction between students and their teachers and peers through a variety of resources and exercises, discussions or instant messaging. These VLEs generate massive amounts of data on the progress and style of students learning. In addition to this, after nine years of providing connectivity, Plan Ceibal has a wide range of data coming from the use of devices (tablets, laptops) and schools WiFi networks. Strategic use of data in the education sector is crucial to inform the teaching and learning process, to reduce achievement gaps and to increase the quality of public education [3].

Today Plan Ceibal is building a business intelligence system, developing preliminary experiences in educational data mining. However, until now, these databases have been collected and processed just for management and operational needs, without analysing multi-platform online user's behavior. In order to expand these experiences and conduct advanced analytics of the educational system, a Big Data Centre for Learning Analytics is being planned [2]. This will build the foundation for future research which can further analyze how technology is being used and how it can be used for teaching and learning practices. This involves new challenges such as the development of new tools, techniques, and

[1] www.ceibal.edu.uy, http://www.fundacionceibal.edu.uy/en

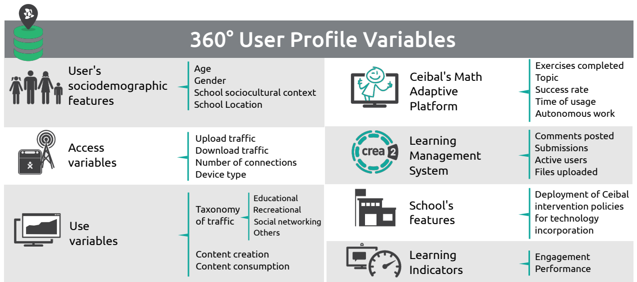

Figure 1: Main variables to be included in the 360° profile.

people's skills; resolving data concerns such as how the data is captured, analyzed and disseminated protecting ethical, legal and societal concerns. In this regard, several efforts are being made by the organization in order to guarantee the quality and integrity of the data management, ensuring anonymity and following ethical and legal guidelines not only in accordance to the Uruguayan legislation but also in agreement with international recommendations [9].

One of the goals of this work is to build a comprehensive user online profile, hereafter referred to as *360° user profile*. This profile, combined with statistical modelling techniques, can help: (i) identify online patterns on aspects such as learning styles, content creation pathways, adaptive contents services; (ii) predict learning behaviors; (iii) measure student's engagement or retention. Also within the scope of the 360° profile, but at a larger scale, this meta-index can provide useful information of the school system as a whole. This higher level of data is expected to provide accurate information for decision-makers regarding how tools (e.g., devices, infrastructure) and services (e.g., connectivity, educational software) are used at a national scale. It is expected that the integration of variables, aggregated or disaggregated, can contribute to understand either general trends or individual based behavior, helping Plan Ceibal make better informed decisions on future educational strategies.

This article is organized as follows. Section 2 introduces the 360° user profile and proposes some of the fundamental questions that motivate its creation. Sections 3 and 4 provide examples of the analysis based on the this profile. Finally, a summary of this work is presented in Section 5.

2. TOWARDS A 360° USER PROFILE

The 360° user profile is the result of integrating key variables that describe a wide variety of aspects related to the online behavior of educators and learners. This profile includes aspects such as: network usage (e.g., what kind of application does the learner prefer, is the learner active in social networks, is the learner active in the educational platforms), educational platforms usage (e.g., if the learner uses the educational platforms, what kind of usage does she/he do? number of exercises completed, number of correct exercises, number of required hints, books consultations, frequency of usage, usage outside the school, proactive or teacher suggested usage, participation in discussion forums, individual or collaborative work), sociocultural background, school context, among others. Figure 1 presents a summary of the main variables included in the 360° user profile.

This profile will be the main input underpinning the analysis in future studies, both at user (e.g., student or teacher) and larger scale levels (e.g., per school or school system), addressing both the need to understand the influence of technology on the learning experience as well as to provide accurate feedback to support evidence based national policies.

When the 360° profile integrates data at a general level (school system) the information analyzed aims to provide large-scale insights to monitor and understand how and when the system is being used. What are the demands and main uses of the educational community at a national level. On the other hand, the individual-based level aims to offer a comprehensive (multi-dimensional) perspective of the user and his educational practices. Although groups of individuals with similar behaviour can suggest patterns, the aim is to provide information that could be useful for learners and educators when monitoring the benefits and challenges of the learning experience, targeting questions such as: What student actions are associated with better performance, satisfaction, engagement, learning progress? What features of an online learning environment lead to better performance learning? When are students falling behind in the course? When a student should be referred to a counselor for help?

In order to illustrate how the 360° profile information is analyzed at these two levels, we provide in the following sections examples of these complementary approaches: 1) A global monitoring system which explores key trends that describe how the network is used by the school system at an aggregated level, longitudinal evolution of the traffic, among others. 2) Measuring a set of variables to understand individual's performance when completing exercises with an intelligent tutoring system for mathematics (e.g., frequency, type and number of exercises completed, type of devices used, use of help, etc).

3. A GLOBAL MONITORING OF THE SCHOOLS SYSTEM

Network usage data forms a central part of the Big Data project at Plan Ceibal. To this aim, a Global Monitoring System (GMS), inspired in [6], is being developed. This GMS will enable the visualization of the network usage, which is an essential input to: (i) assess for what purposes Plan Ceibal's network resources are being used, (ii) capture the dynamics of network usage to understand the existing trends and help adjust Plan Ceibal's supply based on the user's needs, (iii) study how the different family of devices are being used, (iv) provide network usage data at user level to build part of the key components of the 360° profile (e.g., most used resources, most visited sites and platforms, etc). The development of the GMS consists in four main phases, detailed in the Gantt diagram shown in Figure 2.

3.1 Data Sources

The GMS combines access and use variables. Access variables include upload and download traffic (per site, user, or application), number of connections (per site, user, or service), and multiple types of devices (laptops, tablets, smartphones, others). Use variables cover the taxonomy of traffic (pedagogical or educational, recreational or entertainment, among others), origin and destination of traffic (school, home, national, or international), and content production versus consumption. The access variables show which online appli-

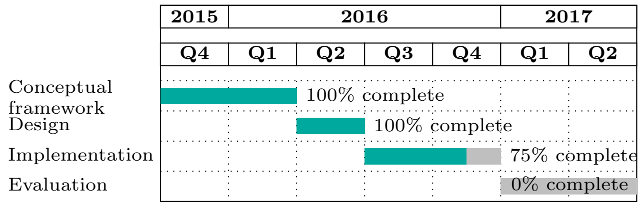

Figure 2: Planning for the Global Monitoring System project.

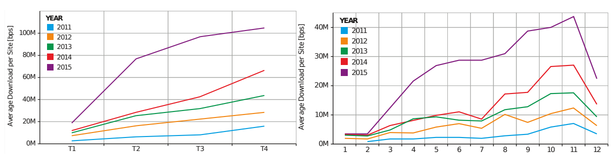

Figure 3: Traffic demand evolution of Plan Ceibal network from February 2011 to December 2015, accumulated by quarters (left) and by month (right).

cations and devices are consuming more network resources (i.e., bandwidth, wifi connections), while the use variables illustrate the types of use (e.g., educational, recreational, type of devices, etc) therefore suggesting the means and motivations for getting online.

3.2 Preliminary results

3.2.1 *Traffic and connections*

The conceptual framework for the definition of global access variables was presented in [7] and an initial analysis of results is introduced in [12]. The latter focuses on the question: what was the evolution of the aggregate demand of Internet access in the Plan Ceibal network between 2011 and 2015 and what is the projection for 2016 to 2019?

The considered variables are download and upload internet traffic and simultaneous connections. These variables were collected hourly everyday, from February 2011 until December 2015. The analysis was conducted with the value of these variables at busy hour[2]. The number of potential users in the network was 625.000, of whom about 120.000 were simultaneously connected during busy hour [4].

Figure 3 presents two charts: the first one, aggregated in quarters, shows a consistent increase in traffic demand from quarter to quarter. In the second chart, the aggregation is done by month, showing a dramatic decrease during the summer break (December to February), and a moderate decrease during the winter break (July). Figure 4 shows the evolution of traffic per site and simultaneous connections during the period 2011 to 2015, considering primary and secondary schools. It is interesting to note that, starting in 2014, the number of connections in secondary schools has grown dramatically, surpassing those in primary schools. This phenomenon correlates in time with the introduction of the Bring Your Own Device Policy in secondary schools (users are now allowed to connect to Plan Ceibal's network their personal devices), which may explain the effect.

[2]60 minute period of the day during which the maximum total traffic load of the network occurs.

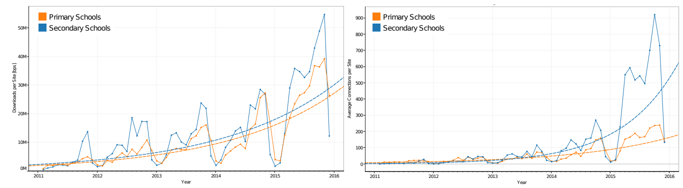

Figure 4: Traffic demand (left) and number of conexions (right) evolution of Plan Ceibal network from February 2011 to December 2015 for primary and secondary schools.

From these preliminary results we observe that Plan Ceibal's internet download traffic has grown 13 times between 2011 and 2015. CISCO VNI reported that global Internet download traffic has grown five times in the same period [6]. In other words, Plan Ceibal's internet consumption has grown about 2.5 times faster than the global Internet. At the same time, download traffic has doubled every 18 months during the period between 2011 and 2015, while upload traffic also doubled every 16 months in the same period. Simultaneous connections per site have doubled every 12 months between 2011 and 2015. Based on these observations, and considering that aspects such as the bandwith capacity, the number of devices and the number of active users are expected to increase, the hypothesis is that the network will continue expanding.

4. TECHNOLOGY ENHANCED LEARNING ASSESSMENT

The 360° user profile will include variables describing socio-cultural aspects (e.g., school socio-cultural level), school characteristics (e.g., full-time, part-time, special education, geographical area, internet connectivity), information concerning the use of educational platforms (e.g., number of users, activities on the platform, type and intensity of use), among others. Furthermore, meta-indexes will be built, from the analysis and combination of these variables, to study more general concepts such as student's school engagement and performance.

Student performance is a major point of interest and has therefore been studied from several perspectives. One such approach is to model student learning curves using cognitive models such as Bayesian Knowledge Tracing [8], Performance Factor Analysis [15], Additive Factor Models [5], among others, studying individual or collaborative work [13]. Another perspective is to follow a more classical data mining approach using classical classification or regression techniques to predict grade scores from various features, such as previous scores, students demographics, extra-curricular activities, high school background, social interaction network, psychometric factors, among others [17]. Furthermore, there exists an extensive bibliography on relationship mining focused on finding relationships between performance and various features, such as the ones previously listed for performance prediction [10, 11, 18, 16, 1].

As an integral part of the creation of the 360° user profile, we will study performance from different perspectives, as well as the effects of technology enhanced learning on performance outcomes. In the following section we present a particular case study of performance which focuses on stu-

Table 1: **Adaptive platform for mathematics (PAM) users and number of performed activities from June 2013 to September 2016.**

	Total		Growth	
Year	Users	Activities	Users	Activities
2013[3]	51.025	4.341.127	-	-
2014	91.685	8.442.303	80%	94%
2015	113.617	31.818.345	24%	277%
2016[4]	120.349	27.346.793	15%[5]	19%

dent's performance on an intelligent tutoring system (ITS) for mathematics learning.

4.1 Performance in an intelligent tutoring for mathematics learning: a case study

We propose to study student's performance on an ITS that has been incorporated into the mathematics curricula through the Plan Ceibal network since 2013. This ITS implements an adaptive platform for mathematics (PAM, for its acronym in Spanish), which proposes series of exercises and suggests particular areas of improvement depending on the user's mathematical skills. Table 1 presents a summary of global usage indicators of this platform. So far this year, 27 million exercises have been made in the platform, 63% of which were performed by primary school students, 25% by secondary school students and the rest being completed by other beneficiaries of Plan Ceibal. It is important to remark that a very heterogeneous use of the platform is observed, as 20% of the total users completed 75% of the total exercises.

By combining data from different sources, including log-files of the user's activity in PAM, we aim at studying the effect of various factors on student's performance and answering fundamental questions such as: the more exercises the better? What role does frequency of exercises completed play here? Do proactive and autonomous students outperform the rest? Do schools receiving supportive Plan Ceibal policies have an increased average performance than the rest? To what extent the engagement of the teacher with the platform affects the performance of the student? and if so under what circumstances? Is higher performance in the platform related to particular patterns of network usage?

4.1.1 Data Sources

The considered dataset consists of 120.000 students in 1760 schools. The variables to be studied can be classified into five main categories: socio-cultural (e.g., school's socioeconomic context), geographical (e.g., urban or rural school), educational (e.g., the deployment of Ceibal intervention policies, such as the presence of support teachers to promote technology incorporation at school), technology (e.g., type of device used by the user) and user-specific (e.g., whether the user has a proactive attitude towards using PAM or not). Table 2 shows a detailed description of some of the considered variables. The performance indicator is here defined as the total number of correct exercises completed by the student in PAM in a given time period, for instance half or a complete school year.

[3]Data from June to December.

[4]Data from January to September.

[5]Growth corresponding to September 2015 to September 2016.

4.1.2 Methodology

In order to gain further insight into the effect of certain factors into students performance in PAM, we will evaluate possible correspondences between the variables listed in Table 2 and performance. First, the Pearson correlation between each variable and the performance outcomes will be computed to find possible linear dependences between these variables. Second, following a classical data mining approach, we will conduct various regression methods for performance prediction from the considered variables. More precisely, we will assess the prediction capacity of common classification methods such as decision trees, Naive Bayes, SVM and random forests, previously shown to be powerful for educational data mining applications [14, 19]. This evaluation will show the capacity of this set of variables to predict the performance outcomes.

5. SUMMARY AND FUTURE WORK

This work provides an overview of a unique national-scale education and technology monitoring system developed in Uruguay. Different efforts are currently conducted to adopt new metrics to better understand and support users needs in Uruguay's educational system.

The overall goal of the Big Data Centre for Learning Analytics at Plan Ceibal is to build a 360 degrees user profile at a school or at user's level (the later can include students or educators). Approaching these two levels is considered useful to fulfil the information needs of different communities (e.g., educators, learners, school leaders, policy makers, technology developers, researchers, etc.)

Today, the demand of Plan Ceibal's resources is facing a phase of expansion. This is considered a significant opportunity to improve the analysis process in order to better answer the needs of the different stockholders linked with this education and technology national initiative.

Acknowledging that building a multi-dimensional profile demands the selection and integration of key indicators, in this article the authors provide a framework of variables and information sources to consider when building these two stages of the 360° profile.

Noteworthy, the principal goal is to elaborate a meta-index, which provides relevant and accurate data of the teaching and/or learning experiences. In other words, if this analysis does not offer tools either to support the work of educators while they plan their teaching, or to provide relevant data for the learners regarding their performance (i.e., to support self-directed autonomous learning), then the proposed implementation will be suboptimal at best.

As follow, some of the forthcoming challenges that we anticipate during the creation of the meta-index (360 degrees profile) are highlighted:

1. Although Plan Ceibal manages national scale educational information, it is considered strategic to begin with the analysis at two levels (schools and users). The idea is to design mechanisms to compare large scale and individual-based analysis, whose outcomes will be relevant for different stakeholders.

2. The broader the scope of data available to include in the process the higher the responsibility of adopting proper standards to ensure the quality and integrity of the data management, protecting the identity and privacy of users (i.e., privacy-by-design) or following appropriate ethical and

Table 2: Variables used for performance prediction.

Description	Category
School's socioeconomic context	Socio-cultural
School's location (e.g., urban, suburban, rural)	Geographical
School type (e.g., regular, critical socioeconomic context, special care needed)	Educational
Deployment of Ceibal supportive policies	Educational
Teacher engagement with platform adoption in class (measured by the number of Ceibal courses completed by the teacher)	Educational
Device type (laptop, tablet, pc, smartphone)	Technology
Total number of exercises completed	User-specific
Autonomous work (total number of exercises started by user choice, as opposed to those indicated by the teacher)	User-specific
Completion of exercises suggested by the platform in the adaptive learning process	User-specific
Time or Location when the exercises were completed (e.g., school time, leisure time, school hours, off school hours)	User-specific
Frequency in which exercises were completed	User-specific
Network usage profile	User-specific

data privacy guidelines (such as DELICATE [9]).

3. To include educators and school leaders in an early stage of the learning analytic process. In order to ensure their participation but also to improve the design of the analytics with continuous iterations and inputs from the different communities involved.

6. ACKNOWLEDGMENTS

The authors would like to thank Plan Ceibal and Fundación Ceibal for supporting this study, as well as Germán Capdehourat and Enrique Lev for their valuable contributions.

7. REFERENCES

[1] L. K. Allen and D. S. McNamara. You are your words: Modeling students' vocabulary knowledge with natural language processing tools. *In 8th Int. Conf. on Educational Data Mining*, pages 258–265, 2015.

[2] M. Bailón, M. Carballo, C. Cobo, S. Magnone, C. Marconi, M. Mateu, and H. Susunday. How can Plan Ceibal Land into the Age of Big Data? *In 4th Int. Conf. on Data Analytics*, pages 126–129, 2015.

[3] P. Bergman. Technology adoption in education: Usage, spillovers and student achievement. In *2015 Fall Conference: The Golden Age of Evidence-Based Policy*. Appam, 2015.

[4] G. Capdehourat, G. Marín, and A. Rodríguez. Plan Ceibal's wireless network for the implementation of the 1:1 educational model. *Latin America Networking Conf.*, pages 9:1–9:4, 2014.

[5] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Int. Conf. on Intelligent Tutoring Systems*, pages 164–175, 2006.

[6] Cisco. Cisco visual networking index: Forecast and methodology. Technical report, Cisco, 2014.

[7] C. Cobo and M. Mateu. A conceptual framework for analysis and visualization of uruguayan internet for education. *To appear in Interactions, ACM*, 2016.

[8] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[9] H. Drachsler and W. Greller. Privacy and analytics: It's a delicate issue a checklist for trusted learning analytics. In *In 6th Int. Conf. on Learning Analytics & Knowledge*, pages 89–98, 2016.

[10] S. Fancsali. Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. In *In 7th Int. Conf. on Educational Data Mining*, pages 28–35, 2014.

[11] C. Forsyth, A. Graesser, P. Pavlik Jr, K. Millis, and B. Samei. Discovering theoretically grounded predictors of deep vs. shallow level learning. *In 7th Int. Conf. on Educational Data Mining*, pages 229–232, 2014.

[12] M. Mateu. Plan Ceibal 2016: Future scenarios of technology and education. Unpublished Master's Thesis, 2016.

[13] J. K. Olsen, V. Aleven, and N. Rummel. Predicting student performance in a collaborative learning environment. *In 8th Int. Conf. on Educational Data Mining*, pages 211–217, 2015.

[14] Z. A. Pardos, R. Baker, M. San Pedro, S. M. Gowda, and S. M. Gowda. Affective states and state tests: investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 1(1):107–128, 2014.

[15] P. I. Pavlik, H. Cen, and K. R. Koedinger. Performance factors analysis –a new alternative to knowledge tracing. In *Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*, pages 531–538, 2009.

[16] M. O. Z. San Pedro, E. L. Snow, R. S. Baker, D. S. McNamara, and N. T. Heffernan. Exploring dynamical assessments of affect, behavior, and cognition and math state test achievement. *In 8th Int. Conf. on Educational Data Mining*, pages 85–92, 2015.

[17] A. M. Shahiri, W. Husain, et al. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72:414–422, 2015.

[18] E. Snow, L. Varner, and D. McNamara. Tracking choices: Computational analysis of learning trajectories. In *In 7th Int. Conf. on Educational Data Mining*, pages 316–319, 2014.

[19] P. Strecht, L. Cruz, C. Soares, J. a. Mendes-Moreira, and R. Abreu. A comparative study of classification and regression algorithms for modelling students academic performance. *In 8th Int. Conf. on Educational Data Mining*, pages 392–395, 2015.