Evolutionary development: a model for the design, implementation and evaluation of ICT for education programmes

Patricio Rodríguez[1], Miguel Nussbaum[2] and Lioubov Dombrovskaia[3]

[1]Center for Advanced Research in Education (CIAE),
Universidad de Chile
Periodista José Carrasco Tapia Nº 75, Santiago, Chile
prodriguez@ciae.uchile.cl // Tel: +56- 2 -978 2762

[2]Department of Computer Science, College of Engineering,
Pontificia Universidad Católica de Chile,
Edificio San Agustín P4, Av. Vicuña Mackenna 4860, Santiago, Chile
mn@ing.puc.cl // Tel: +56-2-354 4440

[3]Departamento de Informática,
Universidad Técnica Federico Santa María,
Av. Vicuña Mackenna 3939, Santiago, Chile
liuba@inf.utfsm.cl // Tel: +56-2-4326600

This is the pre-peer reviewed version of the following article: Rodríguez, P., Nussbaum, M., & Dombrovskaia, L. (2011). Evolutionary development: a model for the design, implementation, and evaluation of ICT for education programmes which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2011.00419.x/full 2011 in Journal of Computer Assisted Learning

Abstract

The impact of info-communication technology (ICT) in primary and secondary education is still an open question. Following review of the available literature, we classify the causes of the lack of impact on students' attainment in four dimensions: 1) the design and implementation of ICT in educational settings; 2) the evaluation of its impact; 3) the scaling-up of these kinds of innovations, and 4) the cost-effectiveness of technology enhanced learning environments.

Based on this evidence, we proposed the Evolutionary Development Model (EDM), which aims to produce a cost-effective and sustainable ICT for education (ICT4E) programme in three steps: efficacy, effectiveness, and efficiency. In each step, one component of the programme is built and validated in real educational settings. Therefore, the resultant ICT4E programme is ready to be replicated across the school system.

We also show how the EDM guided the development of a programme based on Mobile Computer Supported Collaborative Learning (MCSCL), known as Eduinnova. Finally, we discuss how EDM can serve as an analysis tool for researchers and policy makers.

Keywords: ICT; education; schools, monitoring and evaluation; cost-effectiveness; assessment, scaling-up, design-research; primary and secondary education.

1 Introduction

The vision of using ICT to transform the teaching and learning process in primary and secondary education (Culp, Honey, & Mandinach, 2003; Kozma, 2003) is far from becoming a reality (Moonen, 2008; Reynolds, Treharne, & Tripp, 2003; Robertson, 2002; Tondeur, Braak, & Valcke, 2007; Trucano, 2005). Some studies show little or negative impact on learning in developed (Campuzano, Dynarski, Agodini, & Rall, 2009; Conlon & Simpson, 2003; Dynarski et al., 2007) and underdeveloped (Barrera-Osorio & Linden, 2009; Sánchez & Salinas, 2008; Wainer et al., 2008) countries, while others show a positive impact, mainly in more advanced countries (Balanskat, Blamire, & Kefala, 2006; Harrison et al., 2002; OECD, 2006, 2010). Meta-analysis studies also reflect this contradictory reality (Kulik, 2003; Liao & Hao, 2008). From this point onwards, we will refer to any improvement in students' attainment simply as an "impact" (Cox & Marshall, 2007; Trucano, 2005). We have made this decision for two reasons: 1) The difficulty of measuring any educational initiative in monetary terms (Harris, 2009: Laurillard, 2007; Margolis, Nussbaum, Rodríguez, & Rosas, 2006) and 2) because the question of short and long-term effects on students' attainment still doesn't have a conclusive answer (Burns & Ungerleider, 2003; Cox & Marshall, 2007).

1.1 Dimensions affecting the impact of ICT in education

Having revised the respective literature, we have identified the causes that we feel explain why ICT has not had the effect that might be expected, and have broken them down into 4 dimensions: 1) the design and implementation of ICT in an educational context, i.e. how ICT applications are created and used in the classroom; 2) the evaluation of impact, i.e. the value of ICT in the educational system; 3) the scaling-up of the pedagogic use of ICT, i.e., how programmes based on ICT become widely used within the school system and 4) the estimation of cost-effectiveness, i.e. how to determine success based on learning results while also considering the necessary investment to achieve these goals. What follows is a synthesis of each of these dimensions.

Regarding the **design** of educational programmes supported by ICT, although these programmes should be based on educational theories and pedagogic principles (Lai, 2008; Marshall & Cox, 2008; Reeves, 2008; Roblyer, 2005) to establish a connection between specific use and possible results (Earle, 2002; McFarlane, 2001; Roblyer, 2005), the technology is generally placed before pedagogy (Trucano, 2005; Watson, 2001), in spite of the fact that there are several frameworks to support learning using ICT (Bottino, Chiappini, Forcheri, Lemut, & Molfino, 1999; Kirkman, 2000; Lim, 2002; Luckin, 2008; Mioduser, Nachmias, Tubin, & Forkosh-Baruch, 2003; Mooij & Smeets, 2001; Plomp, Pelgrum, & Law, 2007; Tearle, 2003, 2004).

Regarding the **implementation** of ICT, the majority of countries that introduced ICT in education followed a logical sequence of public policies that began with an “emerging stage”, (Villanueva, 2003) during which schools were provided with the necessary hardware, software, technical support and connectivity (Earle, 2002; Hennessy, Ruthven, & Brindley, 2005; Moonen, 2008; Tearle, 2003; Villanueva, 2003). However, this infrastructure merely enables the incorporation of ICT into pedagogic practices (Earle, 2002; Hayes, 2007; Laurillard, 2007). In a recent review of the literature, Infante & Nussbaum (2010) identified three types of barriers affecting technology integration that needed to be properly managed. The first type of barriers is related to infrastructure and support, such as lack of equipment, technical staff and other resources. The second type of barriers has to do with institutional obstacles, such as lack of leadership supporting the integration of ICT, and lack of training in terms of how to incorporate technology into the classroom. The third type of barriers is related to the opposed needs of students and teachers, within the classroom.

Furthermore, there is an important tension between the educational approach which promotes the use of ICT (a more constructivist pedagogical style) and the transmission-type pedagogies which tend to produce better results on standardized tests (Trucano, 2005; Cox & Marshall, 2007; Law, 2009). Finally, the lack of adequate monitoring of ICT initiatives which offers feedback about lessons learned affects the design of new programmes and educational policies (Anderson & Plomp, 2009; Reeves, 2008; Trucano, 2005).

With regards to the **evaluation of ICT's impact on education,** although many studies have been carried out, they have not produced conclusive answers (Cox & Marshall, 2007; Reeves, 2008) because there are no standard methodologies for measuring their impact (Anderson & Plomp, 2009; Trucano, 2005). Issues related to evaluation can be classified into 3 categories: what to measure, what to measure with, and how to measure.

In terms of what to measure, and what to measure with, we can highlight the following issues:

- Sometimes it is difficult to identify the effects of ICT, because its impact greatly depends on the type of resources used, how they are specifically used and the subject they are used with (Cox & Marshall, 2007; McFarlane, 2001).

- It is necessary to identify how the design and curricular implementation of ICT affect students because this will affect the learning impact (Cox & Abbot, 2004; Cox & Marshall, 2007).
- The teachers' pedagogical approaches need to be studied further because they are what determine the use of ICT, thus affecting students' learning achievements (Cox & Webb, 2004).
- In some cases, assessment instruments don't match results**,** because those instruments don't measure the results expected by ICT (Cox & Marshall, 2007; Roblyer, 2005; Trucano, 2005). For instance, they look for improvement in traditional processes and knowledge instead of new forms of knowledge and reasoning (Cox & Marshall, 2007; Penuel, 2005).
- The instruments used to measure educational results have rarely been investigated sufficiently for reliability and validity (Marshall & Cox, 2008), which can invalidate findings (Chatterji, 2005).

In terms of how to measure we can highlight the following issues:

- It is difficult to isolate the role technology plays in experimental studies carried out in real educational settings (Marshall & Cox, 2008) and this can weaken the obtained conclusions (Pilkington, 2008).
- There are substantial differences between the design and the actual implementation of ICT in

education (Penuel, 2005; Reeves, 2008).

- Lack of explanation regarding results. Experimental methods mainly determine which method (traditional class vs. technological support) was more effective, but not why (Chatterji, 2005; Reeves, 2008).
- Relevance of findings: while there may be differences that are statistically significant, practical relevance is frequently weak (Reeves, 2008), which is expressed, for example, by a very low effect size (Fan, 2001).
- The assessment of the effect that ICT has on learning requires appropriate experimental design with sample sizes that adequately control statistical errors (Agodini, Dynarski, Honey, & Levin, 2003; McDonald, Keesler, Kauffman, & Schneider, 2006; Raudenbush, 2008). Additionally, as the implementation scale grows, the effect size decreases (Slavin & Smith, 2009), which means that finding evidence can be a complex and expensive process (Chatterji, 2005; Marshall & Cox, 2008; Pilkington, 2008; Raudenbush, 2008; Slavin, 2008).

The **scaling-up process** has not been studied in depth in educational programmes, whether they are based on ICT or not (Coburn, 2003; McDonald et al., 2006). There is a tendency to try and repeat what worked locally, everywhere (Dede, 2006). However, the teaching strategies that are successful for one teacher often cannot be generalized, even for faculty members at the same school (Dede, 2006).

The emerging scaling-up models (Coburn, 2003; Dede, 2006) identify four dimensions for this process: depth (changes in classroom practice), sustainability (maintaining these changes over time), spread (diffusion of the innovation to large numbers of classrooms and schools), and shift in reform ownership (schools' adoption of the programme).

Finally, very few rigorous, quantitative studies of the real cost of ICT in education have been conducted (Trucano, 2005). Therefore, the required investment in technology cannot be easily calculated or compared between different countries and schools (Moyle, 2008). Even less is known about the cost-efficiency of ICT, particularly in developing countries (Trucano, 2005), although there are methods to calculate it (Harris, 2009; Margolis et al., 2006; Laurillard, 2007).

1.2 Problem definition, purpose and structure of this article

As has already been mentioned, several frameworks can be used to design applications that support the teaching-learning process within specific contexts. However, three practical aspects remain unsolved:

1. How to determine if an educational technology is ready to face a rigorous evaluation (Chatterji, 2005). This means addressing the issues of what to measure and what to measure with (section 1.1) to avoid wasting valuable time and resources.

2. The cost-effectiveness of an educational technology, because it competes for resources with other alternatives (Harris, 2009; Margolis et al., 2006). This is particularly important for developing countries, because the technology can be a means of compensating for the limitations of conventional education, but its costs should be lower and its impact higher than those of traditional methods (Castro, 2004).

3. How to spread the innovation throughout the school system. The idea of scaling-up should be

incorporated from the very beginning in the design of educational technology.

These three aspects are to be addressed through technology-enhanced instructional design or technology-enhanced learning (TEL) (Dillenbourg, 2008; Kozma, 1991; Roblyer, 2005), i.e., an educational programme based on technology as opposed to a particular technology. Thus, the design, implementation and evaluation of this type of programme should be conceived as a process to generate innovations with a higher chance of success as well as being economically feasible. Economic feasibility means determining if the programme can be implemented given a certain financial reality, such as the restrictions on school and Ministry of Education spending (Margolis et al., 2006). This is because without a reasonable control of the costs, TEL expenditure will consume a disproportionate amount of the limited funding available to education without a commensurate value (Laurillard, 2007).

This article presents an incremental model for the development of educational programmes based on ICT, validating the design through rigorous evaluation of several scales of implementation in representative educational contexts.

The remainder of this article is structured as follows: in Section 2 we define the evolutionary development model that conceptually combines design, implementation and evaluation of educational programmes based on ICT, so as to manage the problems we previously identified. In Section 3, we describe the application of the model using an educational programme based on a specific ICT. The article ends with conclusions and future work.

2 The evolutionary development model (EDM)

We define an ICT for education (ICT4E) programme as composed of a pedagogic model, an intervention and transference. The **pedagogic model** defines tasks that teachers and students should perform within schools, supported by ICT (Bielaczyc, 2006; Coburn, 2003; Luckin, 2008; Penuel, 2005). The **intervention** is the way the pedagogic model is adopted by the school leading to the autonomous implementation of the programme. It is composed of planned activities, such as training sessions for teachers, practical experiences, and classroom observations (Rodríguez, Nussbaum, López, & Sepúlveda, 2010). In addition to this, during intervention it is necessary to perform a **monitoring and evaluation plan** which measures the fidelity of implementation and determines if the pedagogical model was adopted, using process and result indicators (Wagner et al., 2005). The final component of an ICT4E programme is **transference,** which is a course of study for professionals to faithfully carry out the intervention on a massive scale (Rodríguez, 2008).

This conceptualization is the core element of the EDM that breaks down the design, implementation and evaluation of an ICT4E programme into stages where each component is built incrementally. The EDM is based on educational design research methodology (Bielaczyc, 2006; Reeves 2008) that addresses complex problems in real contexts in collaboration with practitioners researchers, experts and other stakeholders (Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). This is an iterative process of design, use, and redesign (Cox, 2008; Reeves, 2008) to build an "ideal" intervention (Van den Akker et al., 2006) and understand how it works (Reeves, 2006).

Our EDM considers three stages:

- **Stage 1**: the pedagogic model is built studying the **efficacy**, which will be achieved if, in a controlled environment, a technology application *can* improve students' results (Agodini et al., 2003).

- **Stage 2**: the intervention is designed and tested for its effectiveness. We will understand **effectiveness** as a technology that *does* improve students' outcomes in the educational context for which it was designed (Agodini et al., 2003).

- **Stage 3***:* transference is carried out, studying its **efficiency**, i.e., the achievement of the desired effects or results with minimum waste of resources. As a measure of efficiency**,** we will look at cost-effectiveness, (Feinstein & Picciotto, 200; Levin, 1983) calculated as the relationship between the effect size and the total cost of the programme (Harris, 2009; Margolis et al., 2006). This will allow us to analyse the practical applicability, replicability and scaling-up of the programme.

Following the design research methodology, formative studies are performed at each stage to evaluate the design (Reeves, 2008), developing and testing the instruments for the summative evaluation (Chatterji, 2005; Marshall & Cox, 2008; Penuel, 2005). At least one formative study should be performed before summative evaluation, so as to determine if the innovation is ready for an assessment of its effects and the cause of these (Chatterji, 2005).

The summative evaluation measures the impact of the consolidated design of each component. If the expected results are obtained, we move to the next stage, where new processes are designed and therefore need to be tested. For example, at the end of the efficacy stage we will have a completely developed pedagogic model with a notion of its effect size. In the case of ICT, the expected effect size fluctuates between 0.25 and 0.35 on a large scale (Agodini et al., 2003; Dynarski et al., 2007). If the obtained results are below this range at this early stage, it is not worthwhile to continue developing the programme. However, if the results are higher, the intervention must be added to the validated pedagogical model so that it may be tested in the effectiveness stage. The complete process is illustrated in Figure 1.

As we move forward through the stages, studies require greater time and implementation scales, making each cycle last longer (depicted by the diameter of each circle in Figure 1). Processes are also documented and specified in detail as part of the necessary transference to scale-up the programme. A detailed description of each stage follows.

2.1 Efficacy

At this stage, the pedagogic model is developed and tested for impact on students' attainment. The pedagogic model's components are focused mainly on practices carried out with technological support in a school context, in frameworks such as curriculum components (Van den Akker, 2003), domains of educational innovation (Mioduser et al., 2003), learner-centric ecology of resources (Luckin, 2008), and social infrastructure frameworks (Bielaczyc, 2006).

These elements must be reviewed through the cycle of formative evaluation that aims to guide the decisions that improve educational software through trials with students in schools (Reeves, 2008). The central elements that need to be evaluated at this stage are:

- The pertinence, i.e. usefulness, as perceived by teachers (Penuel, 2005), alignment with curricular objectives (Bielaczyc, 2006; Coburn, 2003; Cox & Marshall, 2007) and its applicability in different curricular contexts (Coburn, 2003; Reeves, 2008).

- The assumptions on which it is based, according to the supporting learning theories (Roblyer, 2005), the teachers' and students' attitudes, and the students' level of conceptual understanding and cognitive development (Cox & Marshall, 2007; Marshall & Cox, 2008; Penuel, 2005).
- The usability of the learning environment (Cox, 2008; Penuel, 2005), how the material is represented and how the students interpret it (Cox, 2008), and how students interact with software and become involved in a deeper learning activity (Marshall & Cox, 2008). There are guides and methodologies to evaluate these aspects (Reeves, 1994, 2006; Squires & Preece, 1999).
- The literacy in the ICT environment that is necessary for its use, because the way students think about a problem is influenced by how familiar they are with the ICT with which they are working (Cox & Marshall. 2007; Marshall & Cox, 2008).
- The impact on students' attainment, measured in terms of the abilities and knowledge that the innovation hopes to affect (Cox & Marshall, 2007; Penuel, 2005; Roblyer, 2005).

The above information can be obtained from interviews with teachers, classroom observation, expert revision, user observation and usability tests, where the participants vary from experts on the subject to students who are representative of the target audience (Reeves, 2008).

The implementation of summative evaluation at this stage must be long enough to provide evidence that the programme is sustainable through time, and doesn't present problems that threaten its systematic use (Slavin, 2008). Because very short studies have low external validity, some authors suggest a minimum duration of 12 weeks (Slavin, 2008).

2.2 Effectiveness

The intervention stage includes design and implementation of monitoring and evaluation plan among others activities. The main concerns during this stage are:

- The context into which the programme will be implemented (Bielaczyc, 2006; Coburn, 2003; McDonald et al., 2006), expressed in terms of the type and number of students (Penuel, 2005), school organization, the role played by the staff, teachers' experience, and equipment, so as to evaluate the changes over time (Marshall & Cox, 2008). The variables that measure the context must be related to ICT barriers (Infante & Nussbaum, 2010) or facilitators for the school (Rodríguez et al., 2010). The performed evaluation should link the particular learning environment with the broader context (Marshall & Cox, 2008).
- The intervention to be tested at this stage is designed based on the context, and includes activities to be performed and resources to be used. These activities are to be carried out within the school, i.e., diagnostics, working plan, training of the school's professionals, and the monitoring and evaluation plan. The resources are the ICT, materials, and documents (Rodríguez, 2008).
- The fidelity of the implementation, which measures the effectiveness of the intervention, defining critical teachers' practices along with the metrics of their performance (Coburn,

2003; Penuel, 2005, Rodríguez et al., 2010).

- The impact on students' attainment, in terms of the abilities and knowledge that the programme aims to affect (Cox & Marshall, 2007; Penuel, 2005).
- The relationship between context, variability of the implementation and results helps to understand the limits of the programme's applicability and thus to explain the variations in effectiveness during its implementation (Agodini et al., 2003; McDonald et al., 2006; Penuel, 2005).

Understanding the learners' context is essential. Several models can help to conceptualize the educational setting, by identifying key players and their relationships, as well as the facilitators and barriers that the ICT4E programme faces (Luckin, 2010; Nachmias, Mioduser, Cohen, Tubin, & Forkosh-Baruch, 2004; Plomp et al. 2007; Tearle, 2003, 2004). These models serve as a guide when designing the intervention. Also, during this stage, the cycle of formative studies validate the assessment instruments (Marshall & Cox, 2008; Penuel, 2005). Formative evaluations will gather valuable information regarding, for example, whether the students understood the material that was used, if they understood the instructions or had the cognitive maturity required for that task, whether the teachers' pedagogic practices were directed at all the students or not, and if the time used for instruction was enough to understand the tasks (Marshall & Cox, 2008). They can also shed light on which cognitive processes students used to solve problems (Marshall & Cox, 2008) and if the teachers adapted the innovation thus undermining the principles of its design (Penuel, 2005). Also, formative studies would resolve such questions as: what were the initial conditions? What critical factors delayed, precluded, or deformed the original plan?; In doing so, the key factors which affect implementation and intervention would also be determined. Thus, the qualitative data gathered through the summative study can explain the success or failure of the programme, allowing for its improvement (Chatterji, 2005; Marshall & Cox, 2008).

In the intervention, a monitoring and evaluation scheme should be defined; specifying indicators (input, process and outcomes), assessment instruments and a monitoring plan (Rodríguez et al., 2010). Input indicators measure context conditions, indicating which ones must be maintained over time in order for the programme to be sustainable. Process indicators measure two aspects: fulfillment of the intervention and adoption of the practices that the model promotes. Outcome indicators reflect direct results (e.g., the development of skills in teachers, or improvement in the students' attainment). Assessment instruments (e.g. learning tests, guidelines for classroom observation, and surveys) measure the indicators described above. Finally, the monitoring plan determines when the assessment instruments are applied, thus implementing process evaluation.

According to the scaling-up models, the testing of the innovation in several settings helps to discover possible context-intervention interactions and to identify the key context variables that consistently produce the desired impact (McDonald et al., 2006). These successive studies define the iterative cycle at the effectiveness stage of the EDM.

Regarding the duration of the implementation, previous research indicates that schools need time to adopt innovations (Cox, 2008), which differs depending on each TEL. Therefore, the implementation must be long enough for minimal adoption to take place, while also considering that pre-test effects must be controlled. These criteria must be considered by the designers and measured in the monitoring and evaluation plan.

2.3 Efficiency and Scaling-up

While efficiency should always be present in the design process, we will study it here because only at this point is the programme complete. The goal of this stage is to improve its cost-effectiveness, while the massive transference is carried out, and the total costs are calculated. The main elements that need to be evaluated are:

- The fidelity of the intervention in the school, to evaluate the effectiveness of transference to external professionals.
- The fidelity of the implementation by the teachers, expressed in terms of critical practices that they must implement in the classroom (Coburn, 2003; Penuel, 2005, Rodríguez et al., 2010).
- The total cost of the programme, including all the costs associated with the use of computers: hardware, software, administrative costs, licenses, installation, updating software and hardware, training and development, maintenance, technical support, and other costs associated with the acquisition, implementation, operation, maintenance and updating of an organization's computer systems (Moyle, 2008).
- The students' attainment, measured in terms of the abilities and knowledge that the innovation is expected to affect (Cox & Marshall, 2007; Penuel, 2005; Roblyer, 2005).
- The relationship between the variability of intervention, implementation and results, which

explains variations in the innovation's implementation, as well as the effectiveness of the transference.

Just as in the effectiveness stage, the implementation's fidelity is measured to ensure that the pedagogic model is being applied correctly. Here we measure the intervention's fidelity to the previously validated design. The changes in the intervention can affect the teachers' adoption, and therefore the results of final implementation. Evaluating fidelity of implementation and intervention, as well as the results, can establish the causal links between these elements.

Calculating the total cost of the programme requires defining tangible and intangible assets —such as people's skills— in operational terms (Moyle, 2008). For example, at the effectiveness stage we identified abilities and practices that must be developed. The costs associated with these intangible assets will be training for the teachers and the corresponding monitoring and evaluation plan to measure their adoption.

The implementation of summative evaluation should have a similar length to the one carried out in the effectiveness stage because the innovation requires time to be adopted. This study should use a random experimental design (or its quasi-experimental equivalent) for its rigorousness, calculating the expected effect size that the programme will have once implemented permanently. The linear hierarchical models for student, class and school level can help in the analysis of results (Agodini et al., 2003; McDonald et al., 2006; Slavin, 2008), but require appropriate sample sizes so that the statistical tests can detect the desired effect size, as seen in Agodini et al. (2003).

To study the possibility of replicating the programme in the education system we will use the methodology described by Margolis et al. (2006) which uses three attributes: benefits, cost, and feasibility. In EDM, the benefits (in terms of impact on students' attainment) and costs are calculated through impact studies carried out at this stage. With this information, it is possible to determine the feasibility of the educational programme, comparing if the cost is within the government's and schools' range of expenses. If it isn't, the cost effectiveness of the ICT4E programme's activities must be improved. However, after these changes it is necessary to evaluate whether or not the programme still maintains its effectiveness.

2.4 Summary and discussion of the model

Considering that the ICT4E programme is developed and refined through a process of several stages of empirical validation (i.e. it evolves), we can call this an evolutionary development model. Evolution also occurs in the dimensions of analysis, such as quantity (e.g., students, schools, and districts), variety of educational contexts, duration of evaluation studies, and the number of validated components of the ICT4E programme. Table 1 summarizes the main characteristics of each stage, determining the main research goals, the elements to be evaluated, and possible design studies.

In software engineering the terms "evolution" and "evolutionary" were introduced by Gilb (1981) and although many view iterative and incremental development as a modern practice, its application dates as far back as the mid-1950s (Larman & Basili, 2003). In this field, the so-called "evolutionary prototypes" (Carter, Anton, Willians & Dagnino, 2001; Davis, Bersoff &

Comer, 1988; Larman & Basili, 2003), focus on gathering a correct and consistent set of software requirements (Carter et al., 2001).

However, there are crucial differences between these evolutionary prototypes and our EDM. The evolutionary prototypes are limited to software alone, whereas our EDM is based around the development of a TEL built for a specific educational requirement which is then integrated into the school system at the various levels, (subject, classroom, school, districts, regional and national) using evidence-based research (Slavin, 2008).

Given that it does not rule any one of them out, our EDM enables integration with other already existing frameworks at various levels. For example, it is possible to use Evolutionary Software Prototyping (Carter et al., 2001; Davis, Bersoff & Comer, 1988; Larman & Basili, 2003) at stage 1 as a methodology for software development. At stages 1 and 2, frameworks can be used to develop the pedagogical model and the intervention process, as well as to study the context, for example, Activity Theory (Bottino et al., 1999), Analysis Schema (Mioduser, et al., 2003), Learner Centric Ecology (Luckin, 2008) and the Conceptual Framework for SITES 2006 (Plomp, Pelgrum, & Law, 2007). At stage 3, scaling-up models can be used, such as those proposed by Coburn (2003) and Dede (2006). Whichever model is used for the TEL, the intervention and the transference stages, its application and effects should be evaluated, taking into account any feedback which comes out of the evaluation process and building it into the design. At the end of each cycle, summative evaluations which provide solid evidence about the expected impacts should be undertaken.

3 Using the evolutionary development model to create the Eduinnova ICT4E programme

Here, we illustrate the development of Eduinnova's ICT4E programme using the model described in Section 2.

3.1 Efficacy

In Chile, at the time when the programme was designed, the computer to student ratio was 1:36 in elementary schools and 1:26 in secondary schools, with at least 6% of computers being placed within the classroom (Hinostroza, Labbé, & Claro, 2005).

The objective of the efficacy stage (section 2.1) was to develop a TEL that motivated students' active participation in the classroom (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003) through Computer Supported Collaborative Learning (CSCL) (Dillenbourg, 1999). The technological infrastructure that would have allowed a 1:1 ratio within the classroom was not in place in Chile at the time.

Therefore, the first stage of EDM was creating and testing a prototype of activities based on educational principles such as collaborative learning (CL) and constructivism using Personal Digital Assistants (PDAs). These devices offered a highly mobile, low-cost platform with wireless connectivity, making them ideal for working within the classroom. Thus, we designed a Mobile Computer Supported Collaborative Learning (MCSCL) environment to scaffold collaborative work in small groups, transforming the teacher's educational practices in the classroom (Cortez et al., 2005; Rodríguez et al., 2010).

We tested MCSCL activities with specific contents in controlled environments, comparing them with collaborative activities of the same nature, but without technological support (Zurita & Nussbaum, 2004a). Given the positive results for MCSCL activities in the first exploratory study, we carried out three more studies in order to evaluate the learning environment in terms of usability and representation of the material, making collaboration within the classroom practical for teachers. We also measured the efficacy of students' attainment at different levels, in different subjects (language, math and science) and in different contexts (primary and secondary education). A summary of these studies is shown in Table 2.

In terms of experimental design, all the studies were carried out in 1 school corresponding to a low socio-economic level, in Santiago, Chile. Small groups of students were observed using the CL activities, with and without PDAs, over 4 weeks during the year 2002. Both Pre and post tests were applied to measure attainment in Language and Mathematics, according to curricular contents. Specific assessment instruments were also used to evaluate other aspects in these studies.

Altogether, the efficacy studies allowed us to conceptualize how to create pedagogical models supported by MCSCL (Zurita & Nussbaum, 2007), which guided the future development of new applications and contents.

3.2 Effectiveness

At the second stage of EDM, we designed an intervention —using the pedagogic model defined in Cortez et al. (2005)— for the target educational context of the programme: state-subsidized schools, lower-middle socioeconomic strata and poor academic performance. The intervention was implemented and evaluated in two studies in 2004 and 2005-2006.

3.2.1 2004 study

Our first design was a course for teachers on content subjects (10th grade physics) and the use of the MCSCL applications, held before the beginning of the school year

The objective of the 2004 study was to evaluate the design of the intervention stage in a real world context and the impact on students' attainment, the TEL in this case being implemented autonomously by the teachers. This study was carried out during **2004** in **5 schools** (located in the cities of Santiago and Antofagasta) representative of the programme's target schools and which were not running any other external educational programme.

To gather its qualitative results, the study utilized a series of questionnaires and surveys directed at students and teachers, in-class observations (to establish how the teachers integrated the technology into classroom activities) and focus groups of teachers and students. The quantitative methods, on the other hand, included end of semester tests according to the structure of the Chilean curriculum.

More than 1,300 students participated in the study, under a quasi-experimental prospective design with 3 groups: experimental (EG), where students used the PDAs, and two control groups ICGa (same teachers as the EG, with no technology and different students) and ICGb (different teachers and students, with no technology). Student attainment was measured using a validated test (average difficulty: 0.48; Cronbach's α=0.87), which was divided in two according to the contents of each semester (Cortez et al., 2005; Nussbaum & Rosas, 2004).

The following conclusions were drawn from analysis of the qualitative data:

- **Context in which the programme was used:** There are inherent flaws in the type of schools used in the study, in areas such as: planning of classes, use of class-time, communication between teachers, the mastery of subject material and digital literacy, among others. For the pedagogical model of the TEL to work, specific attention must be paid to some of these deficiencies. Moreover, the programme must stem from the core of the organization to guarantee a minimum level of dedication from teachers involved in the programme. . This means clearly defining the roles of key players, while also involving the whole school community in the integration of the technology.
- **Teaching Orientation:** This TEL required a methodological change by the teacher, from a class with an emphasis on transferring subject matter, to one which involves collaborative small group work, where the teacher´s role becomes more that of a mediator. This transition is not an automatic one and therefore a certain amount of "coaching" is required to guide the teachers in their work in the classroom.

- **Technical issues**: we observed that the teachers achieved different levels of proficiency in technological and pedagogical practices, which affected the implementation of TEL in both form and frequency of use in the classroom. Other technological problems (such as lack of connectivity between the 45 student PDAs and the teacher's device) and logistical problems (such as lack of charge in the PDA batteries) also made the process difficult.

- **Research overhead:** A complex experimental design and/or overly frequent external measurements, as were used in this study, can overburden the teachers. It is important to allow the teachers to work under conditions as similar as possible to those of their typical classroom environment.

On the other hand, the quantitative analysis showed significant statistical results, with effect sizes for the EG-ICGb of 0.37 in the first semester and 0.88 in the second semester (see Table 3). However, there were no statistically significant differences between EG and ICGa during any semester which, when added to the fact that these differences were present between ICGa and ICGb (both without PDAs) and the important effect sizes (0.25 and 0.87 respectively), indicates that the impact could not be attributed to the TEL intervention (see Table 3).

3.2.2 2005-2006 study

As our first design for the intervention did not yield the expected outcomes in terms of impact, following the EDM we produced a second design based on the 2004 study results and the evidence found in literature. The 4 main changes were:

1. We made explicit, before the start of the project, the expected pedagogical practices (Penuel 2005) and teachers' skills (Cox & Webb, 2004) in terms of observable tasks and procedures.

2. Based on models for the successful professional development of teachers (Trucano, 2005) and in order to promote teachers' and schools' autonomy in taking charge of the programme (Coburn, 2003), the intervention was designed in 2 phases: previous training to

implementation and on-going, formal and informal pedagogical and technical support.

3. We incorporated a monitoring and evaluation scheme (section 2.2), enhancing instruments from the previous qualitative study, such as classroom observation, to analyse the tasks and procedures described in point 1 above.

4. The software was improved, making it more robust and tolerant to flaws, so that it would work more reliably in the classroom.

The details about this intervention and monitoring and evaluation scheme are discussed further in Rodríguez et al. (2010).

The objective of this second study was to evaluate the new, improved intervention design, as well as the relationship between the variable nature of the fidelity of the implementation (measured as the adoption of the TEL by the teachers assessed through the classroom observations) and the impact on student's attainment. The study was conducted during the 2005-2006 period, again in schools from Santiago and Antofagasta, with more than 1,600 students participating. This time, the assessments of 10th grade physics were at the beginning (pre-test) and at the end (post-test) of the year, so as to control previous differences between students, using the same instruments as in the 2004 study. The matched prospective quasi-experimental design considered 3 groups: Experimental Group (EG), Internal Control Group (ICG, same as ICGa in 2004) and External Control Group (ECG, students from schools without TEL, matched using variables such as socioeconomic level, funding sources and average national standardized test scores). The analyses compared the groups using the net improvement per student between pre and post-tests, shown in Table 3.

In the quantitative study, the EG-ICG effect size was 0.41 and the EG-ECG effect-size was 0.32 (Table 3).We may assert that the impact is attributable to the TEL because the comparison ICG-ECG was not statistically significant (Rodríguez et al., 2010). Regarding the relationship between teachers' adoption and students' attainment, the analysis revealed that the EG teachers with a high adoption rate produced better results in their students than teachers with lower adoption rates, elevating the effect size to 0.56 (compared to ICG) and to 0.45 (compared to ECG) (Rodríguez et al., 2010).

3.2.3 2007 and beyond

Using these two effectiveness studies of the EDM, we were able to better understand the TEL's adoption process in subsidized state schools. By the end of the second study, we achieved an intervention, with validated assessment instruments to measure adoption over time.

A third group of studies, using the validated design for the intervention, proved its effectiveness in tertiary education (Bustos & Nussbaum, 2009; Valdivia & Nussbaum, 2007, 2009) and in other countries: the United Kingdom (Nussbaum et al., 2009) and the United States (Roschelle et al., 2010).

3.3 Efficiency and scaling-up

Finally, in the third stage of EDM, during the years 2008-2009, dissemination of the programme began, involving **30 schools.** The costs were calculated with the improved intervention after the

2005-2006 study. This calculation was done by specifying a typical school where the technology would be applied and determining the total number of classes, teachers, and students that would implement the technology. In this way, the amount of material and human resources needed to satisfy a fixed number of schools was calculated, thereby determining the total cost of the programme per school and the cost per student. The feasibility analysis (see section 2.3) showed that the equipment was too expensive (approximately 75% of the cost per student) for the programme's target schools.

To use the schools' existing equipment to implement the programme, we used *multiple mice* (MM) (Pawar, Pal, & Toyama, 2006) so that several students could share a PC. In this way, three students collaborated to solve a problem on the same computer, each with his/her own input device. The efficacy studies showed results comparable to those obtained on individual platforms (Infante, Hidalgo, Miguel Nussbaum, Alarcón, & Gottlieb, 2009; Infante, Weitz, Reyes, Nussbaum, Gómez & Radovic, 2010). The usability evaluations carried out in both studies showed no significant changes regarding the use of PDAs. This environment was tested for effectiveness in preschoolers on the subjects of language and mathematics (Infante et al., 2010). Statistically significant differences were found, in favor of the experimental group, with effect sizes of 0.52 (t=4.165; p=0.000) and 0.66 (t=5.806; p=0.000) respectively (Rodríguez, 2010).

Thus, Eduinnova's unitary cost was reduced to 25.37% of the original cost, maintaining its effectiveness and making it affordable for the low-income schools targeted by the programme.

4 Conclusions and future work

As we discussed earlier in this article, measuring the impact of ICT4E programmes is still an unsolved problem and a research area that has yet to provide conclusive results (Cox & Marshall, 2007; Reeves, 2008). If we add to this the fact that there is no certainty about the theoretical background of many initiatives carried out (Earle, 2002; Marshall & Cox, 2008; McFarlane, 2001), nor about what specifically happened during their implementation (Marshall & Cox, 2008; Penuel, 2005; Reeves, 2008), the results found are not surprising. Even when an ICT4E programme demonstrates its effectiveness experimentally, the government decisions are weakly related or not related at all to this evidence (Cox, 2008). In general, the effectiveness of educational programmes is only mentioned to justify decisions that have already been made, or opinions that have already been formed (Slavin, 2008).

In this article, we described the evolutionary development model (EDM) of ICT4E programmes, establishing a definition of processes that considers the evidence indicated in the literature on the topic (section 1). The EDM (section 2) aims at the building and rigorous, iterative testing of each of the components of the ICT4E programmes in real educational settings, so as to produce cost-effective and sustainable solutions that can later be disseminated among schools. The proposed model is useful for researchers who design these innovations as well as for those who develop public policies for ICT in education. Within the proposed evolutionary cycle, this design can be improved and re-evaluated, until the specific goals of the component under study are reached.

The EDM allows us to iteratively ensure the internal consistency of the design, theoretical soundness, and alignment with the results expected of an ICT4E programme. It establishes a roadmap to create and validate ICT-based, pedagogic innovations for the classroom, incorporating their cost-effectiveness so that they are both technically and financially feasible. In the analyzed example (section 3), the formative and summative studies for the efficacy, efficiency and effectiveness stages used samples of 1, 5 , and 30 schools respectively, making it possible to test the programme's basic concepts, develop an intervention (associating adoption levels with learning results), calculate the total cost of the programme, and test technological alternatives that improved cost-effectiveness.

Using a process such as the one described, we can build ICT4E programmes based on the specific needs and realities of the schools. These different needs will require different technologically enhanced learning environments (TEL), perhaps resulting in the diversification of the equipment existent in schools, which currently only responds to external, uniform and standardized criteria, most often defined on a technological basis.

However, the formative and summative evaluations used in this methodology need time, especially to verify the depth, sustainability, spread and shift in reform ownership of the programme (Coburn, 2003). For this reason, the development of the ICT4E programmes, which should be continually refined over the course of years (Laurillard, 2007; Reeves, 2008) with continuing involvement of researchers, must not be restricted simply to the kind of generally available research funding which most often encourages short-term projects and the publication of scientific articles, and neither guarantees the necessary time commitment, nor covers the costs that such development implies.

Therefore, the public agencies responsible for ICT and broader education policies must be prepared to support this process. Our model recommends a potential system of grants directed at each stage of the model. The first must promote the design, development, and evaluation of pedagogic models based on ICT (stage 1); the second, intervention processes in projects that have previously proven their efficacy (stage 2); and the third, the packaging and improvements necessary to disseminate effective initiatives within the school system (stage 3) to finally produce ICT4E programmes which can then be used on a wide scale in the school system. Under this model of financing, projects are compared against others at the same stage, in order to determine whether or not they will receive further funding, based on, for example: their comparative cost-effectiveness (Harris, 2009), intervening subjects and their alignment with the objectives of educational system improvement, and the novelty of ICT used to support emerging technologies.

Furthermore, the incorporation of rigorous evaluation standards is necessary for the summative evaluations so as to compare the impact of the programmes at each stage. They are currently being produced by different initiatives, such as: *What Works Clearing House* (http://ies.ed.gov/ncee/wwc/), *Best Evidence Encyclopedia* (www.bestevidence.org), *Comprehensive School Reform Quality* Center (www.csrq.org), Campbell Collaboration (www.campbellcollaboration.org) and *Evidence for Policy and Practice Information and Co-ordinating Centre* (http://eppi.ioe.ac.uk/cms).

Finally, meta-analysis studies (Liao & Hao, 2008) can be carried out on all the ICT4E programmes once they have been disseminated in order to analyze the effectiveness of ICT developed in this way in a particular education system. These studies will complement large-scale and generic studies about the impact of ICT in the school system.

References

Agodini, R., Dynarski, M., Honey, M., & Levin, D. (2003). The Effectiveness of Educational Technology: Issues and Recommendations for the National Study. Mathematica Policy Research, Inc. Retrieved November 17, 2010, from http://www.mathematica-mpr.com/PDFs/edtechrec.pdf.

Anderson, R. E., & Plomp, T. (2009). Introduction. In *T. Plomp et al. (Eds.), Cross-National Information and Communication Technology Policies and Practices in Education* (Second Edition., p. 3—17). IAP-Information Age Publishing.

Balanskat, A., Blamire, R., & Kefala, S. (2006). The ICT Impact Report: A review of studies of ICT impact on schools in Europe. Retrieved November 17, 2010, from http://ec.europa.eu/education/pdf/doc254_en.pdf.

Barrera-Osorio, F., & Linden, L. (2009). The Use and Misuse of Computers in Education: Evidence from a Randomized Controlled Trial of a Language Arts Program1. Retrieved November 17, 2010, from http://www.povertyactionlab.org/sites/default/files/publications/Computers%20in%20Education.pdf.

Bielaczyc, K. (2006). Designing Social Infrastructure: Critical Issues in Creating Learning Environments with Technology. *Journal of the Learning Sciences*, 15(3), 301-329.

Bottino, R., Chiappini, G., Forcheri, P., Lemut, E., & Molfino, M. (1999). Activity theory: A framework for design and reporting on research projects based on ICT. *Education and Information Technologies*, *4*, 279-293.

Burns, T., & Ungerleider, C. (2003). Information and Communication Technologies in Elementary and Secondary Education. *International Journal Educational Policy, Research and Practice*, *3*(4), 27-54.

Bustos, H., & Nussbaum, M. (2009). An experimental study of the inclusion of technology in higher education. *Computer Applications in Engineering Education*, 17(1), 100-107.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts. Retrieved November 17, 2010, from http://ies.ed.gov/ncee/pubs/20094041.

Carter, R. A., Antón, A. I., Williams, L., & Dagnino, A. (2001). Evolving Beyond Requirements Creep: A Risk-Based Evolutionary Prototyping Model. In *Requirements Engineering, IEEE International Conference on* (Vol. 0, p. 0094). Los Alamitos, CA, USA: IEEE Computer Society.

Castro, C. (2004). Are new technologies better technologies? For whom? In *Adapting technology for school improvement: a global perspective* (pp. 39-54). International Institute for Educational Planning - Unesco.

Chatterji, M. (2005). Evidence on "What Works": An Argument for Extended-Term Mixed-Method (ETMM) Evaluation Designs. *Educational Researcher, 34*(5), 14-24.

Coburn, C. E. (2003). Rethinking Scale: Moving Beyond Numbers to Deep and Lasting Change. *Educational Researcher*, 32(6), 3 -12.

Conlon, T., & Simpson, M. (2003). Silicon Valley versus Silicon Glen: the impact of computers upon teaching and learning: a comparative study. *British Journal of Educational Technology*, 34(2), 137-150.

Cortez, C., Nussbaum, M., López, X., Rodríguez, P., Santelices, R., Rosas, R., & Marianov, V. (2005). Teachers' support with ad-hoc collaborative networks. *Journal of Computer Assisted Learning*, 21(3), 171-180.

Cox, M. (2008). Research IT in education. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 965-981).

Cox, M., & Abbot, C. (2004). ICT and attainment: A review of the research literature. ICT in Schools Research and Evaluation Series – No.17, DfES-Becta. Retrieved November 17, 2010 from http://publications.teachernet.gov.uk/eOrderingDownload/ICT and attainment.pdf.

Cox, M., & Marshall, G. (2007). Effects of ICT: Do we know what we should know? *Education and Information Technologies*, 12(2), 59-70.

Cox, M., & Webb, M. (2004). ICT and pedagogy: A review of the research literature. ICT in Schools Research and Evaluation Series – No.18, DfES- Becta. Retrieved November 17, 2010 from http://publications.becta.org.uk/display.cfm?resID=25813.

Culp, K. M., Honey, M., & Mandinach, E. (2003, October). A Retrospective on Twenty Years of Education Technology Policy. U.S. Department of Education, Office of Educational Technology. Retrieved November 17, 2010 from http://www.ed.gov/rschstat/eval/tech/20years.pdf.

Davis, A., Bersoff, E., & Comer, E. (1988). A strategy for comparing alternative software development life cycle models. *Software Engineering, IEEE Transactions on*, 14(10), 1453-1461.

Dede, C. (2006). Scaling Up: Evolving Innovations beyond Ideal Settings to Challenging Contexts of Practice. In In *R.K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences*. Cambridge Handbook of the Learning Sciences.

Dillenbourg, P. (2008). Integrating technologies into educational ecosystems. *Distance Education*, 29 (2), 127.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., et al. (2007). Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort. U.S. Department of Education, Institute of Education Sciences. Retrieved November 17, 2010, from http://ies.ed.gov/ncee/pubs/20074005/.

Earle, R. S. (2002). The Integration of Instructional Technology into Public Education: Promises and Challenges. *Educational Technology*, 42(1), 5-13.

Fan, X. (2001). Statistical Significance and Effect Size in Education Research: Two Sides of a Coin. *The Journal of Educational Research*, 94(5), 275-282.

Feinstein, O. N., & Picciotto, R. (2001). *Evaluation and Poverty Reduction*. World Bank Series on Evaluation and Development (Vol. 3). Washington, D.C: Transaction Publishers, New Brunswick, New Jersey.

Gilb, T. (1981). Evolutionary development. *ACM SIGSOFT Software Engineering Notes*, *6*, 17–17.

Harris, D. N. (2009). Toward Policy-Relevant Benchmarks for Interpreting Effect Sizes: Combining Effects With Costs. *Educational Evaluation and Policy Analysis*, 31(1), 3-29.

Harrison, C., Comber, C., Fischer, T., Haw, K., Lewin, C., Lunzer, E., McFarlane, A., et al. (2002). ImpaCT2: The impact of information and communication technologies on pupil learning and attainment. BECTA. Retrieved November 17, 2010 from http://partners.becta.org.uk/page_documents/research/ImpaCT2_strand1_report.pdf.

Hayes, D. N. A. (2007). ICT and learning: Lessons from Australian classrooms. *Computers & Education*, 49(2), 385-395.

Hennessy, S., Ruthven, K., & Brindley, S. (2005). Teacher Perspectives on Integrating ICT into Subject Teaching: Commitment, Constraints, Caution, and Change. *Journal of Curriculum Studies*, 37(2), 155-192.

Hinostroza, E., Labbé, C., & Claro, M. (2005). ICT in Chilean Schools: Students' and Teachers'

Access and Use of ICT. *Human Technology*, 1(2), 246-264.

Infante, C., Hidalgo, P., Nussbaum, M., Alarcón, R., & Gottlieb, A. (2009). Multiple Mice based collaborative one-to-one learning. *Computers & Education*, 53(2), 393-401.

Infante, C., & Nussbaum, M. (2010). Third order barriers to the integration of technology in the classroom. Working paper.

Infante, C., Weitz, J., Reyes, T., Nussbaum, M., Gómez, F., & Radovic, D. (2010). Co-located collaborative learning video game with single display groupware. *Interactive Learning Environments,* 18(2), 177-195.

Kirkman, C. (2000). A model for the effective management of information and communications technology development in schools derived from six contrasting case studies. *Technology, Pedagogy and Education*, 9(1), 37-52.

Kozma, R. B. (1991). Learning with Media. *Review of Educational Research*, 61(2), 179-211.

Kozma, R. B. (2003). Technology and Classroom Practices: An International Study. *Journal of Research on Technology in Education*, 36(1), 1-14.

Kulik, J. (2003). Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say. Technical Report, SRI International.

Lai, K. (2008). ICT Supporting the Learning Process: The Premise, Reality, and Promise. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary* *and Secondary Education* (pp. 215-230).

Larman, C., & Basili, V. (2003). Iterative and incremental developments. a brief history. *Computer*, 36(6), 47-56.

Laurillard, D. (2007). Modelling benefits-oriented costs for technology enhanced learning. *Higher Education*, *54*(1), 21-39.

Law, N. (2009). Curriculum and Staff Development For ICT In Education. In *Cross-National Information and Communication Technology Policies and Practices in Education* (Second Edition., p. 19—39). IAP-Information Age Publishing.

Levin, H. (1983). Cost-effectiveness: A Primer. New Perspectives on Evaluation. Thousand Oaks. CA: Sage Publications.

Liao, Y., & Hao, Y. (2008). Large -scale studies and quantitative methods. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 1019-1035).

Lim, C. P. (2002). A theoretical framework for the study of ICT in schools: a proposal. British *Journal of Educational Technology*, 33(4), 411-421.

Luckin, R. (2008). The learner centric ecology of resources: A framework for using technology to scaffold learning. *Computers & Education*, 50 (2) 449-462.

Luckin, R. (2010). The Ecology of Resources Model of Context: A Unifying Representation for the Development of Learning-Oriented Technologies In *Mobile, Hybrid, and On-line*

Learning, International Conference on pp. 47-52. Los Alamitos, CA, USA: IEEE Computer Society.

Margolis, J., Nussbaum, M., Rodríguez, P., & Rosas, R. (2006). Methodology for evaluating a novel education technology: a case study of handheld video games in Chile. *Computers & Education*, 46(2), 174-191.

Marshall, G., & Cox, M. (2008). Research Methods: Their design, applicability and reliability. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 983-1002).

McDonald, S., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-Up Exemplary Interventions. *Educational Researcher*, 35(3), 15 -24.

McFarlane, A. (2001). Perspectives on the relationships between ICT and assessment. *Journal of Computer Assisted Learning*, 17(3), 227-234.

Mioduser, D., Nachmias, R., Tubin, D., & Forkosh-Baruch, A. (2003). Analysis Schema for the Study of Domains and Levels of Pedagogical Innovation in Schools Using ICT. *Education and Information Technologies*, 8(1), 23-36.

Mooij, T., & Smeets, E. (2001). Modelling and supporting ICT implementation in secondary schools. *Computers & Education*, 36(3), 265-281.

Moonen, J. (2008). Policy from a global perspective. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (p. 1171—

1178).

Moyle, K. (2008). Total cost of ownership and total value of ownership. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 615-631).

Nachmias, R., Mioduser, D., Cohen, A., Tubin, D., & Forkosh-Baruch, A. (2004). Factors Involved in the Implementation of Pedagogical Innovations Using Technology. *Education and Information Technologies*, *9*(3), 291-308.

Nussbaum, M., & Rosas, R. (2004). Enlaces Móvil: Mobile technology in the classroom. Final report. Enlaces, Center for Educational Technology, Ministry of Education, Chile.

Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face Collaborative Scaffolding. *Computers & Education*, 52(1), 147-153.

OECD. (2006). Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us.

OECD. (2010). Educational Research and Innovation: Are the New Millennium Learners Making the Grade?: Technology Use and Educational Performance in PISA 2006.

Pawar, U. S., Pal, J., & Toyama, K. (2006). Multiple Mice for Computers in Education in Developing Countries. In *Information and Communication Technologies and Development, 2006. ICTD '06*. (pp. 64-71). Presented at the Information and Communication Technologies and Development, 2006. ICTD '06.

Penuel, W. R. (2005). Recommendations for Evaluation Research within Educational Innovations. Center for Technology in learning, SRI International. Retrieved November 17, 2010 from http://ctl.sri.com/publications/downloads/Rec_Eval_Research_within_Edu_Innovations.pdf.

Pilkington, R. M. (2008). Measuring the impact of Information Technology on Student's Learning. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 1003-1018).

Plomp, T., Pelgrum, W., & Law, N. (2007). SITES2006–International comparative survey of pedagogical practices and ICT in education. *Education and Information Technologies*, 12(2), 83-92.

Raudenbush, S. W. (2008). Advancing Educational Policy by Advancing Research on Instruction. *American Educational Research Journal*, 45(1), 206 -230.

Reeves, T. (1994). Evaluating What Really Matters in Computer-Based Education. In In M. Wild & D. Kirkpatrick (Eds.) Computer Education: New perspectives. (pp. 219-246). Perth, Australia: MASTEC. Retrieved November 17, 2010 from http://www.educationau.edu.au/evaluating-what-really-matters-computer-based-education.

Reeves, T. (2000). Enhancing the Worth of Instructional Technology Research through "Design Experiments" and Other Development Research Strategies. Retrieved November 17, 2010, from http://it.coe.uga.edu/~treeves/AERA2000Reeves.pdf.

Reeves, T. (2006). Design research from a technology perspective. In *J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.) Educational Design Research*. London: Routledge.

Reeves, T. (2008). Evaluation of the Design and Development of IT Tools in Education. In *J. Voogt. G. Knezek (Eds.), International Handbook of Information Technology in Primary and Secondary Education* (pp. 1037-1051).

Reynolds, D., Treharne, D., & Tripp, H. (2003). ICT - the hopes and the reality. *British Journal of Educational Technology,* 34(2), 151-167.

Robertson, J. (2002). The ambiguous embrace: twenty years of IT (ICT) in UK primary schools. *British Journal of Educational Technology*, 33(4), 403-409.

Roblyer, M. (2005). Educational Technology Research That Makes a Difference: Series Introduction. *Contemporary Issues in Technology and Teacher Education*, 5(2), 192-201.

Rodríguez, P., & Gómez, F. (2007). Mobile technology in the classroom: development of a world class product and standard for transforming pedagogical practices. Final report. CORFO Innova Chile 05CT5IP-02.

Rodríguez, P., Nussbaum, M., López, X., & Sepúlveda, M. (2010). A Monitoring and Evaluation Scheme for an ICT-Supported Education Program in Schools. *Journal of Educational Technology & Society*, 13(2), 166–179.

Rodríguez, P. (2008). *Standard for an ICT for education programme*. Enlaces, Center for Technological Education. Ministry of Education, Chile Retrieved November 17, 2010 from http://intranet.redenlaces.cl/index.php?id=11414&no_cache=1&uid=2549&param=publico.

Rodríguez, P. (2010). *A methodology for replication, scaling-up and transference of ICT supported pedagogical models*. PhD in Engineering Thesis. Department of Computer Science, College of Engineering, Pontificia Universidad Católica de Chile.

Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: impact on students' mathematics learning. *Educational Technology Research and Development*, 58, 399–419.

Sánchez, J., & Salinas, A. (2008). ICT & learning in Chilean schools: Lessons learned. *Computers & Education*, 51(4), 1621-1633.

Slavin, R., & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Systematic Reviews in Education. *Educational Evaluation and Policy Analysis*, 31(4), 500-506.

Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education--What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher*, 37(1), 5-14.

Squires, D., & Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11(5), 467-483.

Tearle, P. (2003). ICT implementation: what makes the difference? *British Journal of Educational Technology*, 34(5), 567-583.

Tearle, P. (2004). A theoretical and instrumental framework for implementing change in ICT in education. *Cambridge Journal of Education*, 34(3), 331-351.

Tondeur, J., Braak, J. V., & Valcke, M. (2007). Curricula and the use of ICT in education: Two worlds apart? *British Journal of Educational Technology*, 38(6), 962-976.

Trucano, M. (2005). Knowledge Maps: ICTs in Education. ICT and Education Series. Washington, DC: infoDev / World Bank. Retrieved November 17, 2010 from http://www.infodev.org/en/Document.8.aspx.

Valdivia, R., & Nussbaum, M. (2007). Face-to-Face Collaborative Learning in Computer Science Classes. *International Journal of Engineering Education*, 23, 434-440.

Valdivia, R., & Nussbaum, M. (2009). Using Multiple Choice Questions as a Pedagogic Model for Face-to-Face CSCL. *Computer Applications in Engineering Education*, 17(1), 89-99.

Van den Akker, J. (2003). Curriculum perspectives: An introduction. In *J. van den Akker, U. Hameyer, & W. Kuiper (Eds.) Curriculum landscapes and trends* (pp. 1-10). Dordrecht: Kluwer.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational Design Research*. London: Routledge.

Villanueva, C. (2003). Measuring ICT Use In Education In Asia And The Pacific Through Performance Indicators. Presented at the Joint UNECE/UNCTAD/UNESCO/ITU/ OECD/Eurostat Statistical Workshop on Monitoring the Information Society: Data, Measurement and Methods (Geneva, 8-9 December 2003). Retrieved November 17, 2010 from http://www.unece.org/stats/documents/ces/sem.52/6.e.pdf

Wagner, D. A., Day, B., James, T., Kozma, R. B., Miller, J., & Unwin, T. (2005). *Monitoring and Evaluation of ICT in Education Projects: A Handbook for Developing Countries*. ICT and Education Series. Washington, DC: infoDev / World Bank. Retrieved November 17, 2010 from http://www.infodev.org/en/Publication.9.html.

Wainer, J., Dwyer, T., Dutra, R. S., Covic, A., Magalhães, V. B., Ferreira, L. R. R., Pimenta, V. A., et al. (2008). Too much computer and Internet use is bad for your grades, especially if you are young and poor: Results from the 2001 Brazilian SAEB. *Computers & Education*, 51(4), 1417-1429.

Watson, D. M. (2001). Pedagogy before Technology: Re-thinking the Relationship between ICT and Teaching. *Education and Information Technologies*, 6(4), 251-266.

Zurita, G., & Nussbaum, M. (2004a). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers & Education*, 42(3), 289-314.

Zurita, G., & Nussbaum, M. (2004b). A constructivist mobile learning environment supported by a wireless handheld network. *Journal of Computer Assisted Learning*, 20(4), 235-243.

Zurita, G., & Nussbaum, M. (2007). A conceptual framework based on Activity Theory for mobile CSCL. *British Journal of Educational Technology*, 38(2), 211-235.

Zurita, G., Nussbaum, M., & Salinas, R. (2005). Dynamic Grouping in Collaborative Learning Supported by Wireless Handhelds. *Educational Technology & Society*, 8(3), 149-161.

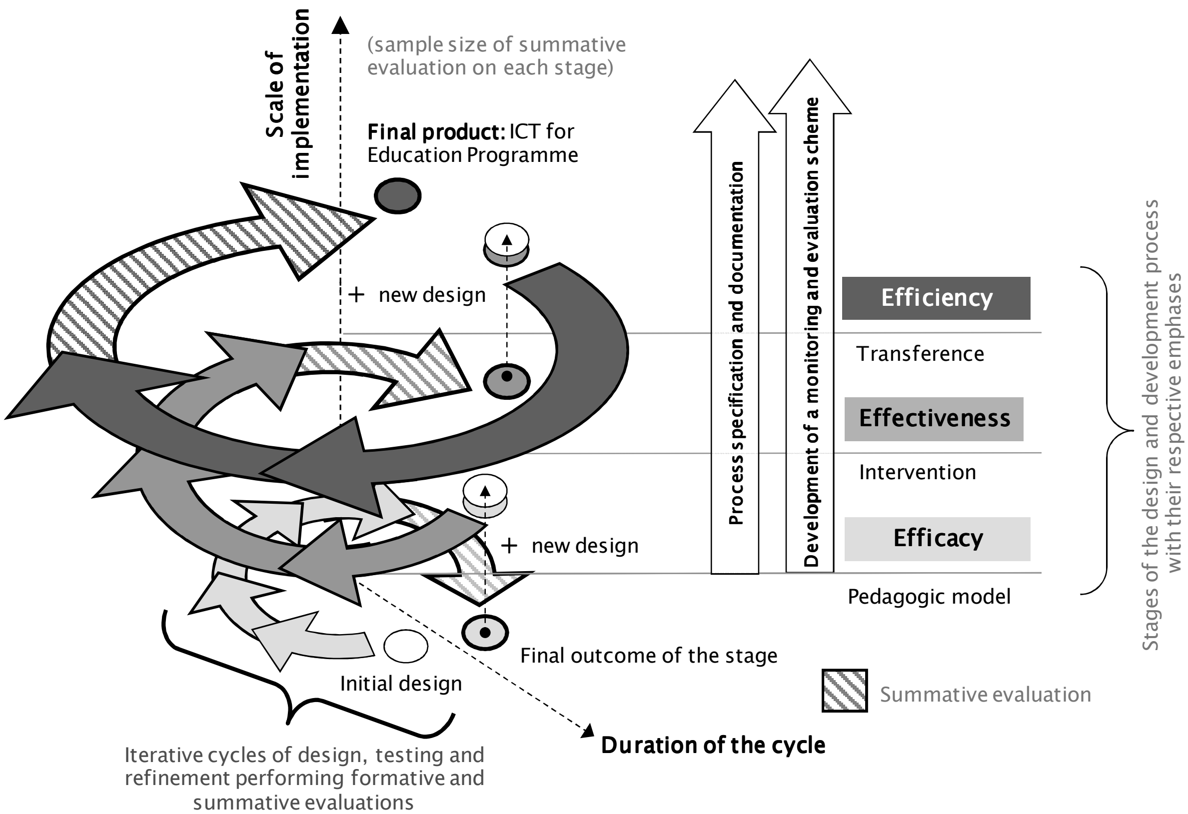

Figure 1: The evolutionary development model for ICT4E programmes

Table 1: Main research objectives (adapted from Marshall & Cox 2008), elements that are to be evaluated and possible design studies for each stage of EDM.

Stage	Main research goals	Main elements to be evaluated	Design study (summative evaluation)
1. Efficacy **(Pedagogic model)**	• Measure impact on learning • Assess effects on learning strategies and processes • Determine effects on teacher practice	• Pertinence • Assumptions it is based upon • Learning environment • Literacy in the ICT environment • Impact on students' attainment	• Matched prospective small-scale quasi-experimental design
2. Effectiveness **(Intervention)**	• Measure impact on learning • Determine adoption of the programme • Assess effects on learning strategies and processes • Determine effects on teacher practice	• Context • Intervention • Fidelity of implementation • Impact on students' attainment • Relationship between variability in implementation and results.	• Matched prospective medium-scale quasi-experimental design • Medium-scale randomized experimental design
3. Efficiency **(Transference and calculation of total costs)**	• Measure impact on learning • Determine adoption of the programme • Assess effects on learning strategies and processes • Measure attitudes towards ICT • Determine effects on teacher practice	• Fidelity of implementation and intervention • Total cost of the programme • Impact on students' attainment • Relationship between variability in implementation and results	• Large-scale randomized experimental design

Table 2: Efficacy studies carried out throughout the development of Eduinnova's programme

Objectives	**Grade and subject**	**Results** (EG = Experimental Group, CG = Control Group; δ = Effect-Size)
To demonstrate if constructivist MCSCL environments are created (Zurita & Nussbaum, 2004a)	First grade Language	All the evaluated principles can be seen in EG, while only a few can be seen in CG Impact on learning: EG > CG: δ post-test = 0.98*
To establish if MCSCL activities improve collaboration with technological support (Zurita & Nussbaum, 2004a)	First grade Language and Mathematics	All the problems seen without technology were overcome by using MCSCL Impact on learning: Mathematics: no difference; Language: EG > CG; δ post-test = 0.78*
To study the configuration of groups for an MCSCL activity (Zurita, Nussbaum, & Salinas, 2005)	First grade Mathematics	Different configurations have different impacts on the socio-motivational aspects evaluated in the study Impact on learning: EG > CG; δ post-test = 1.01*
Impact on students' attainment (Cortez et al., 2005)	Ninth grade Physics	Individual vs. collaborative comparison: δ test = 0.49

* No statistically significant differences were observed between groups in pre-test

Table 3: Summary of the results from summative evaluations, obtained at stage 2 of the Eduinnova programme development.

Study	**Results of the 10th grade physics test** (EG = Experimental Group, ICG = Internal Control Group; ECG= External Control Group)	**Independent sample t-tests comparisons*** (δ = effect-size)

2004 (Nussbaum & Rosas, 2004; Rodríguez, 2010)

Group	Post-test (%)					
	1st semester			2nd semester:		
	N	Mean	Std. Dev.	*N*	Mean	Std. Dev.
EG	553	56.52	14.82	665	46.49	13.15
ICGa	193	54.71	14.09	607	46.38	13.25
ICGb	581	50.34	18.28	155	35.19	11.70

- ICGa: same teachers as the EG, with no technology and different students
- ICGb: different teachers and students, with no technology

- First semester:
  - EG= ICGa (no significative)
  - EG> ICGb; t=6.26; δ post-test = 0.37
  - ICGa> ICGb: t=3.40; δ post-test = 0.25
- Second semester
  - EG= ICGa (no significative).
  - EG> ICGb: t=10.56; δ post-test = 0.88
  - ICGa > ICGb: t=10.33; δ post-test = 0.87

2005-2006 (Rodríguez & Gómez, 2007; Rodríguez et al., 2010)

Group	N	Pre-test (%)		Post-test (%)		Improvement (%) [gain per student]	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
EG	274	29.52	9.89	42.70	10.75	13.17	13.59
ICG	130	28.41	12.65	36.07	10.76	7.66	13.28
ECG	87	29.18	12.56	37.93	14.76	8.75	14.45

Pretest: EG = ICG= ECG (no significative)

Improvement:

- EG > ICG; t=3.83; δ gain = 0.41
- EG> ECG; t=2.60; δ gain = 0.32; (1-β=0.74)
- ICG= ECG (no significative)

* With $p < 0.01$ and statistical power above 0.8 (two tails), except where indicated.