A Spatio-Temporal Model for the Evaluation of Education Quality in Peru

by

Juan P. Alperin

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Arts
in
Geography

Waterloo, Ontario, Canada, 2008

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The role of information and communication technologies in the development of modern societies has continuously increased over the past several decades. In particular, recent unprecedented growth in use of the Internet in many developing countries has been accompanied by greater information access and use. Along with this increased use, there have been significant advances in the development of technologies that can support the management and decision-making functions of decentralized government. However, the amount of data available to administrators and planners is increasing at a faster rate than their ability to use these resources effectively.

A key issue in this context is the storage and retrieval of spatial and temporal data. With static data, a planner or analyst is limited to studying cross-sectional snapshots and has little capability to understand trends or assess the impacts of policies. Education, which is a vital part of the human experience and one of the most important aspects of development, is a spatio-temporal process that demands the capacities to store and analyze spatial distributions and temporal sequences simultaneously. Local planners must not only be able to identify problem areas, but also know if a problem is recent or on-going. They must also be able to identify factors which are causing problems for remediation and, most importantly, to assess the impact of remedial interventions. Internet-based tools that allow for fast and easy on-line exploration of spatio-temporal data will better equip planners for doing all of the above. This thesis presents a spatio-temporal on-line data model using the concept or paradigm of space-time. The thesis demonstrates how such a model can be of use in the development of customized software that addresses the evaluation of early childhood education quality in Peru.

Acknowledgements

When four years go by between the time you started your Masters and the time your thesis is accepted, it is inevitable that there are many people to thank. Such is the case for this Masters student, for whom writing a thesis flowed about as quickly as molasses.

First and foremost, I must thank my supervisor, Dr. G. Brent Hall, who continued to believe that this was possible, even when I no longer did. He has been my mentor, my travel companion, my therapist, and occasionally, my need for therapy. I am indebted to him for all the time and effort he spent ensuring I actually *learnt* from this process.

Secondly, it would be impossible to neglect thanking my fellow slaves, Michael G. Leahy and Nathan J. Engler. Their hard work made EduCal a reality. More importantly, they were able to empathize over how hard this all was, bet beers over who could write faster, and stay in the lab enough to make me feel guilty about leaving. They have been invaluable resources and good friends over the years, and certainly into the future.

The work and research in this thesis and the related EduCal project was made possible by funding provided by the Ford Foundation. I would also like to acknowledge the work and partnership of the Peruvian NGO, Alternativa.

The rest of the acknowledgments are not for work directly related to the thesis, but they are the people who managed to help me maintain my sanity, my motivation, or even those who just made me feel better about not doing anything, thereby letting me get to the point where I once again could.

My parents, Luis and Norma, who kept ~~not so~~ subtly suggesting that I keep at it and for supporting me constantly, through everything I do. My siblings, Cecilia and Will, who continue to teach me everything between how to be an academic to how to make Manhattans. Tambien, gracias a mis tíos, tías y a mi abuela Chela, por las comidas, las oficinas, y las camas que me han permitido concentrarme sin las preocupaciones de la vida cotidiana.

In somewhat chronological order, I would also like to thank the following friends and loved ones: Lisa, for being my partner in being lost and found at the same time, as well as for having shaped me into a socially responsible individual. Michelle, without whom I would not have left

Uganda to end up back in Waterloo to start this whole process; Michel, who got me through a rough patch, accompanied me during the moments of quitting and during a spectacular summer, and who made me feel smart by losing at Trivial Pursuit; Christina, who put up with me during some awful moments, helped me be gainfully employed, and who crocheted me a blanket; Patricia Fernandez, por la compañía en tierras desconocidas y por la *a la cena*; Maura Patricia, por el amor y el cariño este último año y por ser la que más ha vivido la tesis a mi lado; and Ali, for getting excited about my stupid ideas and for giving me the famous phrase *its practically writing itself*, even though I accidentally believed it for the first three years.

Finally, a big thank you to everyone that has put up with my arrogance, cynicism, loudness, short fuse, long rants, and few showers over the last four years, without ever wavering in your love and support. Do not think for one second that everything you have done for me has gone unnoticed.

And one last thank you to all, for never again asking *how's the thesis coming along?*

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

The growth of information and communication technologies (ICT) over the last ten years has played a pivotal role in the development of modern society. As these technologies have become more widespread in their presence and adoption, there has been an associated increase in the collection of data and substantial innovations in the means for storing, retrieving and sharing data across various types of analog and digital networks. The adoption and integration of information technology into almost every facet of daily life continues to serve both as propellant for further ICT innovation as well as a facilitator of exchange of information within today's global economy.

Within the context of the emergent network society (Castells, 2000), the popularization of the Internet has been instrumental in the spread of information by connecting over a billion users (an estimated 1,114,274,426 in 2007) (Nielsen//NetRatings, March 10, 2007) who are both producers and consumers of data. In order to accommodate the incredible growth in the number of Internet users (estimated at 208.7% between 2000 and 2007) (Nielsen//NetRatings), the Internet infrastructure has also grown to the point where there are currently approximately 22,474,669 Web Servers (SecuritySpace, March 1, 2007) that provide information to Internet users. Internet backbones (high capacity lines that form the main arteries in the network) have also increased in capacity and in number. The huge investment in the physical infrastructure of this system increases the ease and speed with which information can be transferred between any number of hosts, thus acting as a catalyst for information exchange.

The growth both in size and importance of the Internet has entrenched the organization of global society into what Castells (2000) coined as a 'network society'. The term network society refers to the organization of societies in such a way that the degree of 'connectedness' (mainly through lines of communication) is significantly more important than physical distances in terms of determining human interaction. Although this structure relies heavily on ICTs for it to be able to function, it remains as bound to cultural, economic and political factors as previously existing social structures. That is, the network society concept or paradigm does not replace previous social paradigms, rather it transforms them to be connected through the medium of real virtuality (Castells, 2000). Real virtuality is defined by Castells (2000; p.404) as "a system in which reality itself (that is people's material/symbolic existence) is entirely captured … in the world of make believe, in which appearances are not just on the screen through which experience is communicated, but they become the experience". Thus, the network society can exist anywhere that real virtuality can be found. The central concept to the network society is the fact that membership is open (and exclusive) to those segments of society, independent of the level of economic development, which are able to become connected to the network that binds its disparate geographic, social and political nodes together.

Contemporary societies are therefore characterised not only by the presence of ICTs, but also by conscious and unconscious attempts at establishing a place on the network. ICTs themselves have not caused these social changes to come about, but without the presence of ICTs the changes would not be possible (Castells, 1998). It is apparent that ICTs and the associated data that flow between nodes on communications networks impact all sectors of global society in a direct (i.e. by their incorporation in daily operations) and indirect manner (i.e. through their effects on economic development). The perception that ICTs improve efficiency has bought about changes in organizations at all levels of social development from the family to national governments and international agencies. These changes have resulted in a shift toward the incorporation of ICTs into daily organizational functions and often also into structures, and this, in turn, has had a domino effect that has resulted in information driven economies, and the need to become information rich in order to advance on the global stage.

The integration of technology and mass acceptance of ICTs as tools for improving the efficiency of information management through the creation of information systems has lead to the collection of extremely large volumes of digital data resources being available to individuals and organizations via the global accessibility of the Internet. The high demand for ICT access has lead also to a decrease in the prices of computers and other communications device hardware (such as mobile telephones) and software, allowing individuals and small organizations to become users as well as, in some cases, producers of information. The result is the largest collection of data and information resources to exist in the history of humanity.

One of the most important ways that this enormous collection of data and information has helped to improve the efficiency of management society is through its effects on planning and decision-making. In this context, digital information is used as a resource for everyday decisions such as form of dress depending on downloaded weather forecasts, for routes chosen for travel due to downloaded road conditions, as well as far longer term and influential choices that fall within the domain of local, regional and national level planning decisions. The improved capacity for decision-making can lead to a variety of desired and potentially also undesired outcomes. In the case of education planning, two examples of desired outcomes are a more cost-effective education system and a greater overall quality of education. The latter area of concern is the specific focus of the research investigated in this thesis.

1.2 Existing Planning Tools

In order to avoid ad hoc decision-making that may have no grounding in reality, the objective is to equip planners with information and tools that allow them to assess reality using the digital data the foundation. Planners can then user prior knowledge and experience in combination with this discovered knowledge in order to make decisions. However, the quality of the decisions they make is constrained, in large part, by the quantity and quality of the knowledge they are able to attain about the issue they are dealing with. Although, currently, more data are being collected about all manner of things than ever before, there is a need to turn these data into information and ultimately into knowledge that can be used to inform the decision processes. In order to accomplish this, planners have become engaged in using a variety of information processing tools, including use of the ICTs to extract knowledge from large data sets, in an easy and timely manner.

This need has lead to the development and improvement of decision support systems (DSS). A DSS can be formally defined as a computer system that support decision makers in unstructured or semi-structured situations to analyze complex information that helps them to make better informed decisions (Scott Morton, 1971). Such systems are best understood as what they state, namely a means of supporting decision processes rather than taking them over. Although DSS were originally used for financial portfolio management and then business management (Power, 2007), they have gained popularity in areas such as education, medicine and health care, the military, environmental policy, and other areas that involve risk management, resource allocation under conditions of scarcity or constraint, and similar tactical and strategic decisions (Power, 2007). However, despite their increasing use, these systems have not yet reached maturity and continue to lack important knowledge extraction capabilities. In particular, there is still a shortage of DSS tools that adequately include the spatial and temporal dimensions that affect many areas of application such as those mentioned above, among others.

Spatial data are an important component of information systems since as much as 80 percent of data have a spatial component (Klinkenberg, 2003). In this context, the term spatial refers to a geographic characteristic that defines where something is in real world space both in absolute terms and also in relation to other similar and different things. Geographic information systems (GIS) were developed to accommodate storage and analysis of spatial data. GIS continue to be developed and improved for the purpose of better integrating, storing, editing, analyzing, and displaying geographically-referenced information.

In the mid-1980s, with substantial improvements in computer architecture and processing power, the proliferation of DSS evolved into the domain of spatial problem solving and the concept of a spatial decision support system (SDSS) was coined (Armstrong, Densham, & Rushton, 1986). By the early 1990s, SDSSs were typically included in reviews of the GIS field (Densham, 1991) and today SDSS are routinely used in many application domains. Despite this increased activity surrounding SDSS, spanning more than fifteen years, there are still few truly successful implementations in practice. Moreover, the temporal aspect of decision-making is essentially non-existent in SDSSs, given the prevailing emphasis on dealing with spatial rather than spatio-temporal data and problem solving that involves consideration of both space *and* time.

In addition to describing absolute and relative distributions, spatial data can describe conditions on, below, or above the surface of the earth. While these data are useful for addressing a variety of decision problems, they are of limited use when processes that occur over both time and space are of interest to a decision maker. Numerous examples exist of spatio-temporal problems, that is, problems that have a spatial extent and temporal duration. For example, education planning requires the use of information from both the temporal and spatial dimensions, since education systems are distributed over space and education itself is a temporal process. The inclusion of time, or process-oriented phenomena, changes spatial aspects of school performance into a spatio-temporal problem and this increases the complexity of derived models and their analysis. Since this thesis focuses on spatio-temporal analysis of school-based education quality, the nature of education as a temporal process is now briefly discussed.

1.3 Education

Education must be thought of as a process, rather than as a static concept. As such, changes in the quality of the education received by students must be assessed continuously through the course of their school careers and through the sequence of planning cycles that characterise school administration and funding. The quality of education that a student receives cannot be determined only by the status of the academic institution they attend or by their experience at it for a single time period such as a day, a week, a semester or an academic year. Rather, students' education quality is a function of the cumulative experiences gained throughout their schooling experience across multiple years. Hence, to assess quality, it is important to be able to determine whether problems in the-education process are recent or on-going, as well as to determine which schools have declining or improving performance in quality-related indicators over time. These types of time-related questions are a critical part of education quality assessment and education planning. In this sense, education planning must be both proactive (must take into account future needs) and reactive (must manage problems and resources made available by the past situations). In order to introduce systemic change, with the aim of improving the educational experience of students in the system at any given instance of time, as well as across multiple time periods, there must necessarily be a consideration of pattern (space) as well as process (time). In this way schools systems can evolve with efficiency as well as quality as the cornerstones of their growth.

Time plays another important role in planning. Specifically, the ability to assess the outcomes or impacts of an intervention allows planners to learn which strategies are effective for remedying each situation. Without having comparisons over time within different areas, such impact assessments are impossible to evaluate, as there is no baseline for comparisons. As an example, the success of education quality improvement plans is dependent in large part on the ability of planners to assess, intervene, and then reassess the situation in the schools they are responsible for.

The formulation of effective education quality improvement plans should play a key role in the form and function of the approach that is used to achieve progress in education delivery in economically and socially lagging countries. Education is considered to be a fundamental human right as encapsulated in the United Nations Charter of Human Rights (UN, 1948), and it is well understood that it is an important means for the social development and for the growth of societies (UNESCO, 2002). This makes the improvement of education systems, through improving the quality of education received by children, a priority throughout the world. This improvement is of particular importance in developing countries, especially in those places where there is significant sub-national variability in education standards and where there are many children who do not receive an education that meets minimum international standards.

Such countries typically have limited human resources, and in these contexts good planning has the potential to create a significant and positive impact on the education system, not to mention a basis for future growth of the national economy as well as the foundations for a democratic and stable society. Equivalently, there is no room for mismanagement, and under the presence of a shortage of resources a fine balance must be found between education provision relative to other priorities in human and environmental health and economic development. In order to address these concerns, there is an important and immediate need to develop tools to support decision-making within the education sector.

1.4 A Way Forward

As stated earlier, GIS and SDSS are used widely in economically advanced nations. However, these technologies fall short of addressing spatio-temporal problems in three inter-related ways. First, as mentioned above, both conventional GIS and SDSS give decision-makers static, cross-sectional 'views' of reality for a single instance or period in time. Second, few available GIS or SDSS allow decision-makers to access spatially and temporally aggregated and disaggregated data, as conventional spatial databases lack an explicit temporal dimension in their schema or conceptual design. Third, even with a the creation of a specifically devised spatio-temporal database schema, the tools available in virtually all current GIS and SDSS are geared either only or primarily toward spatial analysis and do not easily accommodate investigations of relationships over space *and* time without substantial customised programming.

In traditional DSS, aggregated and disaggregated data were first made accessible through a series of tools collectively known as on-line analytical processing (OLAP) (Codd, 1995). In this context, the term "on-line" is synonymous with "on-demand" and OLAP is the complement of the more familiar relational database concept of on-line transaction processing. OLAP tools have grown to the point where they are standard functions in virtually all enterprise-level databases. They facilitate monitoring and analysis of data that are mined from the multiple dimensions typically stored in such archives. However, OLAP tools alone are not sufficient to address all of the concerns regarding spatio-temporal data outlined above as they were initially developed for databases in which space was not a consideration, while time was restricted to organisational concepts such as payroll periods and strategic planning analysis.

To extend the concept of OLAP into the realm of both space and time Bédard (1997) introduced the concept of spatial OLAP (SOLAP), by adding an explicit spatial dimension to the other dimensionalities of large databases. In addition Bédard and his colleagues at Laval University have, since 1997, worked consistently to program OLAP-type functions that apply specifically to the spatial dimension and that facilitate rapid data mining to produce spatio-temporal knowledge that would otherwise be very difficult to discover. Despite these important achievements, SOLAP tools have yet to become a mainstream technology that is within the reach of government planners who have to deal with complex spatio-temporal data that characterise important activities such as the delivery and assessment of education. The concept and implementation of SOLAP tools and functions offer substantial advantages in this regard. Unlike their aspatial equivalent, SOLAP tools and functions provide access to aggregated and disaggregated spatial data through the provision of "roll-up" and "drill-down" functions that allow users to drill vertically within the spatial dimension of a multidimensional database. Moreover, using a combination of drilling down into space and "drilling across" through a temporal dimension, it is possible to refine queries successively to extract important and previously unknown information for planning decision making.

This thesis implements and uses a SOLAP-based approach to develop a multi-dimensional database schema and associated model, and enriches it with highly flexible query functions derived from data mining to address both time and space in the analysis of school-based education performance. The model and its associated functions provide information that is fundamental to the planning tasks of monitoring and adjusting need-driven responses to improving the quality of early childhood education. The final requirement in this process is a medium with which to deliver such a SDSS to its users.

Associated with the growth of interest in the concept of DSS in the spatial domain, interest has become firmly entrenched in the use of the Internet and the World Wide Web (Web) as a facilitatory medium to promote access to and participation in decision-making (Kinsgton, Carver, Evans, & Turton, 2000; Carver, Evans, Kingston, & Turton, 2001; Adrienko, 2001). As noted earlier, the emergence of the Internet and the associated Web as the prevailing ICT has had profound implications at all levels of all societies (Castells, 2000). These implications are especially significant for developing countries where the presence of a digital divide in technology access has had a paradoxical impact. On the one hand the relative absence of Internet access in most developing countries, especially in countries in Latin America and Africa, has had the effect of damaging their prospects of advancing economically by curtailing their full interaction with economically advanced societies in the global economy. However, on the other hand, rather than damaging the prospects for advancement of specific countries, the Internet can equally be seen as an opportunity for promoting and facilitating decentralised administration and planning by utilising the modest but generally good coverage of Internet points of presence through the telephone systems that exist with all countries.

Hence, use of the Web constitutes a medium through which geographically remote communities can access rich electronic data resources about their own communities and, with the assistance of tools built around the concepts of SDSS and SOLAP, produce improvement plans for activities such as education. The task required to facilitate this is to develop tools that are accessible, easy to use, sophisticated while not being baffling, and easy to understand.

Local and regional branches of Ministries of Education under a decentralized government structure, for example, are extremely well situated for utilizing such tools for enhancing social development. Hence, a Web-based SDSS focused on education quality and its assessment, that operates over the Internet, can serve to transfer administrative responsibility to local (district) and regional levels of decision making. In addition to bringing the decision making and data closer together, respectively to their points of authority and collection, such an approach also provides potentially the flexibility of allowing individual schools, school administrators, teachers, parents associations, researcher and individual parents to participate in decision making, as long as they can access a computer that is connected to the Internet.

The ubiquitous nature of this core ICT even within rural areas of some of the least developed countries in the world, and the prospect that over time this penetration will increase, makes the reality of an Internet-based SDSS for decentralised education planning and management a viable possibility. The ideas, tools, and the data model presented in this thesis seek to turn this possibility into a reality for the first time.

1.5 Thesis Objectives

This thesis uses the nation of Peru as a case study. The intent is to construct an information-based suite of decision support tools that improve Peruvian education planners' abilities to evaluate the quality of school-based education by accessing and analysing school performance data within a multi-dimensional spatio-temporal data model that represents the various levels of the Peruvian education system. Moreover, the approach used facilitates potentially much greater participation in education planning than would otherwise be possible as both highly disaggregated data and tools with which to analyse them are available to anyone with access to the Internet.

The thesis satisfies this general intent simply by harnessing the power of the Internet and the Web, and integrating large education databases and SOLAP tools within an online SDSS. More specifically, the thesis shows that despite the importance of good education planning for economic and social development, education planners currently lack tools with which to perform relevant data analysis. It is argued that development of a functional spatio-temporal education data model and associated knowledge generating tools can positively contribute to fill this gap in education planning.

The specific objectives of the thesis are as follows:

- To specify for the case study of the nation of Peru the design of a data model, which is both powerful and intuitive to use for the extraction of spatio-temporal knowledge from large databases.
- To build a practical, functional instance of the spatio-temporal data model using largely freely available data published online by the Peruvian Ministry of Education.
- To build a Web mapping interface that allows education planners to access comprehensive spatio-temporal data within the data model.

- To enhance the Web interface with intuitive tools that allow data mining operations, specifically drill down (disaggregate), roll up (aggregate), drill across, slice and dice (section the model), to be used to execute simple and complex queries over both space and time.
- To apply all of the above in developing a comprehensive assessment of current spatial and temporal education conditions for a study area within Peru.

It should be noted that while Peru is the case study used for the thesis, the proposed data model and tools are sufficiently generic in nature to be applicable to the education systems in any country, independent of its state of social and economic development. This model must be sufficiently robust to facilitate the extraction of knowledge from complex questions that may be posed from a wide range of perspectives (both expert and non-expert users). Moreover, it is equally important to note that the application that is developed is implemented using strictly Open Source (OS) software components derived from the spatial and aspatial communities, supplemented with extensive OS programming to substantially enhance the core modules and orientate them toward the education domain.

Special emphasis here is given to the extraction of *knowledge*, as opposed to information, from data. In this context, knowledge is viewed as the final form of the transformation of data into information and then to knowledge-based understanding of complex causal and associated factors within the education domain.

1.6 Thesis Structure

In order to achieve the above objectives, the thesis is broken into four subsequent chapters. The following chapter lays the foundations for the topic of education quality by outlining the outcome of the major international conferences on education. Chapter 2 also explores the concept of education as a spatial and temporal process, as well as the nature of information systems that can be used for education planning. Chapter 3 outlines an operational data model that meets the needs of education planners in the assessment of Education Quality, involving the dimensions and concepts previously set out in Chapter 2. With the education-related concepts and data model in place, Chapter 4 describes the high-level design of the st-DSS tool, EduCal. The open source technologies chosen, the interaction between the components, the database model, and the user interface are all described in detail. The EduCal tool is subsequently used to conduct a comprehensive analysis of the conditions found within the department of Cusco, in Peru, from the perspective of a regional education planner. In order, these four chapters present the reasons and motivations for the creation of a new educational planning tool, the conceptual model needed for its creation, the details of how open source technologies can be combined to implement the conceptual model, and example of how this implementation can be used for the creation of new education plans.

Chapter 2

Education Planning, Space, Time and Decision Support Systems

This chapter provides the conceptual and technical foundations for the thesis by sequentially addressing the central issues of early childhood education planning in space and time, and the representation of these in DSS. These issues are considered first in isolation of each other and then integrated in the ultimate section, where a comprehensive conceptual framework for spatio-temporal education assessment is presented.

The discussion commences with a discussion of early childhood education in developing countries. A summary of the objectives and outcomes of successive international conferences on Education For All (EFA), hosted by the United Nations Education and Scientific Co-operation Organization (UNESCO), is provided first and the concept of education quality is then explored. Following this, time, space and their presence in decision support systems are introduced both as independent concepts and then combined together. Finally the chapter integrates the initial review with a data-centred discussion to present and discuss the framework that is used for the remainder of the thesis.

2.1 Education For All

Education is a vital part of human experience. Hence, the presence of accessible basic education is an important and necessary condition for individual development and for the cumulative effects of education to be reflected in long-term sustainable development of societies. The importance of providing an accessible and good quality early childhood education in all countries has been the subject of several international conferences over the last fifteen years. In fact, the consensus reached at the two primary world education summits, namely the *World Conference on Education for All* (WCEFA) in Jomtien, Thailand (1990) and the *World Education Forum* (WEF) in Dakar, Senegal (2000), serves as a testament to the level of international agreement on the need for and importance of improved primary education systems on both ethical and economic grounds.

The concept underlying the first of these summits, namely Education for All (EFA), has served as the driving force behind all global childhood education planning and development since that time. The goal of the WCEFA was to set an agenda for all participating countries to meet the basic learning needs of all children, youths, and adults by the year 2000 (UNESCO, 1990). The term, 'basic learning needs', was used throughout their discussion and is explicitly defined as:

> "Basic learning needs...comprise both essential tools...and the basic learning content required by human beings to be able to survive, to develop their full capacities, to live and work with dignity, to participate fully in development, to improve the quality of their lives, to make informed decisions, and to continue learning." (UNESCO, 1990)

This definition captures the idea that education must be empowering, while, at the same time, implying that a broader definition of education must include both skills development and personal growth. The definition suggests also that access to education is a fundamental human right and that receipt of a fulfilling and good quality education is a basic prerequisite for individuals to be able to enhance their own quality of life while also contributing to improving the overall standard of the society they live in.

The accompanying *Framework for Action to Meet Basic Learning Needs* that emerged from the WCEFA recommended that countries lay out specific priorities, plans and time-bound objectives, that sought to produce improvements in early childhood education from international through national to local levels within each of the participating nations. In order to guide policy makers, six target objectives were identified (UNESCO, 1990), namely:

(i) to expand early childhood care and developmental activities, including family and community interventions, especially for poor disadvantaged and disabled children,

(ii) to achieve universal access to, and completion of primary/basic education by the year 2000,

(iii) to improve learning achievement such that an agreed percentage (e.g., 80% of 14 year-olds) attains or surpasses a defined level of necessary learning achievement,

(iv) to reduce rates of illiteracy (e.g., to half of 1990 levels by the year 2000), with emphasis on female literacy to reduce current disparities,

(v) to expand the provision of basic education and training in essential skills required by youth and adults, with program effectiveness assessed in terms of behavioural changes and impacts on health, employment and productivity,

(vi) to increase the acquisition by individuals and families of knowledge, skills and values required for better living and development, made available through all education channels, including mass media and other forms of modern and traditional communication, and social action, with assessments of effectiveness made in terms of behavioural change.

Achievement of the above objectives was to involve the active participation of agencies from many sectors and levels of government, multilateral lending agencies, development organisations, non-governmental organisations (NGOs), families, teachers, schools, and students. With this understanding, the general agenda specified at the conference, and a commitment by governments and lending agencies to shift priorities to increase education funding, the goals of EFA in 1990 seemed possible.

However, despite high initial optimism, the aspirations of EFA were largely unfulfilled by the target year of 2000. The assessment of EFA, in preparation for the WEF summit, confirmed that although some countries showed significant improvements at the national level (e.g. improved school enrolment rates, decreases in illiteracy, and participation of young girls in the education process), these gains were highly variable spatially both between and within countries (UNESCO, 2000b).

Hence, a new document was produced at the WEF, named the *Framework for Action*. The Dakar Framework (DF) outlines six goals, all of which were intended to reaffirm the vision established in Jomtien in 1990, as well as commit the 180 participating countries to take action. It is important to note here, that while the WCEFA established a broad framework for setting ambitious targets, it lacked clear national and sub-national strategies that would allow these objectives to be realised fully. In contrast, the DF sought to take the spirit of Jomtien and add practical strategies to it that provided specific guidelines and programs for countries to implement in relation to the practical goals. The DF outlined the following six goals (UNESCO, 2000a):

(i) to expand and improve comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children;

(ii) to ensure that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete free and compulsory primary education of good quality;

(iii) to ensure that the learning needs of all young people and adults are met through equitable access to appropriate learning and life skills programmes;

(iv) to achieve a 50 per cent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults;

(v) to eliminate gender disparities in primary and secondary education by 2005, and achieve gender equality in education by 2015, with a focus on ensuring girls' full and equal access to and achievement in basic education of good quality; and

(vi) to improve all aspects of the quality of education and ensure excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.

These goals generally reiterate those laid out in the original framework of 1990, but reflect some of the lessons learned during the intervening decade. The goals were purposely left to be general in order to allow for countries to set their own National EFA plans and targets relative to the differing national contexts and states of readiness (UNESCO, 2000). Despite this generality, it is noteworthy that *education quality* is explicitly included among the goals in the DF. Moreover, and importantly, based on the six original goals, the lessons learned, and the global situation that prevailed at the turn of the century, the DF identified twelve further goals which all the governments, organizations, agencies and associations represented at the WEF pledged to strive to achieve (UNESCO, 2000), crafting strategies to suit their own national conditions:

(i) to mobilize strong national and international political commitment for education for all, develop national action plans and enhance significantly investment in basic education;

(ii) to promote EFA policies within a sustainable and well-integrated sector framework clearly linked to poverty elimination and development strategies;

(iii) to ensure the engagement and participation of civil society in the formulation, implementation and monitoring of strategies for educational development;

(iv) to develop responsive, participatory and accountable systems of educational governance and management;

(v) to meet the needs of education systems affected by conflict, national calamities and instability and conduct educational programmes in ways that promote mutual understanding, peace and tolerance, and help to prevent violence and conflict;

(vi) to implement integrated strategies for gender equality in education which recognize the need for changes in attitudes, values and practices;

(vii) to implement as a matter of urgency education programmes and actions to combat the HIV/AIDS pandemic;

(viii) *to create safe, healthy, inclusive and equitably resourced educational environments conducive to excellence in learning with clearly defined levels of achievement for all;*

(ix) to enhance the status, morale and professionalism of teachers;

(x) *to harness new information and communication technologies to help achieve EFA goals;*

(xi) *to systematically monitor progress towards EFA goals and strategies at the national, regional and international levels; and*

(xii) to build on existing mechanisms to accelerate progress towards education for all.

Despite their comprehensiveness, these objectives still remain quite general and lack clear practical directives as to how they can be achieved. However, they provide somewhat more direction for improving the provision of early childhood education than the ambitious objectives of ten years earlier. From the DF, strategies (iii), (iv), (viii), (x) and (xi) (italicised for emphasis) are particularly relevant to this thesis. As stated in Chapter 1, the thesis seeks specifically to harness the potential of new ICT tools to assist education planners in their decision making to achieve improvements in education quality from the local level upwards. Moreover, the approach adopted encourages the participation of multiple stakeholder groups in the formulation, implementation and monitoring of strategies for educational development by developing responsive, participatory and accountable systems of educational governance and management. Hence, the thrust is to create practical tools and techniques that allow progress toward achieving the EFA and DF goals and objectives. Central to the goals of both summits, although perhaps not given the attention it deserved, is the concept of education quality.

Education quality, described above in the DF as excellence in learning and mentioned explicitly in the sixth EFA goal, deserves further attention relative to the focus of this thesis and to education provision in general. This concept emerged as a priority from the WEF and in the DF as experts acknowledged in 2000 that many of the EFA efforts had focused on the quantity of education provided at the expense of the quality of education received by students (UNESCO, 2000b). Clearly, simply putting students in schools and counting this an achievement, when the actual learning outcomes, for a variety of reasons, are very low or in many cases non-existent, is only at best a partial margin of progress. In this regard, the importance of education quality was also recently highlighted in the 2005 EFA Global Monitoring Report. This report, *Education for All: The Quality Imperative*, affirms the importance of education quality on personal incomes, economic growth, non-cognitive skills, and behavioural change (UNESCO, 2005a). However, the report also recognizes the difficulty in defining and measuring education quality.

This important concept and the conceptual underpinnings of its definition and subsequent measurement are discussed in the following section.

2.2 Education quality

It is clear that providing a high quality early childhood education is a key foundation for building responsible and informed societies. Given that education comprises, among other things, a set of processes and outcomes that are defined qualitatively, merely increasing the number of children attending schools achieves very little in real terms if no qualitative improvements in education processes and outcomes are achieved (UNESCO, 2004). The lack of understanding or consensus on a definition of education quality has contributed to the neglect of this concept in national education policies, as it is much easier to establish quantitative goals and targets for the education sector without having to tackle the harder task of evaluating and monitoring changes in a concept that is inherently difficult to define and measure (UNESCO, 2004). Hence, establishing a framework for measuring education quality is critical relative to the task of monitoring progress towards achieving the EFA goals noted earlier.

Although it is difficult to define unambiguously and precisely education quality, the concept can be best understood by examining the word 'quality' and then placing commonly understood meanings of quality in the context of education. The word quality has two basic and related meanings. First, it can be used to mean "a general excellence of standard or level" (Oxford, 2005). In this sense, something is attributed the condition of 'quality' if it possesses the characteristics required to achieve a level or standard of general excellence, i.e. a product of quality, a person of quality, a process of quality. In the context of education, neither the characteristics that are required to achieve quality nor their levels are universally agreed upon. Hence, what constitutes a quality education in one country may be unacceptable in another when treated objectively and evaluated by the same criteria (Peters, 2002).

The second definition of quality provides greater assistance in that it refers to "a distinctive attribute or characteristic possessed by something" (Oxford, 2005). This definition suggests that it possible to determine certain observable/measurable characteristics that must be present in order for something to be considered of good quality. For example, in the context of education there must be a given ratio of students to teachers for each grade taught, there must be more teachers with approved qualifications than without, the ratio of students per computer must be at a certain level within each school, the number of hours of instruction spent on core subjects per unit of time must be at least equal to a set number and so on, for the education provided to be considered of good quality. With this definition, it is possible to establish universally agreed criteria and attributes and then set standards on these, from first principles, and apply these relative to local conditions. Hence, education quality can be monitored and assessed systematically according to a set of locally relevant criteria and attributes. Putting both definitions together allows the components of a quality education to be established (the second) and then observations may be evaluated and contrasted as to their level of quality both in absolute and relative terms (the first), once acceptable standards are agreed upon and achieved.

The first definition alone cannot easily be used to evaluate education quality, as it is impossible to capture the concept of quality properly without establishing the attributes and criteria that are required for the evaluation. Similarly, the second definition cannot be used alone, as the levels or standards with which to evaluate education must first be determined and then agreed upon as being locally relevant to a given context. Further, once both definitions are combined, it is essential to incorporate the perspectives of multiple stakeholders not only in terms of criteria with which to measure quality but also in terms of the levels of each criterion. Without these multiple perspectives it is not possible to set operational parameters that have local relevance and local meaning for defining a quality education.

Measuring quality from the multiple perspectives that exist in the education sector poses a major challenge. An important task is therefore to create a framework that permits groups and individuals representing different stakeholder groups (parents, teachers, school administrators, planners, bureaucrats etc.) to participate in evaluating the education system they are involved with. Although levels of knowledge of the education system are likely to vary substantially between stakeholders, it is beneficial if each stakeholder group understands the internal processes within the system as well as its inputs and outputs (Scheerens, 2005). Numerous studies (Bradford, 1991; Reynold, Teddlie, Creemers, Scheerens, & Townsend, 2000; Flowerdew

& Pearce, 2001; Hall & Peters, 2002) have also begun to recognize that the local geographic context of schools play an important role in determining process oriented factors as well as the potential outcomes (i.e. longer-term objectives) of education delivery. It is therefore useful, in order to evaluate education quality, to develop or adopt a model that divides education quality along a number of complementary dimensions. Such a model has been proposed by Scheerens (2000) and adapted successively by Peters (2002), Leahy (2005) and Engler (2007). This model includes five interrelated dimensions, namely inputs, processes, outputs, outcomes and context, as shown with some examples of the many possible indicators that can be used to assess each component (Figure 2.1).

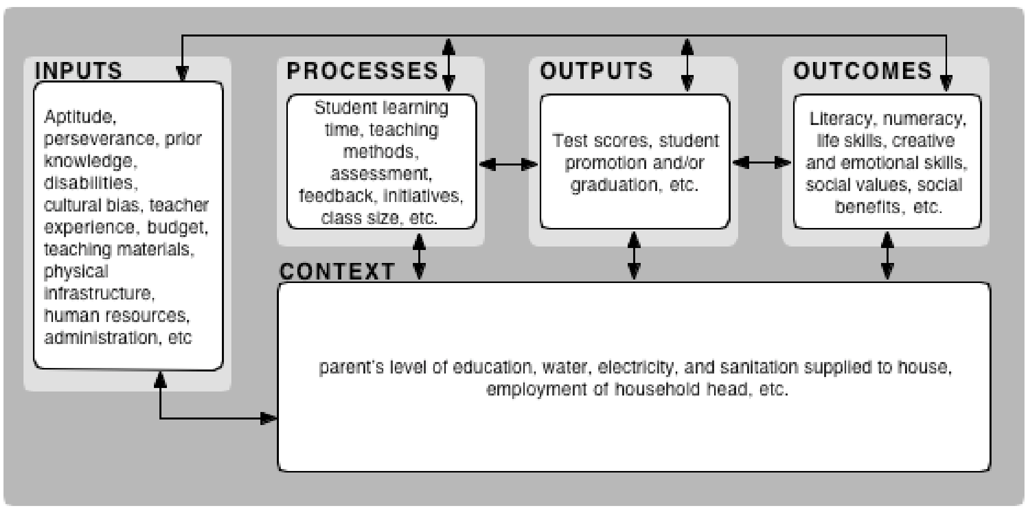


Figure 2.1 – Five-component Conceptual Model of the Education Process (Adapted from Leahy, 2005).

The adapted Scheerens model recognizes that, despite the traditional emphasis on outputs and outcomes, measurement of inputs and processes are also required to analyze education quality (Bertrand, 2003). At the same time, the model is general enough that it can be used at national, regional and local levels of analysis, thus incorporating the spatial dimension of education systems as well as education quality. With this conceptualization, education quality assessment can be focused on a specific geographic level of analysis or include multiple levels, and, depending on the perspective taken, a different set of applicable indicators can be considered at each level or across all levels. The model is also general enough that it can be used by a wide variety of stakeholder groups, all of which have different aims and goals in education system delivery and education reform to meet the goals of EFA, DF and national education policies.

It is important to acknowledge the multi-disciplinary aspect of educational research and to identify explicitly three perspectives through which to understand educational reform, namely the educational, economic, and political perspectives (Ridell, 1999). The educational perspective is primarily concerned with school effectiveness in delivering education, school improvement and the teacher-student interface. The economic perspective is focused on the use of production functions (in order to predict efficient combinations of educational inputs) as well as responsiveness of institutions and individuals on external incentives. Whereas the educational and the economic perspectives have reasonably well defined foci, the political perspective is more ambiguous. In the political perspective, efficiency and effectiveness are measured in different ways depending on the goals and objectives of each criterion both objectively and according to the subjective interpretation of these criteria by all of the current stakeholder groups. The difficulty of accommodating these criteria is further exacerbated by the micro/macro duality of all three, which must also be considered when selecting relevant indicators to represent and assess education quality.

Although the three perspectives are unique, it is possible for a single person to take on multiple perspectives at once. For example, an education planner working for a regional branch of a ministry of education is primarily concerned with the education perspective, but must also bear in mind the economic and political implications of the planning decisions taken. The role of the education planner, as compared to others that influence education systems, is particularly complex for this reason. Education planners are those who are primarily responsible for improving education delivery (both in quantity and in quality) and, as such, they are given special attention throughout this thesis.

As the managers of the education system, education planners must balance a broad range of priorities, each stemming from the three perspectives outlined above. The role of the education planner is then to balance all of the competing priorities in the day-to-day management of the education system. These priorities include, amongst others, educational effectiveness (i.e. the real impact of that the education system has on pupils), cost effectiveness (i.e. ensuring there are adequate resources for all components of the system), equity (i.e. the provision of a certain standard of education to all groups in society, including minorities and marginalized groups), and politics (i.e. education policy and government spending are influenced by political factors and vice versa). Thus, the answers to the same questions posed by education planners may vary depending on which perspective is prioritized. For example, education planners need to routinely answer questions such as the following, all with potentially different answers:

i. Which schools are the most/least effective?
ii. Which students are not receiving sufficient teaching hours?
iii. What schools (or school programs) should be expanded/closed?
iv. What economic sectors are children being best prepared for?

In order to arrive at the answers, education planners (and other education researchers) can draw upon the five-component model presented in Figure 2.1. While each component identified in Figure 2.1 represents part of the education system, it is important to note that each can be of greater, lesser or equal importance depending on the perspective taken. Each perspective presents different sets of agendas and priorities, and as such, they must be represented by appropriate and sufficient indicators (Ridell, 1999). These indicators should be chosen to reflect the relative importance of each component of the education system shown in the model. Conversely, the indicators chosen must also be appropriate and sufficient to represent adequately each of the five model components relative to the different perspectives considered.

The choice of indicators from each of the different components must therefore be careful and deliberate. The choices should, for example, take into account that education itself is mostly an individual process and that education quality is best measured on a student-by-student basis, while changes are usually made at the system or sub-system level. This can often be complicated by the lack of available indicators to measure some of the aspects of education which are of interest, as is discussed at the onset of chapter 3. The choice of indicators must also take into account that the context component is fundamentally different from the other four components of education quality. Context includes the many factors external to school-based education delivery that influences the outputs and outcomes of education inputs and processes. These factors include, among others, the socioeconomic status of the learner's family background, education level attained by the parents, the health and nutrition of students during the school year, and the policy and administrative framework of the education system as this also includes factors that affect the learning environment, but that remain fundamentally outside of it (Peters, 2002).

The vast majority of these factors cannot be directly affected by Government *education* policy as they are multi-faceted and diffuse, while being highly variable geographically. Further, as shown in Figure 2.1, contextual factors have reciprocal, self-reinforcing influences on the other, more linearly and more education system-based and structured input, process, output, outcome components. Their influence serves both to affect the other components, while also being affected by them. The success of improvements in the quality of education provided to students will, in time, be reflected in greater opportunity for accumulation of wealth and improvements in family-based quality of life and reduction in poverty. This in turn will improve, for example, the home-based learning environment, which will allow children to improve their performance at school and rather than the cycle producing a downward spiral of fortunes for lower socio-economic families and the societies they live in, an upward spiral will have an opportunity to evolve.

In creating the conditions necessary for an upward improvement spiral, the socio-economic and cultural characteristics of students and their families must be considered. Families with limited incomes are less likely to be able to afford to send their children to school, to nourish them properly, and to afford required school supplies. Further, parents with a limited or no education themselves are likely to be less capable in encouraging and guiding children with their schoolwork. Hence, contextual factors that describe conditions in the ex-school environment that students live in are essential considerations in evaluating the overall quality of education.

Since, the context of schools is highly variable over space, context is an inherently geographical variable, whereas the input, process and output components tend to be associated with fixed points in space, namely the locations of schools. In this sense, school points and small area contextual data represent the two basic spatial characteristics of the education system. As, noted earlier, all of the five components in Figure 2.1 are temporally variable. Thus, in order to comprehend the performance of an education system relative to its quality, the dynamics of both space and time must be considered.

2.3 Education as a spatio-temporal process

The contextual component of an education system is fundamental to understanding its spatial dimension. In effect, the context relates to the set of facts or circumstances that *surround* a location or an event (Merriam-Webster, 2004). Here, the word *surround* implies 'encircle' and 'envelop' as in the characteristics of the environments surrounding each school and each student's home as well as the learning environments within the schools and the homes of students. The spatial extent of the contextual factors can be proximate (such as the immediate area surrounding an individual school, or the characteristics of the dwelling unit a student lives in) or be distal (such as the characteristics of all households and houses within a district or all schools within a particular local school administrative unit). As the geographic scale of spatial and numerical aggregation decreases (or covers an increasingly large area), so does the variability of individual observations. However, when the individual observations are summed into a composite index the individual variability is lost in the aggregated effect expressed by the index value(s). Clearly, with large areas such as nations or even sub-national regions, localized variation is effectively lost and the only comparisons that can be drawn are comparisons that relate to the level at which the data are summarized (i.e. in this example between nations or between large regions within nations, such as provinces or states in the North American context).

The structure and policies of government ministries of education (both national and sub-national, branch offices) similarly affect the spatial variability of educational contexts (Ochoa & Bonifaz, 2002). These contexts generally determine the distribution of financial resources and the power structures that determine educational management practices. At the same time, the prevailing social, cultural and economic conditions affect individuals in a direct fashion at local levels. The spatial dimension is fundamentally important for planning school districts (Maxfield, 1972), studying variation in school performance (Fotheringham et al., 2001), understanding school and social segregation (Byrne & Rogers, 1996), and assessing education quality (Leahy, 2005; Peters and Hall 2004; Peters, 2002), among other aspects of education provision.

In addition to the spatial dimension, education takes place over a certain time period such as the academic year and/or the time an individual spends in the education system during their life course. From the viewpoint of a student and in relation to the education system itself, this period typically extends from a child's initial enrolment in the system until they graduate, either from one level of the system to another or completely from the system. From a systemic viewpoint, this period is continuous, but in reality it is typically broken into discrete planning cycles that may relate to semesters or terms, the annual school year, or longer-term horizons, such as the time frames evident within the objectives of the WCEFA and DF conferences. From all perspectives, there are a variety of impacts, goals, plans, and experiences that occur during whatever cycle is under observation within the guise of the education system being examined. There are also impacts, goals, plans and experiences that are highly localized (within a household or school), and there are those that are broader in their spatial manifestations. In short, education systems and education delivery span both space and time and, as such, these dimensions must both be explored to assess and monitor holistically levels and changes in education quality. The following section explores both concepts of space and time relative to education delivery.

2.4 Space and time concepts

The concepts of space and time have been subject to debate for centuries, by philosophers and physicists alike. Rather than reviewing the contributions to this debate or seeking to extend it, it is more pragmatic for the purposes of this thesis to define space, time and the relation between these in the context of education system analysis.

2.4.1 Space

In order to understand the relationship between space and the management of education systems, the language, concepts and analytic methods of geography can be used to frame an operational view of an education system and how it functions. At its most basic level, geography is concerned with the two fundamental concepts, namely geographic entities or locations (people, places, objects) and phenomena (events, processes, actions) or processes. Together these transform into the spatial and temporal properties and relations that are characteristic of reality (Couclelis, 2000). With respect to an education system, the two most fundamental geographical or spatial entities are schools and household locations in geographic space. Phenomena could include, among other things, the opening or closing of schools, the enrolment of students, the appointment of new teaching staff or the transmission of students through the system. Education systems can therefore be viewed as a series of spatial entities that interact with one another over space through various processes that occur over both space and time. These entities both initiate and are subject to various phenomena. Viewing, studying and understanding the relationships that exist between them is an essential part of education system analysis and subsequent management through planning processes.

In order to allow an analyst to understand these relationships, it is important that entities and processes are adequately modelled to reflect accurately the various issues that can arise. Given the hierarchical nature of education systems within nested administrative boundaries, the choice of a vector-based, or point, line and area form of spatial data model, to represent geographical entities is the most intuitive to use. With the vector data model, the atomic spatial units are schools and households and these can be represented by points which can be embedded within the boundaries of encompassing administrative and political areas (such as catchments, school boards or districts, and arbitrary political areas such as cities, regions or provinces).

There are many ways in which modelling of spatial phenomena and relationships can be misrepresented or misunderstood. A common such problem is known as the ecological fallacy (Robinson, 1950; Blalock, 1964). This refers to the incorrect assumption that relationships that are found among aggregated data will also be present at lower levels of aggregation, or that relationships found at a given level of aggregation (both numerical and spatial) will also be evident at successively higher levels of disaggregation. While higher-level analyses are always possible with successively disaggregated data, the same is not true for lower level analyses with successively aggregated data. This problem is of particular relevance to the analysis of hierarchical education systems, where there exists significant variability in education quality at the successively lower levels of spatial segregation relative to larger areas. In this case, the ecological fallacy can be avoided by the inclusion of multiple levels of spatial and numerical aggregation which provide the ability to switch quickly and easily between different levels for analytic and management/planning purposes.

Examples of variances at different levels of spatial aggregation can be easily found in the Peruvian education system. Looking at one of the variables shown in Figure 2.1, namely the experience of teachers, Figure 2.2 shows that there exists a certain degree of variance at the department level. Disaggregating these data to the next smallest set of political boundaries (provinces) reveals that there substantially more variability *within* each one of the departments that was evident at the department level itself. In particular, the case of the department of Cusco, which had a high proportion of primary school teachers with titles, it is evident that the provinces within Cusco are in the lowest quintile of teachers with titles for the country. This type of discrepancy between different levels of spatial aggregation could lead to a misrepresentation of facts in systems that do not provide access to all pertinent levels of aggregation. Furthermore, these two sets of figures aptly demonstrate another common problem that can lead to data being misinterpreted or misunderstood, namely the modifiable areal unit problem (MAUP) (Openshaw, 1984).

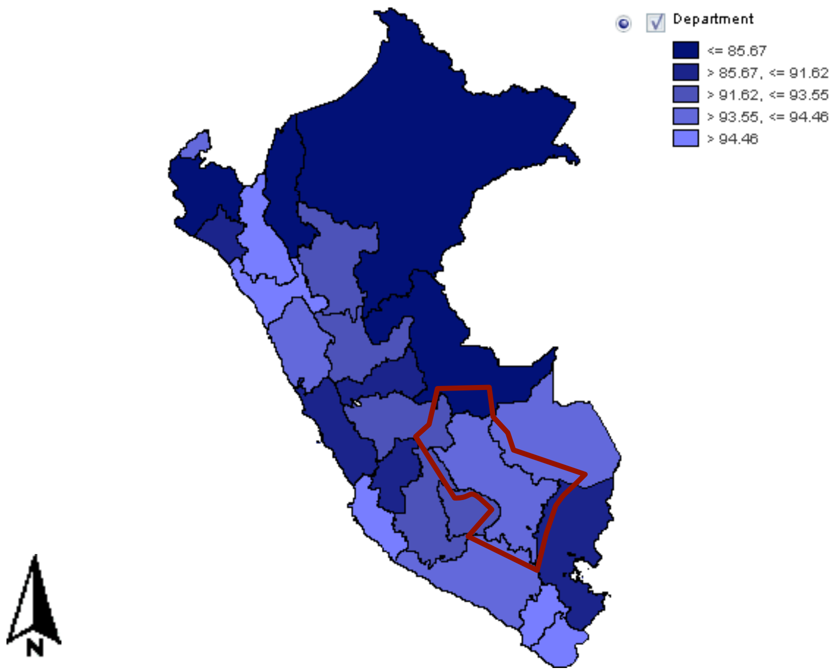


Figure 2.2 – Percentage of Teachers that have Titles by Department (2004)

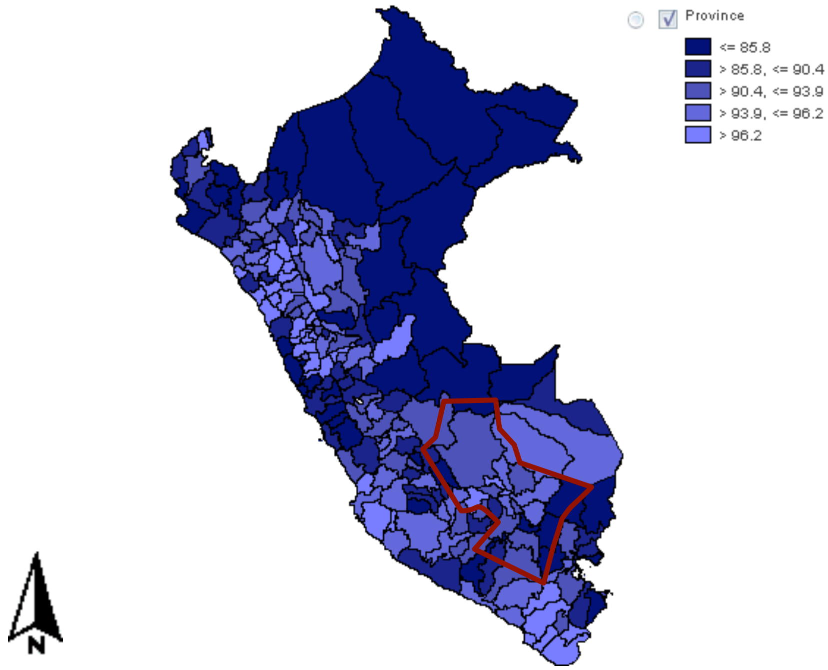


Figure 2.3 – Percentage of Teachers that have Titles by Province (2004)

The MAUP arises from the various ways in which continuous space can be subdivided into discrete entities. Both the scale and the geometric representation of these discrete entities will lead to different interpretations of the underlying data that make up the continuous space under study. In the case of the maps presented above, it can be seen that many of the provinces in the south-east with a low percentage of teachers with titles are aggregated into the same department. A different aggregation might have grouped some of these provinces into neighbouring departments (with high values), which would have eliminated the darker region in the department-level map. Hence, the existing political boundaries for the provinces within each department in Peru affect the way in which the data presented in Figure 2.2 are aggregated and this in turn affects the outcomes of analyses that are performed on the data.

Such issues must, at the very least, be considered in order to ensure that users of a DSS (or other form of information system) are not negatively influenced by the way spatial entities are modelled. Similar caution must be taken when modelling phenomena that describe the education system itself. As phenomena affect and are affected by spatial entities, the latter acquire a past state, a present state and also a future state. For example, the quality of education available at a particular school can improve from one year to the next due to changes that are implemented through the planning process. Since the evaluation, change and re-evaluation processes are inherently temporal, issues surrounding the representation of time in an education information system are important. These are issues are discussed in the following section.

2.4.2 Time

Given the role of time mentioned in the previous section, it is useful to define the concept of time within the context of education systems and their management and planning. One definition, namely "a non-spatial continuum that is measured in terms of events which succeed one another from past through present to future" (Merriam-Webster, 2005), is pertinent in the context of education quality analysis. This definition is suitable, although it explicitly excludes spatial continua, because it introduces the concept of a *continuum* and *sequential events*, which in turn introduce the concept of causality ("the relationship between cause and effect" (Oxford, 2005)).

Causality and planning are intrinsically related. Hopkins (2000) describes how plans work in order to explain their causal efficacy according to five aspects, namely as agendas, policies, visions, designs and strategies. These five aspects of implementation serve as a categorization of plans as well as different types of planning tools. Agendas and policies help to order priorities, while visions, designs and strategies offer different representations of future interdependencies (Hoch, 2007). All five aspects of plans assume a type of outline for what the desired future state of the system being planned should be. In evaluating the effectiveness of a plan, a planner must be able to compare the past with the present in order to discern the causes and effects of actions and of any other contextual factors that have changed. It follows from this understanding that time must be adequately represented in the *process* of planning in order to implement an information system that is directed toward use in planning (such as a DSS), and that can be used to formulate strategies, see them implemented and subsequently evaluate their success.

In the context of education planning, examples can easily be found that demonstrate the importance of incorporating time in any information system directed toward the planning process. Using a similar measure from the examples shown in Figure 2.2 and Figure 2.3, namely the percent of primary teachers that are trained, Figure 2.4 shows that there exists a certain degree of variance from year to year throughout each continent, region, and economic class.

Focusing on South America, the graph shows that the number of trained primary school teachers has fluctuated greatly from year to year between 1998 and 2001, with a decline of 15% in 2001.

The cause and effect implications of this flow two ways. First, the implications of the number of trained teachers on the quality of education received by children could prove to be severe. For example, a decline in trained teachers may potentially be highly correlated with lower performance scores in standard tests and higher desertion rates (student loss) from the school system. Second, if this is true, the causes of the fluctuations in the number of trained teachers are of utmost importance. For example, monetary constraints could be addressed by prioritizing training programs in future budgets, while low enrolment in teacher colleges could be addressed by actively promoting the profession. Regardless of what the cause and the remedial action taken, the evaluation of the intervention will necessarily involve doing temporal comparison of the related indicators. As this simple example demonstrates, the ability to track changes over time is a key part of the planning process and should therefore be adequately represented in any education information system aimed at decision support.

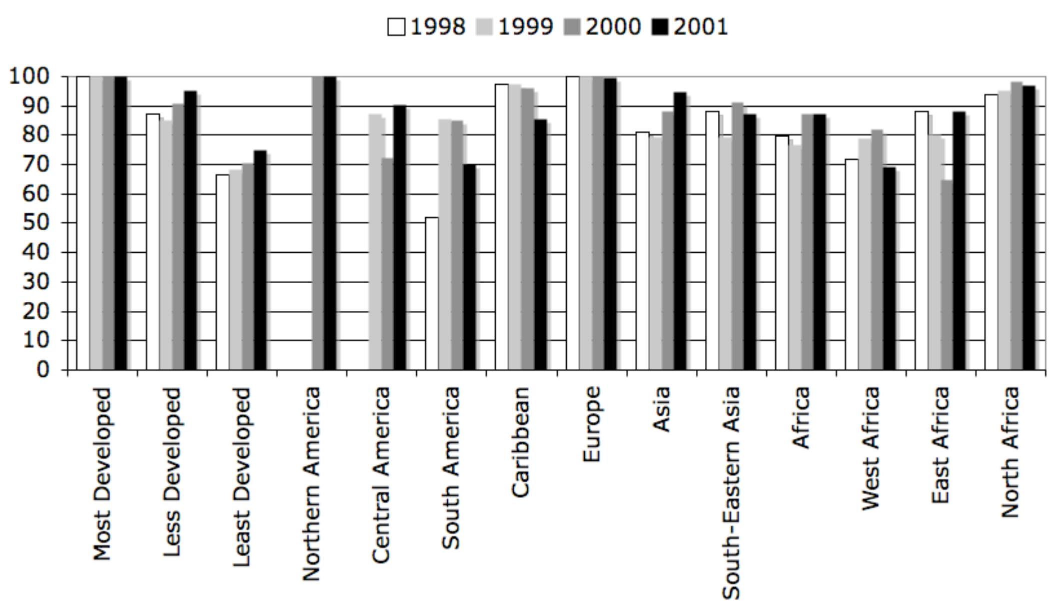


Figure 2.4 – Percent of Primary Teachers that are Trained, 1998 – 2001

Weighted averages calculated using available national-level teacher population vs. percent trained data.

Note: data unavailable for Australia/New Zealand. (World Bank, 2002)

In order to capture a timeline, and to accommodate the planning process, an education information system must employ some form temporality in its database design and implementation. There are three main definitions of time for temporal databases, namely transaction time (TT) which is the time when an event is recorded in the database, valid time (VT) which is the time when a fact is true in the real (or modelled) world, and user time (UT) which is anything else that is time-related (Kokkotos, Ionnidis, Panayiotopous, & Spyropoulos, 1995, Combi & Montanari, 2001). An additional definition of time is event time (ET) (Chakravarthy & Kim, 1994), also known as decision time (DT), which is the time in the real (or modelled) world when a fact is generated or terminated (Combi & Montanari, 2001). The previous three forms of time are typically modelled as intervals, while the latter is modelled as an instant in time.

Regardless of which definition of time is chosen as the basis for a database design within information systems in general, the result is that time is typically separated into discrete periods. This is contrary to the perception of time as continuous and puts the onus on the designer and the user(s) (decision-maker(s)) to understand and take into account the implications of the time-model that is deployed. The utilization of multiple definitions of time in the same system can facilitate this task by providing sufficient flexibility in the input and output of information for the user to reconstruct real-world events. For example, consider the following scenario where a child falls sick (A) and a week later is too sick to attend school (B). The end of the school year arrives two weeks later and since the child has not recovered nor written exams, his/her teacher decides that he/she must repeat the year (C). The child's illness prevents him/her from starting at the beginning of the next session (D), but he/she does register after a month (E). This timeline is depicted in Figure 2.5 through a time sequence and the accompanying transaction time and decision time.

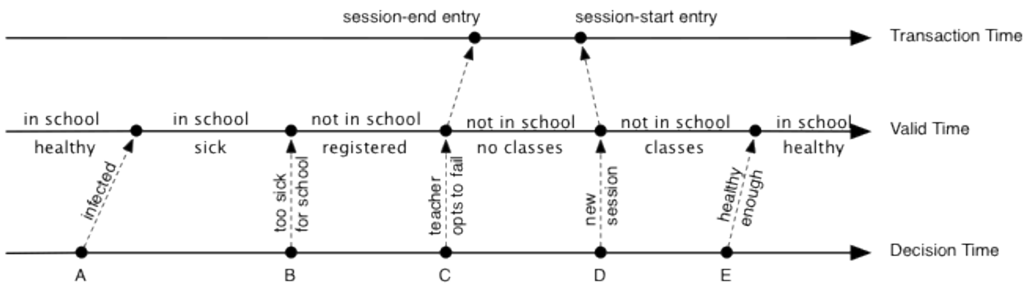


Figure 2.5 – Different Conceptualizations of Time in an Education System

In Figure 2.5 the school only records the student's situation at the end of one school session and at the beginning of the next (TT). A DSS that only includes these data from this system would give a planner the impression that the student has abandoned school and has not returned. Only at the end of the following session would the system show that the student has returned. For this example, it could be argued that only one of VT or DT is needed in order to reconstruct reality with an accurate representation of events, since the points of one can easily be inferred from the state of the other. For example, the DT point A (where a student is infected) necessarily implies that the student will be sick at a slightly later time; conversely, given the VT point where a student falls ill necessarily implies that at some previous point the student was infected. Given this relationship between VT and DT, a system that contains all of the points of either VT or DT would allow a decision maker to reconstruct real-world events fairly accurately, but would do so by adding complexity to the database design requirements and to the analysis of temporal processes.

Given that this thesis deals with the analysis of processes that are both spatial *and* temporal, the adoption of a simplified temporal dimension is necessary in order to limit the burden that the user must bear when reconstructing and analyzing the real world from the data found in the information system. The following section proposes an approach that allows for space and time to be joined, for the benefit of the analyst, under the unifying concept of space-time.

2.4.3 Space-time

In order to assess the performance of an education system in terms of the quality of the education received by students as well as monitor changes in quality over time, it is important to develop a conceptual framework that unites both space *and* time and expands the temporal definition provided in the previous section. An operational model must be derived from this framework that allows planners and decision makers to assess education quality through various lenses and across a temporal as well as a spatial dimension.

Such a framework must necessarily treat space and time as inseparable. In the world of physics, time and space were viewed as independent dimensions until Einstein's work in the early 20th century. Einstein directly challenged the assertion of independence by suggesting that "space and time adjust themselves in an exactly compensating manner so that observations of the speed of light yield the same result, regardless of the observer's velocity" (Greene, 2004). Although scarcely observable in our everyday experience, experimental results have shown that this claim is true and that, in fact, space and time can never be separated. These results imply the existence of a four-dimensional manifold known as space-time, within which all events occur. It is through this concept of space-time that education system analysis is framed throughout this thesis.

To explain a complex concept very simply, space-time is a container that encompasses every place and every moment in time simultaneously. Figure 2.6 shows two students running into a school in a region of space-time. In this simplified diagram, the students and the school exist in a two dimensional world, where their height extends along the y axis and their width along the x axis. The third dimension, t, is the temporal component of the space-time manifold. Under the space-time paradigm, in Figure 2.6 the students' initial location (all outside), their final locations (two inside and one outside) and every state in between exist simultaneously. Viewing and studying the region of space-time surrounding the student can yield an understanding of the event.

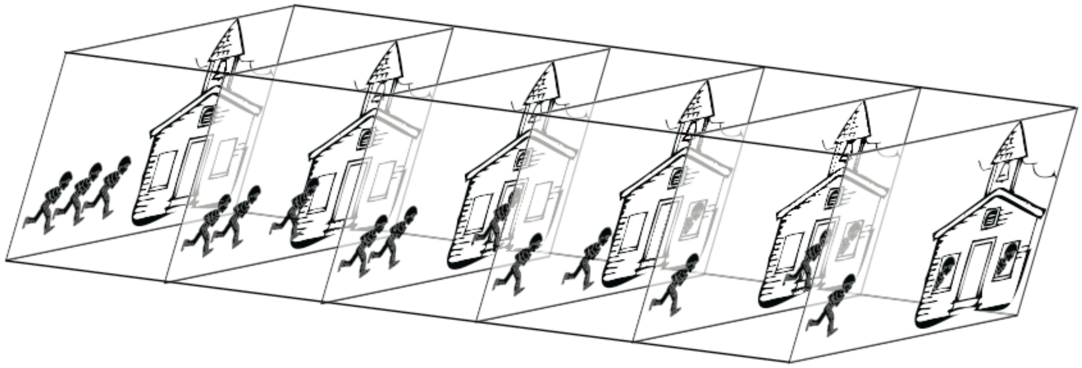


Figure 2.6 – Students entering a school in a region of space-time
(Spaces between panels are for visual clarity only and are not meant to suggest that time is discrete)

Inferences from this can be made about the events that take place within a given region of space-time by observing slices of the region or by projecting the region along one of its multiple axes. For example, a vertical slice of the region of space-time in Figure 2.6 captures the number of students currently in school at one instant in time. To know all of the events that have happened along a given timeline, a slice that is perpendicular to the line of interest can be taken. To know if a student has entered or exited the school, a projection along the time axis between a starting and an ending time can be observed. This idea of slicing and projecting a region of space-time in different ways in order to answer questions about events that have taken place is explored further in Chapter 4.

One further aspect of the dissection of the space-time manifold must be highlighted here. There exists a fundamental difference in what dissections can be made of the space-time manifold in the in the physical world, the one Einstein refers to, and of a conceptual 3D manifold that can exist inside an information system that models the real world. While in the real world all dimensions are continuous and can be dissected at infinitesimally small intervals, a modelled world is subject to the limitations imposed by the underlying data. In particular, dimensions tend to be made up of discrete elements (e.g. years or smaller periods in the case of time) that restrict the slicing of the manifold along these predefined intervals of limited granularity. In other words, the space-time manifold can be manipulated in a wide variety of ways, but these ways must necessarily correspond to the subsections defined by the underlying data being used to model reality. However, despite this difference, it remains true that studying and manipulating the pertinent regions of space-time can lead to an understanding of real-world events.

The existence of space-time, under Einstein's terms of general relativity, fundamentally changed the understanding of the relationship between space and time and of the physics of the universe. Although the revised laws of physics do not affect how an individual makes decisions, the concept of space-time plays a pivotal role in decisions that involve the interaction of these two dimensions. In this context, the interdependence of space and time is as important in decision-making as it is in understanding the principles of theoretical physics. Events are distributed over space-time and, as such, decision-makers need access to the entire region of space-time that encapsulates the system they are interested in observing and modifying. Such access can only come about from the provision of appropriate tools that facilitate the manipulation of data that describe fundamental characteristics of the system in question, including both space and time.

Some of the technological foundations needed for the type of tools described above have been under development since the 1990s. In particular, there has been a significant amount of work in the development of both spatial (see, for example, Guenther and Buchmann, 1990; Güting, 1994; Medeiros & Pires, 1994) and spatio-temporal databases (for example, Abraham & Roddick, 1999; Peuquet, 2001; Mokbel, Ghanem, Aref, 2003). However, in order for database technologies to be useful to decision makers, they need to be embedded within a decision support tool that exploits their capabilities. As this thesis demonstrates, the data structures, access techniques, and query languages developed for both purely spatial and hybrid spatio-temporal databases can be applied in the implementation of decision support tools with analysis functions that facilitate the type of data manipulation required by decision makers dealing with both space and time.

Parallel to the development of database technologies has been the increasing presence of DSS with spatial capabilities. However, this class of decision support tools, commonly known as spatial decision support systems (SDSS), has certain limitations. These limitations are discussed in the following section.

2.4.4 Spatial Decision Support Systems

A considerable amount of work has gone into building decision support tools in multiple and diverse domains. In particular, SDSS have evolved from the more conventional and more widely used DSS to assess the spatial dimension of decision problems and thereby provide support for decision makers when spatial data and spatial problems are of central importance in the decision context. Although SDSS have traditionally focused only on the spatial dimension of the space-time manifold discussed in the previous section, it is possible to design a SDSS in such a way to incorporate consideration not just of space but the extended concept of space-time. This section examines the limitations of traditional SDSS with their focus only on space.

A SDSS can be formally defined as "an interactive, computer-based system designed to support a user or group of users in achieving a higher effectiveness of decision making while solving a semi-structured spatial decision problem" (Jankowski, 1999). There exist many spatial decision domains other than education system analysis. Problems in which geography plays a critical role include evacuation planning (e.g. de Silva & Eglese, 2000; Billa, Shatti, Mahmud, & Ghazali, 2006), land suitability assessment (e.g. Bojórez-Tapia, Diaz-Mondragón, & Ezcurra, 2001; Marinoni, 2006), housing relocation planning (e.g. Johnson, 2001; Sriraj, Minor, & Thakuriah, 2006), health resource allocation (e.g. Hall & Bowerman, 1996; Ibaugh & Rushton, 2003) and many others. Given the importance of spatial knowledge for these types of decision problems, it is clear that traditional DSS (as defined in Chapter 1) lack the ability to manage and utilise spatial databases as well as the spatial models needed for the evaluation of the spatial component of decision problems. In the past, this left decision makers with only the spatial data management tools and analytic techniques available in spatial information technologies (SIT) to explore the spatial aspects of a planning problem.

Although contemporary SIT such as GIS provide a comprehensive array of tools for spatial data management and processing, they still do not provide, out-of-the-box, after more than twenty years of development the type of models that are required to assist with decision-making (Densham, 1991). SIT are therefore in their native form insufficient for decision-making without substantial customisation through macro programming or scripting. Moreover, DSS typically do not consider the management or analysis of spatial data. This serves to highlight the fact that spatial planning problems are inherently difficult to solve.

These challenges lead to the formation of SDSS by integrating SIT within a conventional DSS data and model base. Considerable growth in research, development and applications of SDSS has occurred since the early 1990s (e.g. Carver, Evans, Kingston, & Turton. 2001; Rinner, 2003; Sharifi & Herweijner, 2003). More recently, the Web has facilitated and promoted the use of such technologies for decision tasks. Some authors go as far as suggesting that the Web has become the primary medium for delivering computerized decision support (Rinner, 2003; Bharagave, Power & Sun, 2007). Whether or not this claim is accurate, it is certain that there exists a common motivation to provide greater public access to the data and tools that support decision-making and that the use of the Web facilitates this (Rinner, 2003). Further advantages of using the Web as a delivery medium, specifically in developing countries, are described in Section 3.4.

The two most widely implemented approaches to Web-based SDSS are those that are data-driven and those that are model-driven (Bharagave, Power & Sun, 2007). Model-driven DSS use formal representations of decision models and provide analytical support to decision makers. One of the many examples from this category are the GIS multi-criteria analysis tools that began to appear on the Web as early as 1999 (Malczewski, 1999) and which are comprehensively reviewed in Malczewski (2006). On the other hand, data-driven DSS help users organize, retrieve, and analyze large volumes of relevant data. There also exist numerous examples of tools in this category (c.f. Carver, 2001; Rinner, 2003). The chapters that follow present a further example of such a tool, EduCal, developed at the University of Waterloo, which is in essence a hybrid of the model-driven and data driven approaches.

Despite all of these developments in the realm of SDSS, there have been only very few attempts, with the exception of the work of Bédard et al. (2001, 2003, 2005), to provide a basis for spatio-temporal decision support. As of yet, there is a relative absence of system designs, associated databases, and analytic tools that give a decision-maker the ability to perform evaluations of decision issues that involve considerations of both space and time, within the same design. For example, in the education domain existing Web portals for accessing education-related data, such as UNESCO's Open Education Management Information System (UNESCO, 2006) or the Peruvian Government's MINEDU's ESCALE portal (http://escale.minedu.gob.pe) lack either the spatial or the temporal dimensions. The original version of the EduCal tool (http://gaia.uwaterloo.ca/EduCal) also lacked the temporal component, which has taken prominence as the primary aspect of the tool in its second version. However, recent research has been published that suggests that researchers are now beginning to recognize the importance of time as well as space in many decision problems (Jankowski, Robischon, Turhill, Nyerges, Ramsey, 2006).

The most difficult aspect of the challenges posed by space-time is to design and program a DSS that provides access to both space and time *simultaneously*. In the context of the education component model presented in Figure 2.1, a planner working in a ministry of education must be able to determine easily which schools within a specific area have experienced a decrease in the quality of education over time, while at the same time being able to examine the spatial variability in education quality between schools as well as spatial variations in contextual factors that may have exerted an influence on education quality. These requirements call for a well-designed spatio-temporal DSS (st-DSS) that incorporates space and time in a way that is appropriate for the decision-problem being addressed, as well as being intuitive for decision-makers to use.

The representation of spatial and temporal phenomena in information systems has evolved with the development of various technologies. With increased computing power and improved computational methods, the development of st-DSS is closer to becoming a practical reality than ever before (Jankowski, Robischon, Turhill, Nyerges, & Ramsey, 2006). The growth of generic software, in particular spatial databases, has allowed spatial phenomena and their relationships to be managed in conjunction with other factors and in Chapter 4 a model is presented, using current Open Source software (OSS) technologies, to implement a st-DSS that meets the requirements relative to the problem of education quality assessment that is addressed in this thesis. It is possible to overcome the technical barriers to assessing space-time problems in terms of software production. However, in order to render software useful it is also necessary to design an appropriate operational data model that can subsequently be used to create the st-DSS. Such a model is presented in the following chapter.

2.5 Summary

This chapter has laid the conceptual foundations for the remainder of this thesis. These foundations started with the international context surrounding the Education for All initiatives, specifically the results of the WCEFA of 1990 and the DF of 2000. This context lead to a discussion on education quality and subsequently to the view of education as both a spatial and temporal process. It was proposed in this chapter that education systems be modelled through the unified concept of space-time, in which space and time are combined within a single manifold. The challenges presented by this type of representation of education systems within information systems were then explained. Finally, this chapter concluded by outlining the need for a new category of tools, st-DSS, that provide access to the space-time manifold to education planners and others interested in the assessment of education quality.

Chapter 3
Operational Data Model

This chapter introduces a data model that forms the basis for a st-DSS to support education planners in the assessment of education quality across both space and time. The model is based on data available for the Peruvian education system, but is generic and extensible enough to be applicable to most education systems around the world, with only changes in details rather than in basic design. In order to give a concrete basis for the data model, the available Peruvian education data sources, their formats and the problems in using the data encountered are first explained. This is followed by a description of the data model itself and the type of space-time operations that are made possible with it.

3.1 Available Data

Many education systems throughout the world routinely collect, and make accessible, data regarding the public education system (for example, DiNIECE, 2007; Statistics Canada, 2007; MINEDU Bolvia, 2007; MINEDU Perú, 2007). Since, the case study used in this thesis is the assessment of education quality in Peru, the type, structure, and quality of data collected routinely by the government of Peru and that eventually form the building blocks of the data model developed for this thesis are discussed in the following sections.

3.1.1 Sources

The primary source for all education-related data in Peru is the Peruvian Ministry of Education (MINEDU). At the lowest level, the MINEDU collects data about each individual student through student registration forms known as the *Ficha Unica del Alumno* (Unique Student Record). These forms are completed upon enrolment through a short interview with parents and contain information specific to the student such as name, age, and gender as well as basic information about the student's home and family characteristics. Information about student performance and behaviour is kept, in an informal way, in each teacher's *cuaderno de control* (control notebook). The MINEDU distributes these notebooks with the intention that they become a standardised resource to record information about student behaviour and performance for later recall when preparing report cards. However, these potentially valuable qualitative and quantitative student-level data are not reported back to the MINEDU and so they are only exploited within the classroom for the management of students by teachers. More formally, student performance is recorded in the report cards that are sent home at the end of each school term. The information in report cards is obtained through ongoing data collection for each student through in-class teacher designed testing and standardised Ministry-mandated testing. However, no national mechanism exists for quality control of these report cards, and thus the standard of grading and the quality of the information contained in them varies significantly from region to region, as well as from school to school within regions.

Differences in the method and system of grading are also common. The information in both the *cuaderno de control* and in the end-of-term report cards could potentially, if reported properly, be an invaluable resource for assessing pedagogical practice, the impact of policy reform, or more generally the efficiency and effectiveness of the educational processes that take place at the classroom level.

Data pertaining to personnel and infrastructure are also recorded at the school level through mandated surveys. Teachers and school administrators, at the behest of the MINEDU, collect this information for the *Censo Escolar* (School Census). The survey is carried out in full every two years and with less detail in-between. The census comprises the most thorough educational dataset available in Peru. It contains relatively good quality general information such as the *Código Modular* (School Identifier), the school's address and contact information of the school administrator as well as infrastructure information regarding the number of classrooms, libraries, computers, desks, etc. Information regarding the characteristics of the school personnel such as their training and the type of employment (i.e. full/part-time, contract/salaried) is also recorded. A summary of the collected data sets, their formats and their availability is presented in Table 3.1.

Contextual socio-economic data are primarily collected by the Peruvian National Statistics and Information Institute (*Instituto Nacional de Estadística e Informacion,* INEI) in the form of the *National Census of Population and Housing*. The census, which was last enumerated in full in 1993 and with a representative sample enumerated in 2005, contains data for each household. The 2005 census was invalidated due to poor enumeration control and repeated in 2007 under the administration of a new director of the INEI (BBC, 2007). These contextual data are complemented by data found in the *Ficha Unica del Alumno* (Unique Student Record), which records whether or not the student works outside of the classroom. The third and final source of contextual data includes the height and nutrition surveys done in 1993, 1999 and 2005. However, these surveys only include small samples of students from each school included in the sample frame.

Dataset	Content	Source	Date(s)	Format	Availability
Basic school statistics	Basic descriptive information and summary education data for schools	MINEDU	1998-2006	Digital DBase files	Publicly accessible on-line from the MINEDU website
Annual school census	Detailed data about students, staff and infrastructure for schools	MINEDU	1998-2006	Digital DBase files	Publicly accessible on-line from the MINEDU website
Student registration forms	Descriptive information about student and family characteristics	MINEDU	n/a	Primarily stored on the original paper registration forms, rarely digitized	Effectively not available, except through direct context of local offices
Education budgets	Information about school budgets, their sources, teacher salaries, etc.	MINEDU	n/a	Not recorded in any formal way	No known source of this data is readily available
Student height and nutrition surveys	Sampled data of nutritional status and height of students in grades 4 and 6	MINEDU	1993, 1999, 2005	Digital DBase files	Publicly accessible on-line from the MINEDU website
Student grades and test scores	Data on the performance of students	MINEDU	n/a	Not recorded in any formal way	No known source of this data is readily available
National standardized test results	Sampled standardized tests of student performance	MINEDU	1996, 1998, 2001	Digital DBase file	Publicly accessible on-line from the MINEDU website
National population and housing census	Contextual data about the socio-economic and demographic characteristics of the Peruvian population	INEI	1993, 2006	Digital text and Digital DBase files	Microdata are not available, although data aggregated at the city block level or higher may be purchased from INEI

Table 3.1 – Availability of education quality and context data collected in Peru
(Updated from Leahy, 2005)

As shown in Table 3.1, the various surveys and censuses are collected at different times and time intervals. The lack of coordination when surveys are conducted and which students/schools are sampled complicates the comparison of different datasets over the time dimension of the space-time manifold. Furthermore, even the school census and the basic school statistics, both of which are collected regularly, vary greatly in both their content and format. This makes using the data for monitoring change over time extremely complicated, as some variables are stored in different locations, recorded with different variable names or with different measurements, or not recorded at all for some years. These variances between the surveys of different years cause there to be gaps and fluctuations in the space-time manifold that cannot be corrected. Some of the specific issues encountered when trying to compare data from different years are explained in Section 3.1.3.

Despite the availability of all of the datasets listed in Table 3.1, there continues to exist a shortage of relevant data in Peru that would likely be very useful to education planners. For example, both the available MINEDU data and the INEI data are entirely quantitative in nature and omit important qualitative information such as teaching methods employed in classrooms or information regarding the performance of teachers and/or students (such as annual performance reviews conducted by school administrators or comments on report cards completed by teachers). There is not even any standardized national testing of students that is done on a systematic basis other than the *Crecer*, and even the results of this are generally unavailable for analysis of the performance of students in schools where it is conducted.

However, the quantitative data collected are in general sufficient to calculate a large number of indicators relevant to the assessment of education quality. In fact, an indicator library of 724 individual indicators, each pertaining to the various components of the education model presented in Figure 2.1, was built based on the work of Leahy (2005). It should be noted that this number is partially inflated by breaking down student-related indicators by both grade and sex and by other internal characteristics in the case of other data. The categorization of these indicators yielded 302 input, 229 process, 20 output, 2 outcome and 171 contextual variables with which to assess education quality relative to the derived Scheerens education quality model. As multiple time instances for successive enumerations on the same indicator sources are added into the database, its magnitude can grown very rapidly (i.e. as successive indicators from various input sources are added for new enumerations). Moreover, the database has a certain fluidity to it, as new schools are added and schools that have ceased operations are not updated in new rounds of data enumeration.

As previously stated, the calculated indicators pertain to schools (in the case of education data) that are represented spatially as points and households (in the case of contextual data) which are represented spatially by points aggregated into polygons for city blocks in urban areas and points for *centros poblados* (population centres) in rural areas throughout Peru. With the aid of geographically enabled software such as GIS it is possible to visualize these features on digital maps and perform spatial analyses relative to measured assessments of education quality calculated from selected indicators in the library of indicators.

Spatial data pertaining to the education system are available from the same two sources as the indicator data (the MINEDU and the INEI), although linkages between the two sets of data are poor. In large part, these linkages are difficult to make because of the severe shortage of metadata documentation about their content, how they were created, and how to use them. Furthermore, there are certain limitations imposed by the file formats of all of the spatial data. All of the data are all found in the Shapefile format, a non-topology based format introduced by the Environmental Systems Research Institute (ESRI). Although this format has become the de facto standard in the spatial database world, the format stores no topological information increasing the likelihood of problems with the data (such as the existence of overlaps and gaps between polygons) and causing potential difficulties when performing efficient analyses that draw upon topological relationships between spatial entities such as connected to, beside or adjacent to, within, or surrounded by.

The datasets in Table 3.2 also show all of the levels of spatial aggregation that are relevant to the Peruvian education system. The education system hierarchy explained in the following section defines these various levels of spatial aggregation in more detail, as this is a fundamental aspect of the space-time data model discussed later and its incorporation into an st-DSS for the assessment of education quality.

Dataset(s)	Spatial Feature Type	Source	Date	Format	Availability
Regions, Provinces, Districts	Polygon	INEI	2000	Shapefile	Publicly accessible on-line from GIS data warehouses (e.g. http://www.geocommunity.com)
Regional Management of the MINEDU	Polygon	MINEDU	2000	Shapefile	Generated manually by the NGO Alternativa from district datasets and tabular data
Local Management Units of the MINEDU	Polygon	MINEDU	2000	Shapefile	Generated manually by the NGO Alternativa from district datasets and tabular data
City blocks	Polygon	INEI	1993-2001	Shapefile	Available for purchase from INEI
Population Centres	Point	INEI	1993	Shapefile	Available for purchase from INEI
Population Centres	Point	MINEDU	2000	Shapefile	Available upon personal request from MINEDU staff
Schools	Point	MINEDU	2000	Shapefile	Available upon personal request from MINEDU staff

Table 3.2 – Availability of spatial data layers in Peru
(Adapted from Leahy, 2005)

3.1.2 Hierarchy

As with most countries throughout the world, the Peruvian education system is comprised of a series of nested administrative and political levels that share hierarchical geographic boundaries. Hence, each component of the hierarchy indicates a location of authority from which decisions can be made regarding the part of the education system that fall within that component's jurisdiction. In practice, these boundaries encompass physical locations (e.g. schools and city blocks) or administrative areas (e.g. districts, provinces, departments and, ultimately, the national boundary). The structure of this hierarchy affects the education system in many important ways. For example, the number of levels in the hierarchy defines the number of levels within which decisions and responsibilities can be delegated to different components of the system.

There is an important distinction between the levels in the hierarchy that comprise the formally defined components of the education system and those that are outside of an entity that relates specifically to education management (for example, national and regional government areas). In the case of Peru (and other education systems) there is a close relationship between the education system and other aspects of government-based responsibility and administration. Hence, the type of education decisions made and their motivations at any given level of the hierarchy are likely also to pertain to a political boundary (i.e. a province or multiple districts). Exogenous levels of governing authority may affect education quality indirectly through the allocation of public funding for community-based development efforts, whereas the levels of authority pertaining to the education system itself directly affect the quality of education received by students at schools.

In the case of Peru, the education system shown in Figure 3.1 is comprised of four levels of authority, namely schools, local management units (*Unidades de Gestión Local, UGEL*), regional management units (*Dirección Regional Educaiva, DRE*) and the national ministry. Directly related to these four levels are households, population centres and districts, provinces and departments. As previously mentioned, at the most basic level the MINEDU collects data from the schools while the INEI collects data from households. At the atomic spatial level of the education system (i.e. individual schools) principals, vice-principals and other school staff manage operations and school funds along with input from parent's groups. However, schools are subject to guidelines, regulations and budget constraints mandated by the levels above in the education hierarchy.

Local management units or UGELs are responsible for the formulation of local educational projects, local educational plans and budgets, local school zoning, and the allocation of public investment in school infrastructure and maintenance (UNESCO, 2005b). UGELs are comprised of one or more districts. In the urban areas, districts can be loosely tied to different neighbourhoods while in rural areas districts are somewhat arbitrarily demarcated areas comprising groups of communities. The composition of districts creates a slightly different hierarchy between urban and rural areas. In rural areas, both the MINEDU and the INEI utilize the concept of *population centres* in order to aggregate their microdata. However, in the case of the MINEDU, there exists a near one-to-one mapping between rural schools and population centres while the INEI uses population centres to ensure a minimum population size for every measurement point. Although the population centres defined by each organization can be considered to be the same level of spatial aggregation, there exists no direct correlation between the two. In effect, the two sets of population centre points create two alternative roll-up paths for aggregating data up the spatial hierarchy depending on which source data are being considered. The next highest level of spatial aggregation is the DRE, which shares with its constituent UGELs responsibilities for curriculum development, evaluation and accreditation of schools and teachers (UNESCO, 2005b).

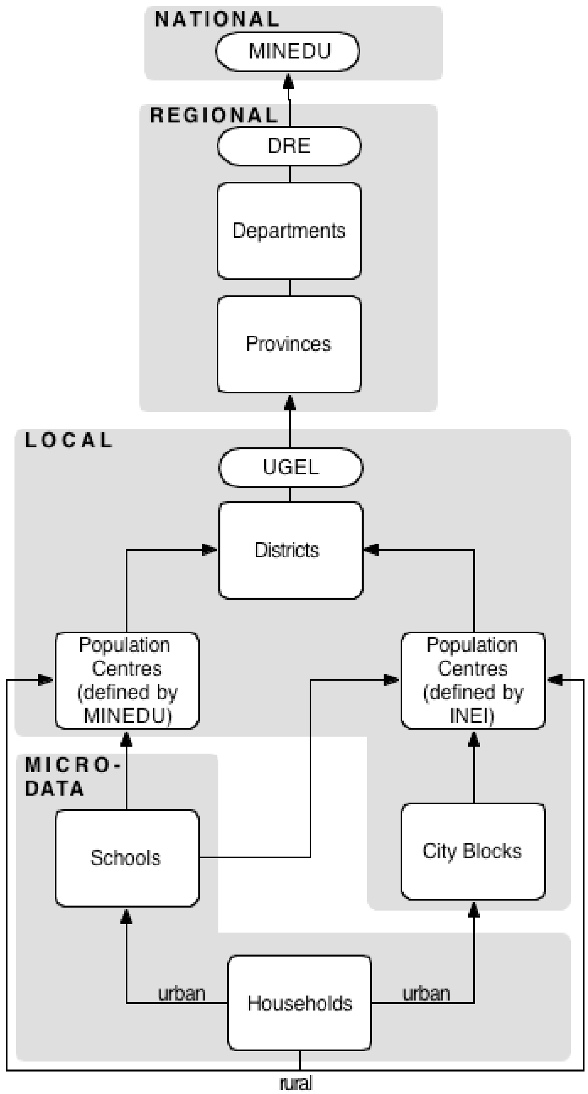


Figure 3.1 – Peruvian Education System Hierarchy

DREs are exclusively concerned with the formulation of regional education projects, education plans and budgets, regional school mapping and country-wide planning of educational supply and programmes of public investment in school infrastructure and maintenance (UNESCO, 2005b). Except for the case of Lima, each DRE is associated with a department (the equivalent of a Canadian province/territory). The department of Lima is the only exception as it has two DREs, one for Metropolitan Lima and another for the rest of the department. Due to the size and variation in the characteristics that exist within a department, it is important also to include provinces (the equivalent of Canadian counties) as part of the education system hierarchy in order to have a level of aggregation with which to detect intra-department variations in education quality. As with the case of the UGELs and DREs, as noted above, regional governments share responsibilities with the MINEDU in issues such as curriculum design, evaluation and accreditation.

While the DREs address problems within their respective regions, the MINEDU is responsible for managing issues that arise from inter-regional variations as well as overall national level monitoring and evaluation of the Peruvian education system. The specific responsibilities of the Ministry are many and varied including developing policies to compensate for or even discriminate in favour of disadvantaged areas or populations, responsibility for the maintenance of information on teachers working in the education system, and the scale of teacher and administrator salaries by level of education and by step (as determined throughout collective negotiations on the financing through the annual public budget), among other things. The MINEDU is also in charge of coordinating projects, such as the introduction of new technologies at schools, education policy and curriculum development, implementing aspects of the EFA national plan, and the Peruvian adult literacy programme (UNESCO, 2005b). In order to accomplish these objectives, the MINEDU has become a large and complex organization, comprised of various secretariats in Lima as well as regional and local administration offices scattered throughout the country.

Two units are primarily responsible for the assessment of education quality. The Education Statistics Unit (*Unidad de Estadística Educativa, ESU*) and the Unit of Education Quality Measurement (*Unidad de Medición de la Calidad Educativa, UMC*) operate under the Office of Strategic Planning and Measurement of Education Quality (*Oficina de Planificación Estrategica y Medición de Calidad Educativa*). All of the data collected by the MINEDU make their way to the ESU and are utilized by the UMC to write and publish reports for the public, to develop plans of action for improving education quality, and to help establish a system of evaluation for the quality of education. Despite the establishment of this unit with the specific mandate of managing and making accessible education data, there continues to be a wide range of problems in making productive use of the available data by local and regional levels of the MINEDU and by the general public, as well as by the UMC itself. Although the data are accessible through the Education Quality Statistics (*Estadística de la Calidadad Educativa, ESCALE)* Web portal (http://escale.minedu.gob.pe), many challenges associated with the use of these data persist. The most significant of these problems are highlighted in the following section.

3.1.3 Data Problems

The challenges associated with Peruvian education data focus primarily on issues related to the completeness and accuracy of the data that are collected. These problems are compounded by the use of unconventional data formats to store the data, an apparent lack of coordination between those involved in the collection of the data themselves, and a lack of proper documentation (incomplete, inconsistent and/or simply non-existent information). Whether these issues are the result of a lack of resources or simply poor information management practices, the challenges presented in this section demonstrate that the education datasets available in Peru are extremely difficult to use in order to produce a large database of indicators to assess education quality within a st-DSS.

As previously noted, the primary problems with the available education data relate to incompleteness and inaccuracy. A lack of communication from the MINEDU back to the schools from which the data are initially collected on how the data can be used and how they can benefit individual schools has caused school administrators to perceive that there is little practical benefit from completing the surveys carefully and completely and submitting them the local and regional offices of the MINEDU on time. The low priority placed on data collection by school staff has caused some schools, particularly private ones, to complete only partially the surveys and, in some cases, not to complete them at all. In other instances, school administrators have been known to misreport values intentionally (e.g. by inflating the number of students in order to ensure sufficient funding).

From a more technical standpoint, the chosen file formats in which the data are found impede their utilization without a considerable amount of skilled labour to transform data into a consistent and usable format. This is particularly true of the school censuses, the most comprehensive of the available datasets, which are fully downloadable from the ESCALE web portal in a series of files in dBase format (.dbf) with an accompanying data dictionary (in portable document format (.pdf)). Although the data format is fully explained in the data dictionary, this information cannot be extracted digitally in a form that facilitates the interpretation of the database files. Each file corresponds to a section of the school census and is comprised of a series of rows, each with five key fields, with the answer to each census question. The lack of any additional metadata accompanying the database files means that any user wishing to use these data must create a series of import scripts to read each record (row) and, by reading each of the five key fields, interpret the values found. It is possible that the MINEDU has, somewhere within their internal systems, a set of scripts that facilitate the construction of a fully normalized relational database for interpretation and querying of the datasets, but no such scripts or filters are available from the ESCALE portal or by inquiring at the MINEDU.

Regardless of whether this is the case or not, a more conventional and practical way of providing these data would be to provide properly integrated and normalized data tables that have exactly one row or tuple per school and the various education indicators stored as columns or fields with an appropriate set of primary and foreign keys. With a properly formatted construction, the constituent files could then be easily imported into any relational database package for querying, filtering, manipulation and subsequent analysis. Moreover, this would facilitate easy linkage to properly coded spatial databases that are associated with school points and population centres and small areas.

In order to deal with these deficiencies an extensive set of filters that had to be created to integrate and clean the 2002 student census data. However, these same filters proved to of little use when importing the 2004 census data. The 2004 data also utilize the same five key fields and are in dBase format, but the values in these fields have changed in many instances and there exists no documentation as to why these changes occurred or how they should be interpreted. Furthermore, not all of the same information is collected from year to year and, in some cases, what appears as equivalent information is, in reality, something slightly different from earlier enumerations. Due to a complete lack of documentation on the changes that occur from year to year in these surveys, a very time consuming process of manually comparing every question in the school censuses had to be carried out in order to arrive at two relational and normalized datasets (2002 and 2004) that can be directly compared to one another over time.

The comparison of the datasets produced by two or more different surveys both at the same point in time as well as over time poses an additional set of challenges. The lack of regularity and coordination in the collection of the various surveys means that the information collected about a school in one survey does not necessarily correspond with the information about the same school in another survey conducted 3-6 years earlier. Furthermore, in the school census, schools are coded with a unique school ID and an additional annex code (to indicate a different location for the same school if the school operates from two places), whereas none of the other data sets contain this annex code. In order to compare data from schools with the annex code and schools that do not have annexes, an annex code of 0 had to be assumed for all schools that fall into this category. This reduces the potential size of the dataset by excluding the latter schools from consideration. Again, there exists no documentation explaining what the correct procedure should be to include or exclude schools with annexes from consideration.

Similar problems exist with the available spatial data in Peru. The multiple datasets from different sources utilize coordinates based on different geographic projections and alternative identification codes for linking the spatial data to attribute/indicator datasets. The lack of any metadata, in particular on how the spatial data were created and how they should be used forces the introduction of potential errors when linking school points to the indicator data as guesswork must often be used. There also does not appear to be any information available regarding the state of maintenance of most of the datasets. Although schools may be opened, closed or moved in any year, there exists no record of how these possibilities are handled. This problem is exacerbated by the fact that there does not exist any street network data with address ranges coded into the network to support the geocoding of school locations. However, this is certainly not the only issue, as it appears that the spatial data have never been fully up-to-date in Peru. There are many schools in the indicator tables for which there exist no spatial data, and nearly all schools in several districts in Lima are missing the identifier codes required to link them to the indicator database. This, of course, reduces the completeness of the derived dataset and introduces the possibility of association error in cases where guesswork must be used to link spatial and attribute databases together.

In addition to the problems outlined above with the linkages between the education data and the spatial datasets, the INEI has failed to create any documentation on the changes they made to their own spatial data in preparation for the 2005 population and housing census. This complicates any linkages between the 1993 population census data and the updated geographic features for the census areas enumerated some twelve years later. Although the grouping of city blocks, the location and number of population centres, and shifted district boundaries are reflected in the spatial data, there have been no adjustments relative to the old census data. The only documentation that exists on this matter is an equivalency table provided by the INEI, for which there is no accompanying explanation on how it should be used in order to maintain the integrity of the contextual data over time.

The types of problems outlined in this section serve to highlight the difference between an ideal and carefully planned and implemented data collection and distribution process and the reality of what can be found in a developing country such as Peru. Naturally, decision makers will be able to make better-informed decisions from complete, up-to-date and well-designed data resources than from data that are none of these. Further to these qualities, ideal data should be well documented using an internationally accepted metadata standard so that both the technical staff who organize the data and end users can quickly and easily achieve a full understanding of what the data represent. Adequate metadata form the basis with which the relationship between the data and the real world can be better understood. Finally, ideal data should be in a digital form that allows easy integration into a variety of commonly used data analysis tools (e.g. in normalized tables compatible with a recognized and widely used relational database package).

The amount of effort and the number of technical challenges that had to be overcome in order to approximate an ideal data scenario from the available data cannot be overstated. It is clear from these efforts that the Peruvian government needs to invest a great deal more thought and action in improving its data management practices if it wishes to provide data in a usable form to its own staff and to the public.

Only once the datasets in their current formats were imported, filtered, cleaned, normalized and consolidated into a relational database organized according to the data model described in the following section was possible to do some worthwhile analysis. Without these additional extensive and time consuming steps the point of collecting the data in the first place must reasonably be questioned according to current practices.

3.2 Data Model

In order to produce a useful st-DSS for education planning, it is imperative to organize the data according to an operational model that facilitates the extraction of information about the education system in general and education quality in particular, relative to the general education model discussed in Chapter 2 (Figure 2.1). The following sections introduce such an operational model based on data warehousing concepts and online analytical processing (OLAP). These two aspects of corporate information technology are aimed primarily at enabling the "knowledge worker (executive, manager, analyst) to make better and faster decisions" (Chaudhuri and Dayal, 1997; p. 1).

The problems outlined in the previous section, regarding the conditions in which education data are found in Peru, make it clear that it is highly improbable that valuable information can be extracted from the data unless the source databases are first transformed and consolidated. A data warehouse, defined as an "enterprise-oriented, integrated, non-volatile and read-only collection of data imported from heterogeneous sources and stored at several levels of detail to support decision-making" (Bédard, 2001; p. 56), serves exactly this purpose. The term *enterprise-oriented* suggests that the data warehouse is a complete source for data that encompasses system-wide information within a large enterprise such as the Peruvian Ministry of Education. The term *integrated* suggests that data warehouses are the result of a series of transformations of heterogeneous semantics, constraints, formats and codings into a homogeneous source (Bédard, 2001).

Two other important and related characteristics in the above definition are that data warehouses are *non-volatile* and *read-only*. These terms mean that once data are placed in the data warehouse they are never changed. Whereas other types of systems might keep only the most recent data, it is crucial for temporal analysis and predictions over both space and time to be viable that historic data remain in the warehouse in a non-volatile (non-changeable) state. Hence, access methods must leave the data unchanged, especially in cases in which the source data systems have the potential to overwrite existing data. The final defining characteristic of a data warehouse is the presence of data at *several levels of detail* (otherwise known as *granularity*). It is essential that decision makers be able to get both a global view as well as to be able to drill down into multiple hierarchical levels of finer granularity, such as the various levels of education system organization explained in earlier sections. All together, these characteristics are indispensable when providing a unified view of heterogeneous data sources to feed decision-support tools with maximum accuracy, completeness and currency.

In fact, it is these same characteristics that make a data warehouse-oriented approach ideally suited to support OLAP-based decision support systems. OLAP is based on the representation of the underlying data as a multi-dimensional cube, sometimes known as a hypercube that is discussed in the following section.

3.2.1 Data Cube

The depiction of data as a multidimensional cube is a powerful and intuitive way to conceptualize the type of data exploration that an education planner needs to undertake in order to extract knowledge from the underlying databases. The data cube itself is comprised of a series of smaller cubes at finer levels of granularity that, at the lowest or atomic level, are linked to the smallest spatial objects (schools, population centres and city blocks) and multiple instances of time (years) to form space-time regions within the overall space-time manifold described earlier. By viewing multiple cubes at a given point in time, it is possible to perform a wide range of analyses including the comparison of any combination of indicators, time periods and places. The smallest cubes can also be amalgamated in varying combinations in order to view aggregated values for space-time regions of coarser granularity within the overall data cube. To make this possible, OLAP-based systems are equipped with the operations required to exploit fully the organization of data along the existing dimensions.

In the case of education data, there exist three major dimensions, namely space, time and attributes (indicators). Furthermore, each of these dimensions has multiple dimensionalities of their own. For example, the spatial dimension can be aggregated and disaggregated according to the nested hierarchy of an education system such as that described in Figure 3.1. The relationship between the three dimensions is such that the attributes are considered to be the dependent variable(s), occurring at the intersection of the other dimensions (the independent variables). This organization of data into a cube, such as that depicted in Figure 3.2, aptly captures the education system by representing space-time regions as three-dimensional (3D) objects. Furthermore, the data cube lends itself to dissection and manipulation in a wide variety of ways, all of which lead to the selection of a different region of space-time for exploration of spatial and temporal trends within the context of a functional st-DSS.

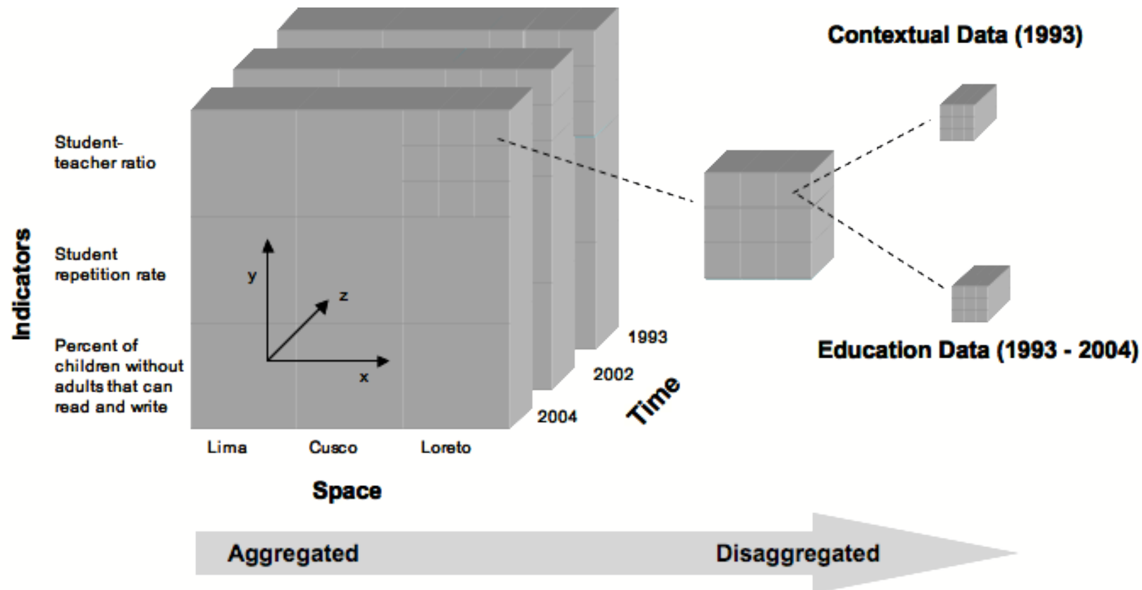


Figure 3.2 – Multidimensional data cube model

Using the dissection of the data cube, various selection actions can be conceptualized as "slicing" along any one of its dimensions. For example, selections by location name slice the cube depicted in Figure 3.2 along the y-axis, selections by attribute slice the cube along the x-axis, and selections involving time slice the cube along the z-axis. The selection of more, or less, aggregated values is slightly different in that the cube is not sliced in any particular direction. Instead, parent cubes are selected in the case of more aggregation and child cubes in the case of disaggregation (Figure 3.2). None of these manipulations of the data cube needs to be performed independently, and, thus, it is possible for a selection to be refined along any number of axes simultaneously.

OLAP is designed specifically to perform these types of operations with a high level of processing efficiency. By design, OLAP systems contain pre-computed aggregations of as many attributes as possible along all of the dimensions. This greatly increases query performance by reducing the number of calculations that must be carried out instantaneously, at the expense of significantly increased storage space requirements. The conservation of storage space is not a consideration in the design of OLAP-oriented systems since supporting rapid *ad hoc* information retrieval is an overriding objective. Using pre-aggregated data, it is possible to attain very fast answers to complex questions (OLAP Council, 1995), which is essential in ensuring that the analyst maintains his/her train of thought and is not distracted by slow response times (Rivest et al., 2001).

Analysts are simultaneously assisted by the multidimensional data structure of OLAP systems, which reflects the cognitive model that database analysts are known to use (Rivest, Bédard, & Marchand, 2001). The specific implementation of the multidimensional structure can vary greatly and be built on top of a wide range of architectures. The end result is always the capability of performing iterative analysis by implementing a standard set of operations through interfaces that eliminate the need for analysts to have an intimate knowledge of complex query languages. These standard OLAP operations are described in the following section.

3.2.2 OLAP Operations

OLAP systems typically provide, through their implementation, access to three main operations, namely *drill-down* (disaggregate values to show more detail with greater variability), *roll-up* (show more aggregated values to show a more global picture with less variability), and *slice* or *drill-across* (show the value of an attribute at the same level of aggregation with constant variability). Performing two slice operations along multiple dimensions simultaneously is sometimes called *dicing*, leading to the expression *slicing and dicing* of the data cube. In some implementations, there also exists a *pivot* operation (swap one dimension for another with constant variability, all other actions being equal), but this operation is not discussed further in this thesis. These operations, combined with some analysis functions, permit access to all the possible views and combinations of the available data, which in turn encourages the emergence and testing of new hypotheses through an iterative process of knowledge discovery (Han, 1997).

An example of how this iterative process of knowledge discovery can take place within a specific implementation of an st-DSS is presented in the following chapter. However, the remainder of this section presents how the OLAP operations mentioned above dissect the data cube presented in Figure 3.2 to provide the building blocks of each iteration. With each operation, the analyst changes the portion of the space-time manifold being examined and, in the process, is exposed to several dimensions at once in order to view their relations to one another.

Perhaps the most basic of the four operations is *slicing*, as it is performed when the analyst wishes to examine the data for a particular attribute. The slice operation leaves the level of aggregation unchanged, but it allows the analyst to change the current portion of the space-time manifold being examined. Slice operations can either select a subset of the current slice of the space-time manifold or change to another slice altogether, but always without changing the level of aggregation. In the example in Figure 3.3 the 3D cube is sliced horizontally in order to sub-select the indicator *student-teacher ratio* from the three indicators shown. When multiple slice operations are executed simultaneously, the operation is termed *dicing*. Figure 3.4 shows an example of the cube diced along both the indicator and the year dimensions simultaneously, providing the analyst with the value of the *student-teacher ratio* for one particular year. One more slice operation, this time along the space dimension, would leave only one cube and would correspond to the value of one indicator, for one year, in one place.

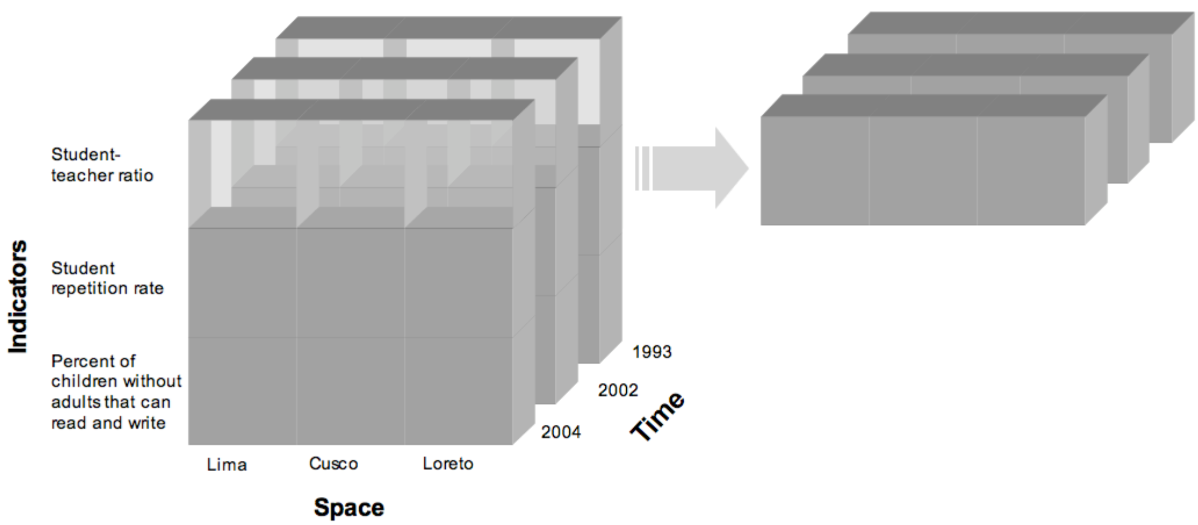


Figure 3.3 – OLAP slice operation

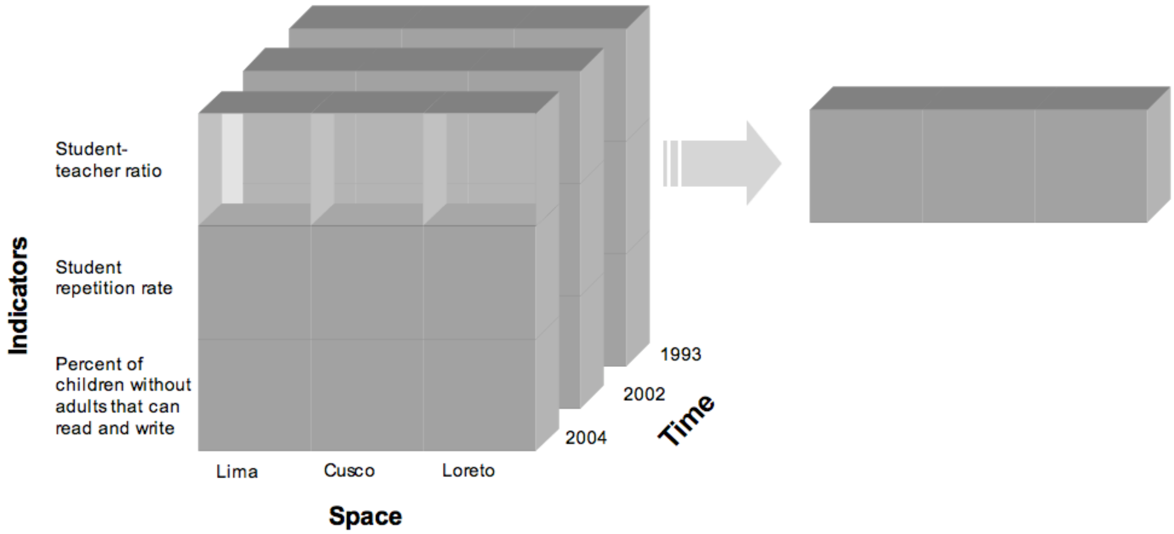


Figure 3.4 – OLAP dice operation

The *drill-down* and *roll-up* operations perform opposite functions to one another. While drilling-down increases the level of detail (granularity), rolling-up decreases it. The hierarchy of the underlying data predetermines the levels of detail accessible through the drill-down and roll-up operators. The analyst can then use these operators to switch between global (aggregated) and local (disaggregated) views of the currently selected dimensions. The example shown in Figure 3.5 shows rolling-up and drilling-down of one level of granularity, first along the space dimension and then along the indicator dimension. In this figure, drilling-down occurs from left to right while rolling-up from right to left. Although Figure 3.5 only shows the change from one level of detail to the next in the hierarchy, it is possible to move from any level of granularity to any other with these operations.

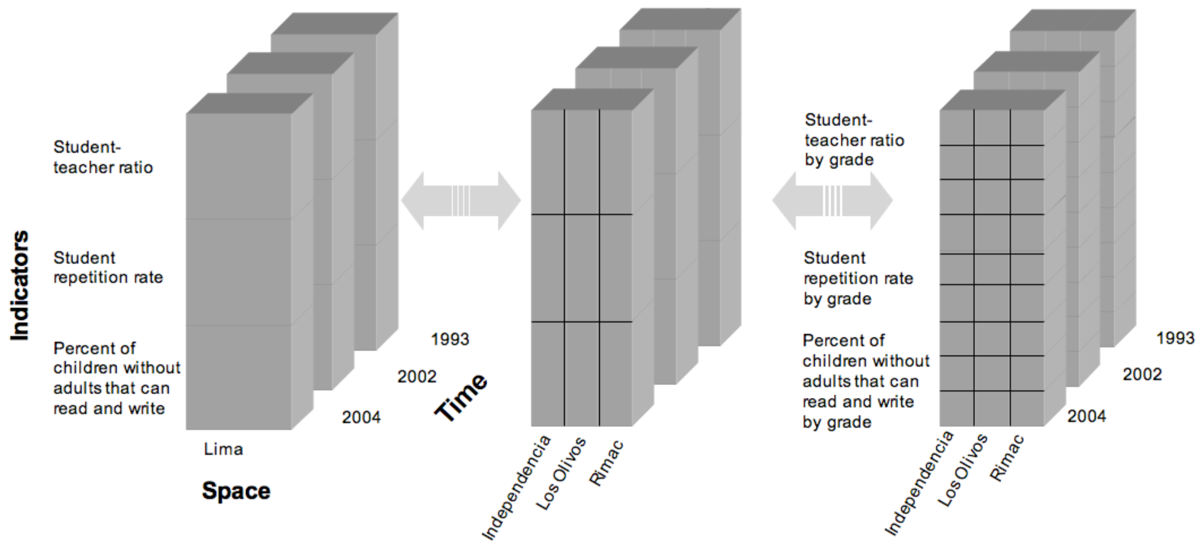


Figure 3.5 – OLAP drill-down and roll-up operations

As can be seen from the examples in Figure 3.3 to 3.5, these operators allow for the selection of different sections of the cube as well as access to all sub-cubes, sometimes called *cuboids* (Thomsen, 1999). By performing these operations in combination with one another, the space-time manifold is at the disposal of the analyst. The use of some additional analytical functions can allow for the combination of variables in ways that allow the analyst to understand the relationships between the data available at the region of space-time under analysis. In this way, the use of OLAP operations and analysis functions enable a series of numerical queries for browsing the multidimensional data of data warehouses. However, although OLAP is recognized as a useful component within current decision support frameworks for numerical and multidimensional problem solving, its potential has yet to be realized in the realm of spatial problem solving.

The spatial dimension of data has some features that distinguish it from other dimensions. In particular, the fundamental difference is that the knowledge within the spatial dimension cannot be extracted through purely numerical methods, and although the spatial dimension has been widely integrated in OLAP systems, this integration has been usually done in a nominal, non-cartographic manner (i.e. using solely place names) (Bédard, 2001). A natural approach to address the particular requirements of the spatial dimension is to combine OLAP with a spatial information technology, such as a GIS. However, the potential of this combination has not been realized due to the difficulty and ambiguity in combining the numerical frameworks of OLAP with the spatial frameworks of GIS (Keenan, 1997).

In fact, Bédard et al. (1997, 2001, 2003, 2005) are among the few who have coupled OLAP and GIS into a DSS. The need for the integration of spatial technologies within OLAP, dubbed by Bédard as Spatial OLAP (SOLAP), is described along with the characteristics of an effective SOLAP system in the following section.

3.3 Spatial Online Analytical Processing

As was alluded to throughout the thesis, to date, there is an inherent need to create appropriate tools in order to gain a better advantage of the spatial *and* temporal dimensions in decision-making. While GIS are the obvious platform for analyzing the spatial dimension, GIS *per se* are not adequate for decision-support applications when used alone, and so alternative solutions must be used (Bédard, Merrett, & Han, 2001). This is due, in part, to the fact that GIS do not provide the multidimensional view of data (and of particular importance, to the temporal dimension) in the way OLAP systems do. However, based upon the OLAP operations outlined in the preceding section, it can be concluded that OLAP offers good support for simultaneous usage of temporal and spatial dimensions, at least when spatial data are treated as another descriptive dimension.

Treating the spatial dimension like any other poses severe limitations because this would not take advantage of the spatial relationships inherent in the data. It is generally understood that there exist three types of spatial dimensions (in the multidimensional sense, not in the geometric sense). Each type considers whether it deals with a geometric spatial reference (such as x,y coordinates), with a semantic spatial reference (such as a place name), or with a combination of the two (such as a street address) (Bédard, Merrett, & Han, 2001). Of these three, only cases where the data are comprised entirely of semantic spatial references can benefit from a DSS based on traditional OLAP.

Although OLAP is effective for decision support in these few cases (see, for example, Caron, 1998), knowledge discovery is hindered as there is no spatial visualization available, practically no spatial analysis, and no map-based exploration (Bédard, 2005). All three aspects are critical in order to address problems that are inherently spatial, such as the topic of this thesis. Spatial planning problems necessitate the extraction of insights from complex spatial data structures and the relationships that exist between spatial objects. Furthermore, this complexity increases when dealing with various levels of spatial aggregation such as the case of hierarchical education systems. The three types of dimensions lead to three types of spatial hierarchies, namely non-geometric, geometric to non-geometric, and fully geometric. Each type of hierarchy is defined by how the spatial dimension type changes as the measure of the dimension is aggregated and disaggregated.

Existing tools such as database management systems (DBMS), GIS and OLAP are capable of processing the different types of data (geometric vs. non-geometric and aggregated vs. detailed), but on their own they continue to be inadequate for spatio-temporal decision support. While OLAP-centric approaches lack the spatial handling functionalities of a GIS, GIS-centric approaches offer limited or no multidimensional data handling functionality. Hence, there is an urgent need to produce an integrated solution capable of handling the aggregated, geometric, and multidimensional data requirements of st-DSS (Bédard, 2005). SOLAP-oriented approaches fill this gap (see Figure 3.6).

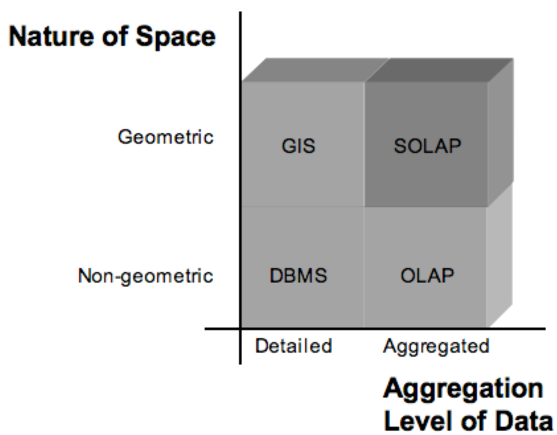


Figure 3.6 – Position of SOLAP with regards to the nature and the aggregation level of data
(Source: Bédard, 2005)

More formally, SOLAP can be defined as "a visual platform built especially to support rapid and easy spatiotemporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays" (Bédard 1997; p94). Implied in this definition is the need to integrate fully both GIS and OLAP concepts, functions and operations into a single system. In more concrete terms, Bédard (2001) identifies several specific desirable characteristics of SOLAP systems. These characteristics are divided into two broad categories, namely visualization of data and exploration of data. A few key characteristics from both categories are noted below.

From the visualization of data category:

- Cartographic and non-cartographic displays

  Access to all dimensions should be available from a common interface. This involves the visualization of both the geometric component through a map interface and conventional displays such as tables and charts.

- Visualization of context data

  SOLAP tools should make it possible to display background maps in order to help the user find their specified information.

- Modification of the graphical semiology

  A SOLAP user should be able to change the graphical semiology (i.e. the symbols used to represent geographical entities) according to their needs. Restrictive semiology prevents users from representing entities in the ways they are used to.

From the exploration of data category:

- Availability of the data exploration operations in all display types

  For example, the drill-down and roll-up operations should be available from all displays (be they tabular or cartographic).

- Availability of spatio-temporal analysis functions (metric and topological)

  In order to analyze spatio-temporal data in a SOLAP system, the appropriate analysis functions must be made available. Traditional spatial operators, such as *adjacent to* and *within a given distance of*, are a minimum necessity and analysis functions, such as buffering and network analysis, should be included where applicable.

- Filtering on the dimension members

  A SOLAP tool should allow the user to restrict the data being visualized to the data of interest so that the user is not overwhelmed by the amount of information given in one display.

With these characteristics in mind, it is possible to take the available education data outlined at the beginning of this chapter and design a st-DSS that fulfills the spatio-temporal analysis needs of education planners.

3.4 Providing SOLAP for Education Planning

SOLAP-based systems with the characteristics outlined in the preceding section have the potential to give planners and researchers the capacity to explore and synthesize the available education data in an efficient and productive matter. More importantly, the data cube model and SOLAP technologies are consistent with the EFA and DF models and objectives set out in Chapter 2. In this sense, the operational data model presented in this chapter can lead towards software that improves the ability of planners and researchers working towards these EFA and DF goals.

While software can be a valuable resource in the process of knowledge discovery, the successful implementation and the use of an st-DSS for education planning will be highly dependent on the degree of accessibility of the appropriate technologies and availability of human resources with the necessary expertise for their use. This is true in any context, but especially true in developing countries where computing resources can be scarce. This is further exacerbated in those instances where there exists an absence of habits surrounding the use of computer systems to satisfy the needs of planners and decision-makers. Unfortunately, processing and time saving advantages available through ICTs are quite often underutilized (UNESCO, 2004b).

The growth of the information society (Castells, 2000) worldwide, propelled by the adoption of ICTs, is leading to the creation of the essential infrastructure and capacity required for a greater utilization of tools to analyze and interpret information as effectively as possible. The limitations that continue to exist in the usage of such technologies are likely to continue to disappear as the information society continues to expand and take hold. Given this trend, a tool such as an st-DSS for the assessment of education quality satisfies the growing demand of an information rich and technologically advanced society.

The potential of such tools is more likely to be realized if the barriers of accessibility are minimized. As mentioned above, these barriers to accessibility could be a lack of physical computing resources, a lack of human personnel with sufficient expertise for their use, or, in many cases, a combination of both of these factors. Software solutions implemented should therefore be designed in such a way that they minimize their need for specialized personnel and hardware. One way to achieve this is to use the Web as a platform and the Internet as the distribution channel.

The use of the Web as a platform and the Internet as a distribution channel for decision support systems reduces the barriers to accessibility in a variety of ways. Firstly, the growth of the Internet (estimated at 433% in Latin America between 2000 and 2007 (Internet World Stats, 2007)) translates into a software distribution channel with a lower cost and further reach than ever before. This makes it possible to make the software available at a relatively low cost. Secondly, Web-based software does not impose additional hardware of software requirements for any computer on the network, as all of these are certain to be equipped with a Web browser. Lastly, the ubiquity of the Internet has lead to people from all sectors of society to be familiar with the types of interfaces and modes of user interaction provided by Web-based software, thereby reducing the need for specialized personnel to operate the software.

As the growth of the Internet continues and the information society becomes more prevalent, the benefits of the user of the Web as a platform for software such as an st-DSS will continue to grow. Therefore, while SOLAP reduces the conceptual barriers to analyzing education data, the Internet and the Web reduce those of accessibility. The following chapter explains in detail how to combine various open source technologies in order to bring about an implementation of a Web-based st-DSS with the SOLAP characteristics previously outlined.

3.5 Summary

This chapter has outlined an operational data model that can improve the ability of planners and researchers to analyze and interpret information effectively. This model takes into account the discussion regarding the EFA and DF model and objectives laid out in Chapter 2, as well as the spatio-temporal nature of education systems. In order to provide a concrete description of the chosen model, this chapter used Peru as a case study.

At the onset of the chapter, the available education data in Peru were described along the difficulties and limitations imposed by this data. It was identified that any information system using the available data would greatly benefit from more complete and organized datasets, but that the available data were sufficient for the creation of a large educational data warehouse. Following this discussion, the hierarchical structure of the Peruvian education system was explained along with the responsibilities at each tier. This education system hierarchy, along with space, time, and indicators, comprise the dimensions that form the proposed data cube model. The standard set of On-Line Analytical Processing (OLAP) operations (roll-up, drill-down, and slice) were then introduced as a companion to this data cube model. A deficiency in the OLAP approach was then identified for spatio-temporal problem domains such as and a new category of software known as Spatial OLAP (SOLAP) was introduced as a solution. Finally, the chapter concluded that, in the growing information society, the Internet and the Web are an ideal medium to disseminate such tools to education planners.

Chapter 4
High Level Design of EduCal Software

The data model and processing tools outlined in the previous chapter need be implemented in conjunction with an adequate software architecture, application model, application logic flow, and database schema. Without a careful design, there would be no chance for a complex education-based st-DSS to be of productive use under the conditions found in a country such as Peru. To meet this need, the EduCal software tool was designed primarily for education planners in Peru. The overall software architecture explained in this chapter is applied in Chapter 5 for the assessment of childhood education Peru, however its design and implementation is sufficiently robust for it to be used with success in most countries.

4.1 Overview of EduCal

EduCal is an acronym for *educación de calidad* (quality education) in Spanish and *education calculations* in English. The software is a Web-based st-DSS that was designed specifically to aid the Peruvian MINEDU and non-government organizations (NGO) operating in Peru in assessing education quality, with the ultimate goal of developing education improvement plans to increase the likelihood of meeting or surpassing the most recent goals of the Dakar Framework for Action (or DF) and EFA. EduCal fits well within the structure of the relatively new decentralized structure of the Peruvian education system and resultant hierarchical organizational model discussed in Chapter 3. Specifically, the tool provides an intuitive map-based interface to a SOLAP-enabled data warehouse containing the data routinely collected by the MINEDU and described in the previous chapter. EduCal adds 'value' to the data by providing access to a flexible scenario-based form of analysis through using the Internet and the World Wide Web (Web) as its access and delivery mechanisms. The use of the Web as the access medium for EduCal draws on the principles of public participatory GIS in the sense that it facilitates access to any interested party with Internet access to spatial data and analysis tools (Hall, Alperin, & Kerrigan-Léon, 2007).

In order to ensure sustained access to EduCal, the tool was developed entirely with Open Source computing technologies. The choice to develop EduCal under the auspices of the Open

Source Software (OSS) movement also is in line with the evident global trend towards the creation of an OSS agenda (The Register, 2005a). The use of OSS for EduCal also guarantees that it the software can continue to have a place in the MINEDU in light of the bill approved in 2005 in Peru that "prohibits any public institution from *buying* systems that tie users into any particular type of software or that limits 'information autonomy'" (The Register, 2005b).

There exists a considerable number of OSS technologies for spatial applications (a summary of these can be found at http://www.MapTools.org). This subset of OSS, commonly known as Open Source Geospatial (OSG) applications, has grown significantly in recent years both in terms of the software packages that are under development as well as the number of users of these tools (Schutzberg, 2005). EduCal draws upon some of the most successful OSG projects yet developed as the software foundations upon which its custom-programmed functions are built. In particular, EduCal utilizes PostGIS (http://postgis.refractions.net/), MapServer (http://mapserver.gis.umn.edu/), and Chameleon (http://chameleon.maptools.org/) to manipulate and deliver spatial data respectively. These OSG tools are used in conjunction with other OSS projects including the popular object relational database PostgreSQL (http://www.postgresql.org/), the world's most popular WebServer, Apache (http://www.apache.org/), and the very widely used server-side scripting language Hypertext Preprocessor (PHP) (http://www.php.net/). Further explanation of all of these technologies, the role they play, and the application programming interfaces (API) used to link them together is presented in the following section.

4.2 High Level Design of EduCal

As noted in the previous section, the combination of OS tools used for the development of the EduCal st-DSS includes the most commonly used set of OSG tools currently available. Each of these tools is fundamental to the creation of mapping capabilities, data storage and retrieval, a user-friendly interface and the linkages between the three. Each component was integrated modularly within the EduCal project to perform a specific task that, when combined in the right way, can provide powerful spatial and temporal decision support functionality within a space-time manifold. The high level architecture of the EduCal software environment is shown in Figure 4.1.

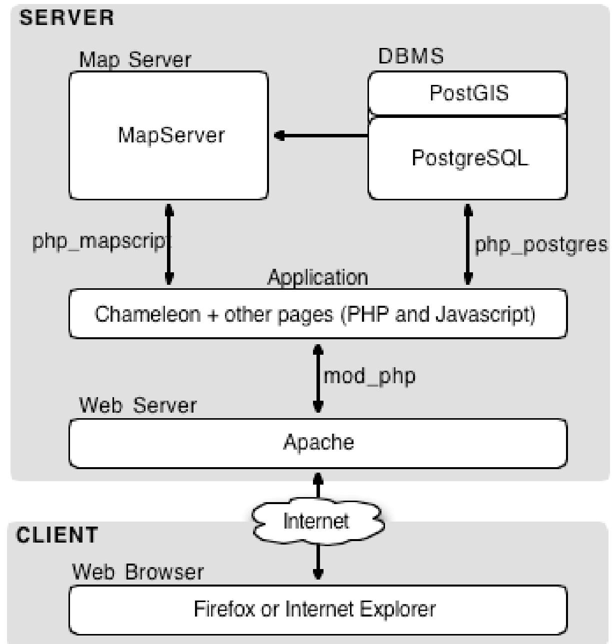


Figure 4.1 – EduCal Architecture

EduCal is a data-driven application and, as such, the software requires a reliable and scalable database back-end that is consistent with the data needs discussed in Chapter 3. This requirement places, at heart of the tool, the PostgreSQL object-relational database management system (DBMS). In its own, PostgreSQL does not have any spatial data handling capabilities, as it is a conventional object-relational database package that provides standard data management and processing functions. This is to say, it cannot store or process spatial objects (essentially, points, lines and polygons) that require the creation and storage of geometries reflecting real world objects and their absolute and relative locations.

To accommodate this need, the PostGIS module is layered on top of PostgreSQL as a spatial extension to the DBMS (Figure 4.1, upper right). PostGIS adds both spatial object geometrical data handling as well as a wide array of spatial operators that enable querying based on spatial attributes. The *intersection*, *contains*, and *distance* operators available within the PostGIS environment are required operations that are discussed later in this chapter.

The combined PostgreSQL/PostGIS spatially-enabled DBMS acts as the back-end for the EduCal software project. On the one hand, it provides the data store for the Web map service (WMS) that is provided by the MapServer component of the tool, while on the other hand it provides data directly to the application layer. The application layer makes direct use of the database's (spatial and aspatial) contents and the WMS utilizes the database to generate map images based upon user generated requests generated from the application layer.

The application layer in EduCal serves two purposes, namely it contains the program control logic and it is responsible for generating the user interface. This layer relies on the Chameleon environment, which provides a highly configurable toolbench for developing Web-mapping applications. Chameleon provides an object-oriented PHP foundation that can be extended in any number of ways. As Chameleon is built on the OS MapServer Project as the core mapping engine, it works with all MapServer supported data formats through a MAP file. A MAP file is a structured text file that describes every aspect of a map, including, but not limited to, the existing layers (including names, visibility, and order) and the symbology used (including shape and colour). These formats include PostGIS layers, Environmental Systems Research Institute (ESRI) Shapefiles and Geography Mark-up Language (GML), among others. This component of EduCal provides the underlying application environment through a series of 'widgets' or tools that execute PHP and JavaScript code snippets. In essence, each component of the EduCal interface (buttons and links) is a Chameleon widget that executes the attached code snippet on response to user-initiated interaction with the tool.

These code snippets make extensive use of two PHP modules, namely php_mapscript and php_postgres (Figure 4.1), on top of the already extensive PHP functions that are used in the basic Chameleon/MapServer environment. These two modules provide the necessary libraries to communicate with the WMS and with the DMBS respectively. More specifically, the PostgresSQL module provides the necessary methods to communicate with the database process running on the host operating system. These methods include *db_connect* to open a connection to the PostgresSQL socket, *pg_query* to execute selections from the database, and *pg_fetch_assoc* to place query results into a regular PHP array structure. Similarly, php_mapscript provides the methods required to communicate with the MapServer process. These methods include the *ms_newLayerObj* function to create a new layer, *setMetadata* and *getMetadata* functions to manipulate the layer's metadata, and general *set* and *get* functions to manipulate other information such as the layer source, colours, and extent. When *php_mapscript* methods are executed, the MapServer process manipulates the MAP file so that, when requested, a new map image can be generated and presented to the user based on the MAP file content.

Although Chameleon comes equipped with a large set of widgets that, out of the box, provide standard map manipulations, more complex functionality was needed for the EduCal project in order to meet the requirements of the tool. An additional series of custom programmed widgets were developed by the EduCal team at the University of Waterloo to extend substantially the standard Chameleon functionality and to provide advanced map exploration, selection and customized analysis tools. The new widgets, all OS in their implementation and distribution with the EduCal source code, provide the drilling operations and the type of st-DSS spatio-temporal querying described in the preceding chapter. The custom EduCal widgets also provide indicator selection and other non-SOLAP-related functionality to facilitate different aspects of user interactions.

As with all Web-based applications, user interactions with the application layer take place through the Internet browser (Internet Explorer, FireFox or Opera, for example) on a client computer. The browser then uses the available Internet connection to communicate with the application layer and back-end that reside on a Web server. As noted earlier, the standard EduCal installation makes use of the OS Apache Web server, although other Web servers (such as Microsoft's Internet Information Server) could be used, if properly configured. The combination of the Apache Web Server and PHP is commonplace. In fact, Apache accounts for over 70% of all Web servers in operation (SecuritySpace, 2006a) and *mod_php* is the most popular Apache module (in just over 40% of all Web servers) (SecuritySpace, 2006b). From the point of view of the Web server, the entire EduCal application is a standard Web page comprised of a series of PHP pages, with different widgets in different folders, that are executed upon request. However, the standard Web application model is not well suited for the types of map and data interactions possible with EduCal. The chosen application model and its implementation are described in the following section.

4.3 Application Model

EduCal software differs from the traditional Web application model in that it makes use of Asynchronous JavaScript and XML (AJAX), a term coined by Jesse James Garret in 2005

(Adaptive Path, 2005) and popularized by applications such as Google Maps™ and Gmail™. The AJAX application model (Figure 4.2) has proven to be highly effective in improving the responsiveness, speed and usability of Web pages, especially where high interactivity with complex objects is fundamental to the function of a Web site. In this sense, the AJAX model is a necessity for Web mapping services such as that used in the EduCal project.

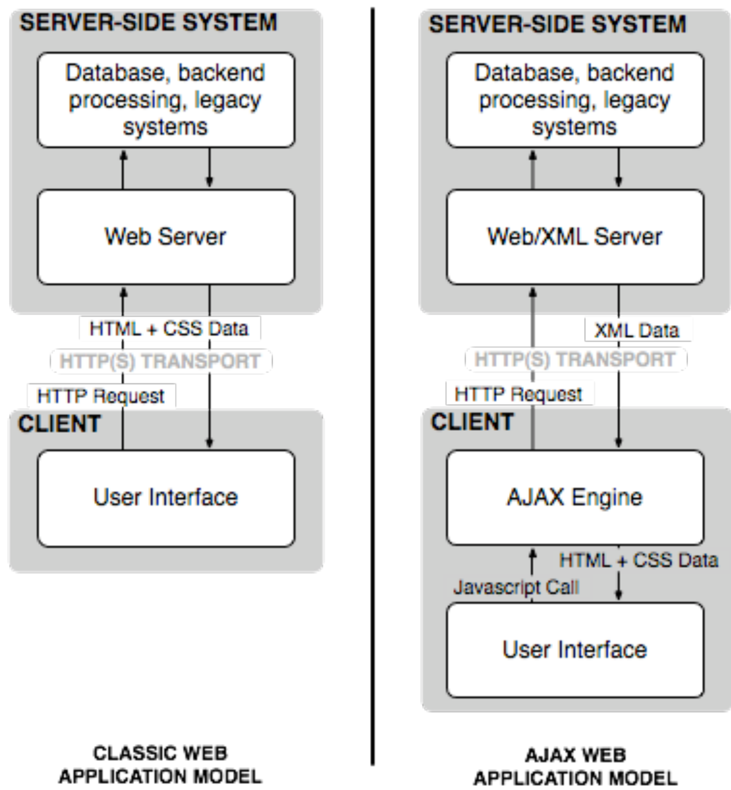


Figure 4.2 – The traditional model for web applications (left) and the AJAX model (right) (adapted from Garrett, 2005)

Under the traditional Web application model (on the left in Figure 4.2), any interaction with the server requires a complete Web page refresh on the client's browser. The obvious disadvantage of this is that, from the moment a request is made, the user must wait for the request to be sent, processed, and an HTML response received. Until the page is fully reloaded, the user cannot continue to interact with the application. In contrast, the AJAX application model (to the right in Figure 4.2), the interactions between the user interface and the server are mediated (asynchronously) by the AJAX engine. This allows for HTTP requests to interact with the server and respond with snippets of extensible markup language (XML) code that contain only the required information. The XML code may involve simple XML-tagged data, but it may also involve JavaScript invocation requests (function calls handled by the AJAX engine). This technique can be used fetch small portions of a page at a time without the need either to disable or reload the rest of the page. The end result is a much more responsive application and user interface from the perspective of the end user. Both of these considerations were fundamental to the design and implementation of the EduCal st-DSS. The increased response time is an advantage on generally old client computers with limited bandwidth connections to the Internet and potentially lengthy wait times for complex analysis results to be returned to the client browser. These facts are of significant importance relative to the development environment encountered in a developing country such as Peru.

In practice, EduCal uses the AJAX techniques described above in three major ways, namely for distributing code load times, for distributing data and information load times, and for maintaining an active and responsive interface during code execution time. From the initial page load, the Javascript and HTML of each widget is loaded in a piecemeal fashion. By sending scripts to the browser only when necessary, bandwidth requirements are distributed over the use of the application yielding a significantly faster application load time. Similarly, when data need to be sent to the client browser, AJAX requests and responses can be used to bring back only the data and information necessary for the part of the interface being interacted with by the user. This is particularly useful for the indicator selection widget, which would otherwise require that the indicator names and categories for over 500 indicators be loaded at once. Finally, maintaining a responsive interface is achieved by using AJAX to process queries in the background of the interface instead of in between page loads. An example of this can be found in the Query Builder discussed later in the thesis, where sample values are loaded for an indicator.

By using AJAX, the database can be queried for a list of sample values while simultaneously allowing the user to type in a value if they choose not to wait for the list to populate. Although this action only takes approximately a second in the case of a partial list and two seconds in the case of a total list, the continuously responsive interface provides a much more fluid user experience. This aspect of the user interface and user experience is of critical importance, as is demonstrated by the important role of selection actions in the application logic flow described in the following section.

4.4 Application Logic Flow

Equipped with the software architecture and application model outlined above, EduCal provides a flexible and intuitive environment for exploring and analyzing school-based education data with the intention of assessing education quality. The application logic flow described in Figure 4.3 explains a cyclical/iterative process for refining selected data and conducting analyses. Both of these processes are vital for education planners and decision-makers to understand and analyze poorly-defined, multi-faceted problems related to education quality, with minimal *a priori* knowledge (OSG, 2005).

Implicit in the application flow described in Figure 4.3 is the need for an analyst to be able to perform selections in a timely and straightforward manner. Decision makers are generally non-technical people with limited technical knowledge (and possibly also interest) required to write expressions using query languages such as structured query language (SQL) to query complex databases. A well-conceived st-DSS must allow the decision maker to analyze the data without being interrupted by the complex manipulations and technical knowledge that are required to seek new knowledge (Bédard, 2005).

The decision maker should only have to think of *what* they wish to select without having to worry about *how* to select it. Furthermore, regardless of which method of selection is used from those possible, the results must satisfy Newell's (1990) cognitive band and appear in less than 10 seconds on the client's browser for them to retain confidence in the use of the tool. Non-responsive applications will interrupt the cyclical/iterative process of end-user interaction and cause them to lose interest in subsequent use of any software package. Preliminary stress testing of EduCal with the components and approach outlined in Figures 4.1, 4.2 and 4.3, showed the software to be capable of performing well within this cognitive threshold, usually with response times of less than five seconds between end-user request and software response.

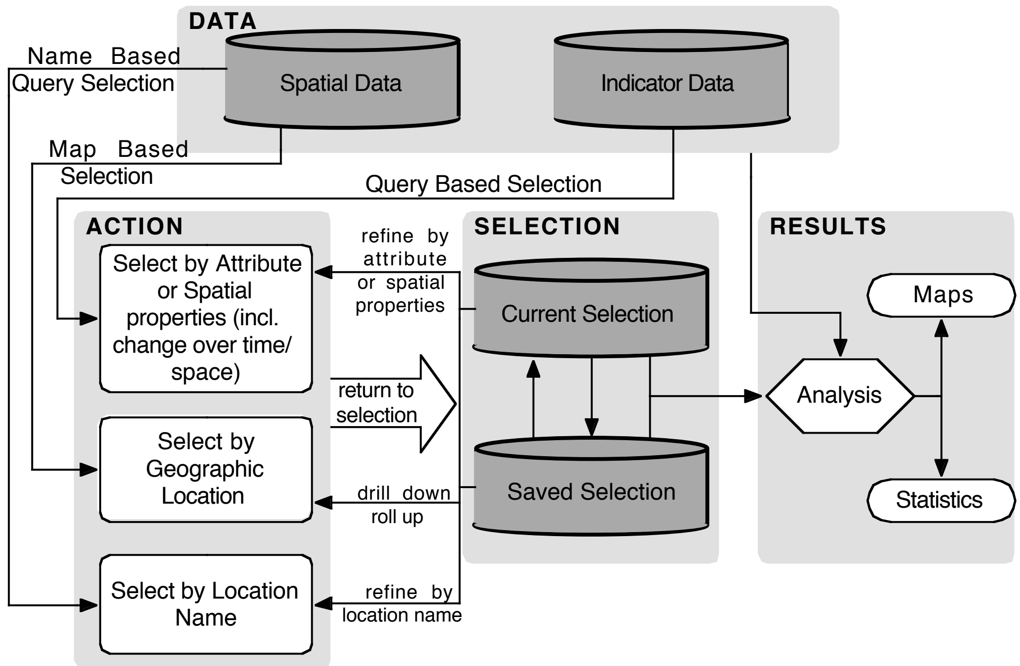


Figure 4.3 – EduCal Application Logic Flow

An end user request is initiated when the user wishes to explore the EduCal database. Data exploration can take two forms, namely users analyze the data directly or they refine an initial selection by one of the three actions on the left of Figure 4.3. However, analysis results are not necessarily the end product of this process. Map and tabular results are intended to serve as a guide for users to choose *which* action to take next. After viewing initial results, the user can return to their current selection and take an appropriate action to refine or further explore their preliminary analysis. This, in turn, will yield a new current selection and thus a new round of analysis can begin. Through each cycle, the ability to save selections allows users to keep track of interesting results to share with other users, to analyse along different lines of investigation, or simply to retrieve for additional analysis at a later time.

Conceptually, each action manipulates the data cube described in Figure 3.2 in different ways. These manipulations correspond to the SOLAP operations described in the preceding chapter. In practice, each action changes the selection along a different dimension through a query (or series of queries) formulated against the data warehouse.

In order to achieve efficient results, a carefully designed database schema had to be implemented along with extensive indexing of both the spatial and aspatial attributes contained in the data store. With the schema in place, the data loaded, and attributes and education quality indices calculated, the specific implementation of SOLAP operations is carried out through a series of efficient SQL statements that can be executed through the user interface and avoiding the need for the user to write potentially complex expressions in native SQL. The database schema, indices, and typical types of database queries relative to the assessment of the space-time manifold in education quality analysis are explained in the remainder of this chapter.

4.5 Database Schema

In order to support the functionality of the EduCal software, its database was designed with four types of tables, namely map (spatial) layers, indicator (attribute) data, metadata, and user (administrative) tables. The map layer tables (also known as feature tables) contain the names, identifiers, and geometries for the features at each supported level of spatial aggregation within the education hierarchy described in Chapter 3. Indicator data tables contain a column for each indicator described in Appendix A, the year and a link to the corresponding rows in the feature tables that allow attributes to be linked to spatial features. The indicator tables are the central part of the database schema, which is based on the *star schema* and the related *snowflake schema*. In fact, most data warehouses use a star schema to represent multidimensional data (Chaudhuri & Dayal, 1997).

In a standard star schema, there exists a single fact table and a single table per dimension. Each row in the fact table contains a primary key formed by concatenating the foreign keys of each of the dimensions. The fact table, at the centre of the star, contains measurements (indicators) that can be aggregated in various ways. The dimension tables form the points of the star, which provide the basis for the aggregations of the fact table. In a star schema, dimension tables tend to be highly denormalized while the fact table is the only one that remains in a normalized form. Under the related snowflake schema, some or all dimension tables are normalized to reduce redundancy.

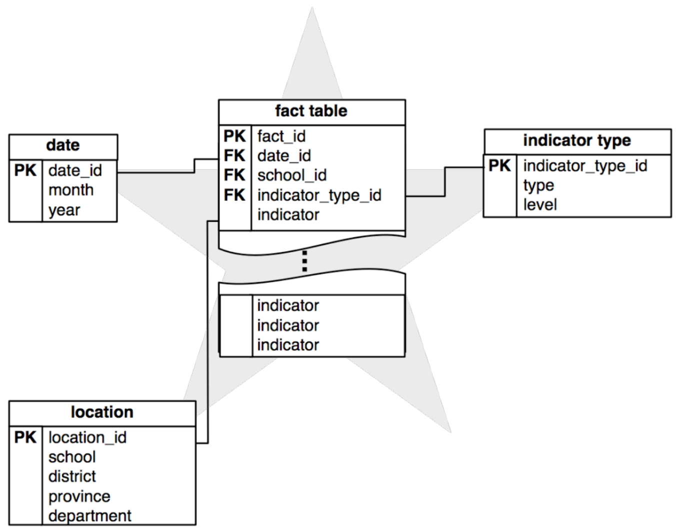


Figure 4.4 – Star Schema with EduCal Tables

Database normalization refers to the process of splitting up all non-simple elements into their atomic (non-decomposable) sub-elements (Codd, 1970). Codd (1971) subsequently defined the first three normal forms, where each is more restrictive than the next. In the typical star schema (Figure 4.4), dimension tables are in the first normal form (only requiring that each table have a primary key, thus ensuring no two identical rows) and the fact table is in third normal form (requiring that non-key attributes be dependent on the primary key and that there exist no dependencies between non-key attributes in the table) (Ahmad, Azhar, & Lukauskis, 2004). The snowflake schema differs in that the dimension tables are normalized to the third normal form (Figure 4.5).

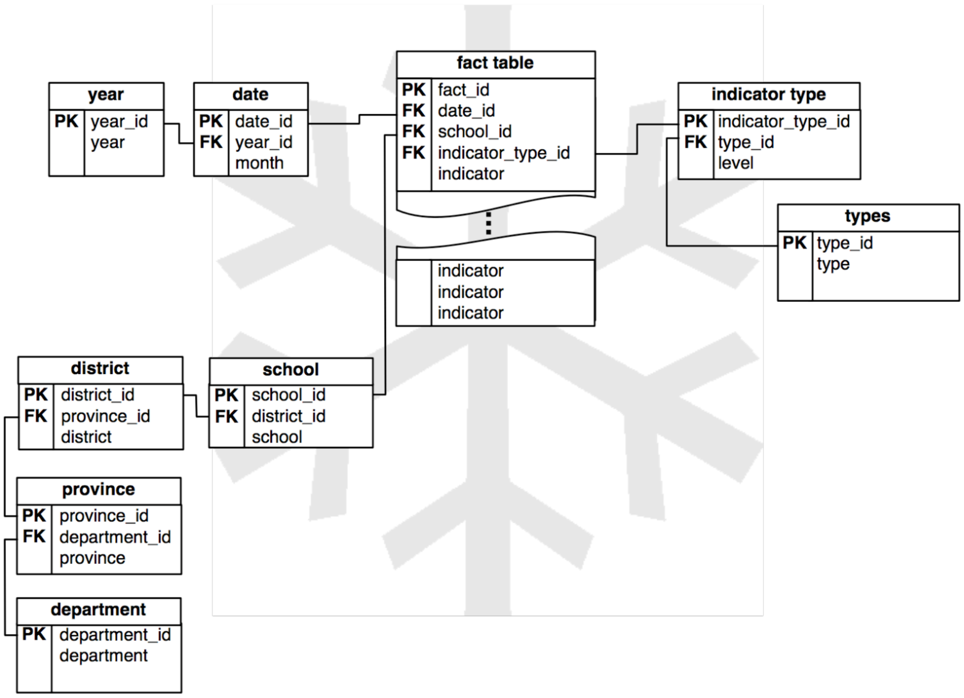


Figure 4.5 – Snowflake Schema with EduCal Tables

The normalization of tables is highly desirable in transactional systems because it ensures logical consistencies and reduces data redundancy. However, one of the main advantages of denormalizing tables with a star schema in a data warehouse is that it reduces the number of relationships between tables and therefore the number of potentially time-consuming joins required to complete and return user queries. Structuring multidimensional data with a star schema has been shown by Kimball (1996) to result in 80% of queries involving only a single table. In the case of the EduCal database, a variation of the snowflake schema was implemented which normalizes all tables while retaining simple querying (with few joins) with highly efficient results in the majority of cases.

The EduCal database schema (Figure 4.6) was optimized for the types of queries likely to be most commonly performed by users of the software (MINEDU planners at various levels of the system as well as school administrators, teachers and parent's groups). The normalization of all dimensions separates the various aggregation levels of the education system into different tables and, due to the nature of most common queries in EduCal, this approach proves to generate efficiencies both in query execution and in database maintenance. Specifically, the central role of spatial data in the analysis process offered by the application logic flow revealed by the database schema (Figure 4.3) places a high demand on the map layer tables one level of aggregation at a time, and this demand can be satisfied by the execution of look-up queries on a relatively small indexed table. Furthermore, as the application is requested for multiple layers of the map to be drawn simultaneously, multiple DBMS processes can serve these requests concurrently without having to share a table resource.

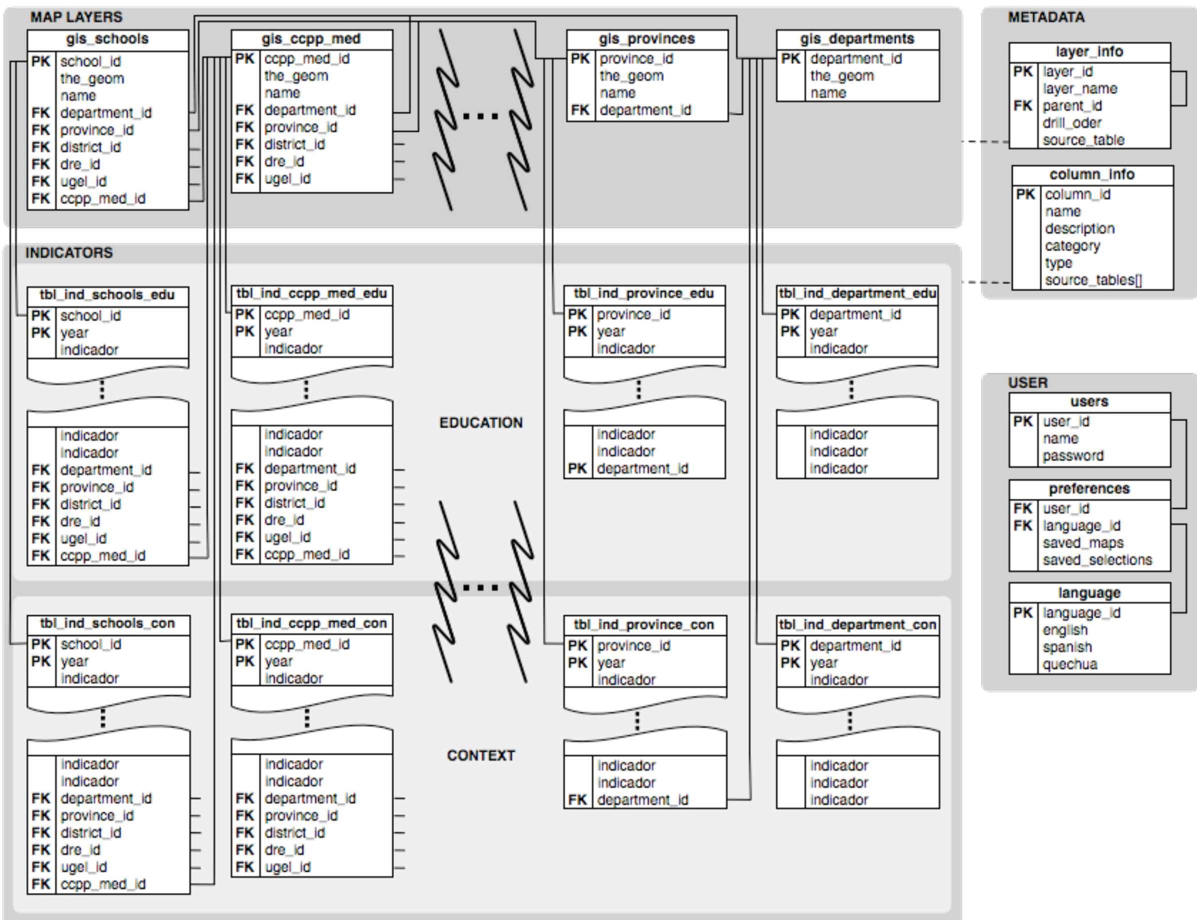


Figure 4.6 – EduCal database schema

As noted earlier, the chosen database schema makes extensive use of database indexes in order to make table joins and table searches more efficient. Indexing is a common database performance-enhancing technique that works by the use of a data structure (an index) that represents the data in one or more database columns in such a way that it reduces the number of disk accesses required to find a given record (row). Without the use of indices, access methods require an exhaustive search that, in the case of large volumes of data, is extremely inefficient.

Thus, a typical DBMS (for example, PostgreSQL, MySQL, Microsoft SQL Server) implements indices based on a classical data structure known as a B-tree to store pointers to table rows.

A B-tree index takes advantage of the natural ordering of the data (be it numeric or lexicographic) to reach the sought after row or rows efficiently. From an information standpoint, an index is a redundant component of the data representation (Codd, 1970) that simultaneously improves response times for database table queries and joins. In the case of EduCal, the data redundancy aspect of indices was not a limitation. Thus, in order to minimize response times, the database was indexed on every column that could potentially be queried, including database keys, indicator columns, and geometry columns.

Geometry columns require a different type of index than alphanumeric ones due to the inherently different nature of geographic data. Unlike alphanumeric data, geographic data do not have an implicit natural ordering and so a different indexing technique is required. In this case an indexing approach similar to the B-tree, known as an R-tree was used (Guttman, 1984) to organize spatial data in a tree-like structure. The main goal of an R-tree is to organize spatial data in such a way that a search will visit as few spatial objects as possible before returning the required result. In this way, B-trees and R-trees serve the same objective. However, while B-trees use the natural order of lexicographic data, the R-tree structure is based on a heuristic that orders spatial objects based on minimum bounding rectangles (MBR) of different subsets of the spatial dataset. The choice of MBRs is based on the assumption that objects that are in geographic proximity to one another are more likely to be queried together than those that are further apart. An example of a set of MBRs and the resulting R-tree is shown in Figure 4.7.

The top half of Figure 4.7 shows a series of white boxes, each representing a spatial object, and a series of grey rectangles representing various MBRs. In this example, each MBR is composed of at most three spatial objects (for example, R3 and R7) or at most three other MBRs (in this case R1 and R2). The index is composed by calculating the MBRs and then calculating the resulting R-tree, as is show in the lower portion of the figure. The tree structure represents the possible search paths, where each branch stemming from a node leads to all the objects and MBRs contained within that node. At the end of each branch is always a spatial object with pointers to the data associated with that object.

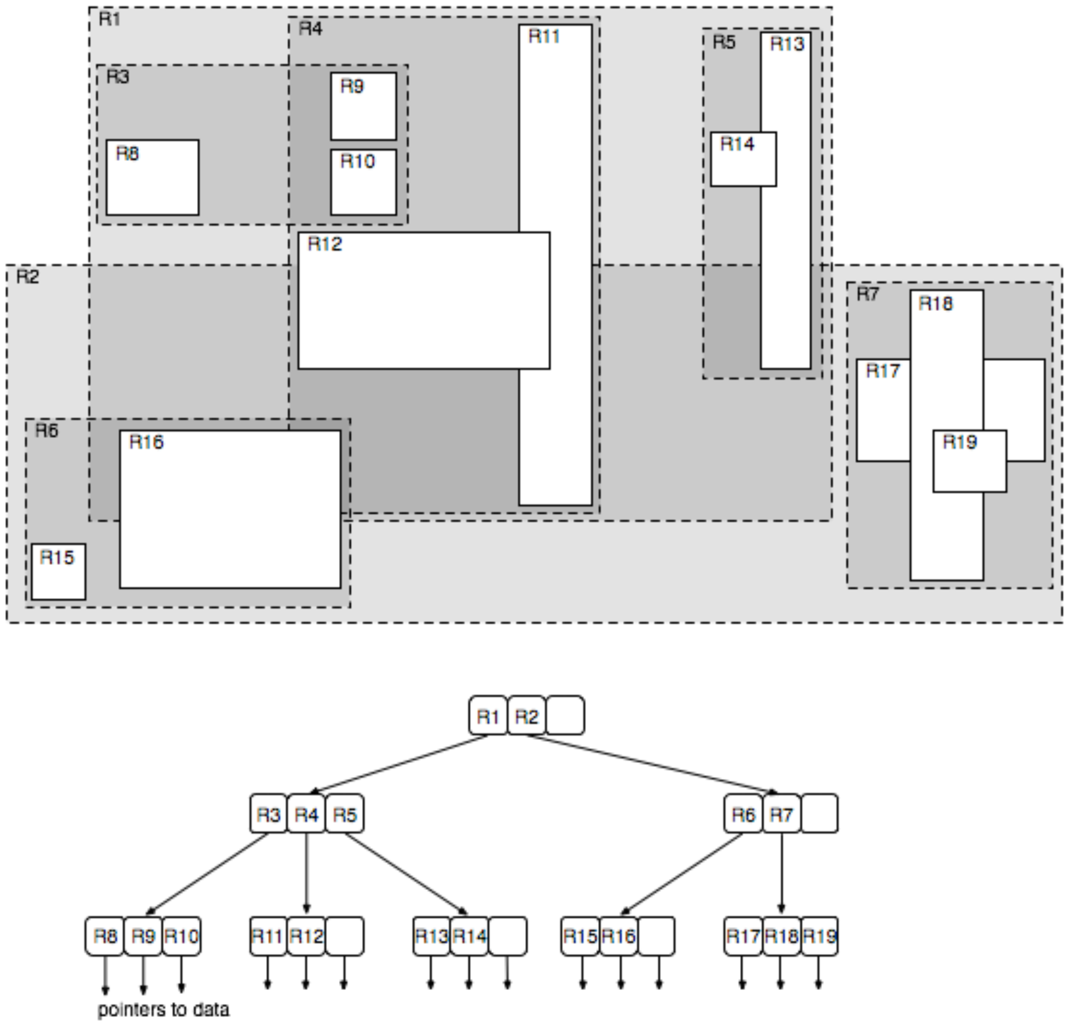


Figure 4.7 – Minimum bounding rectangles and the resulting R-Tree (Gütting et al., 1994)

Important variations exist on the original R-tree structure, such as the R+-tree (Sellis, Roussopoulos, & Faloutsos, 1987) and the R*-tree (Beckmann, Kriegel, Schneider, & Seeger, 1990), which use different heuristics for determining which MBRs to compute. PostGIS provides an implementation of the R*-tree variant for its spatial indexing. All of these variants, and other spatial indexing structures, are capable of supporting many different spatial operations through a *filter and refine* strategy (Gütting, 1994). *Filtering* works by traversing the R-tree using the MBRs, on which spatial predicates can be calculated quickly, thanks to their regular shape, to narrow down the set of spatial objects that need to be operated on. The *refine* step then uses the exact geometry of each object returned by the *filter* step. This strategy is highly effective in improving the speed of query processing by limiting, during the filter step, the number of objects that are used in the more complex and computationally expensive geometry calculations. However, it is clear that despite the efficiencies that can be gained by having an effective filtering structure and algorithm, spatial indexes are inherently more computationally intensive than aspatial indexes as they only limit the number of complex operations that must be computed in the best cases and do not reduce the number of operations in the worst cases.

As explained in the preceding section, the EduCal logic flow involves a geographical selection that users subsequently refine by making changes along one or more of the three data cube dimensions or by drilling down/rolling up the spatial dimension. Due to the central role of maps in the logic flow and the fact that spatial queries are more computationally intense than their aspatial counterparts, a secondary method for accessing spatial objects was created in the EduCal data schema to minimize the number of purely spatial queries that need to be performed. In order to turn some spatial queries into aspatial ones, numeric identifiers were assigned to each spatial feature at every level of spatial aggregation. Although, only a small subset of the possible spatial queries can benefit from the numeric identifiers, these queries are those that are performed most often within EduCal and thus provide a significant improvement in the overall performance of the software. In particular, the numeric identifiers are used to manage sets of spatial objects and, optionally, for drilling/roll up operations.

Selection sets in EduCal are the result of a variety of queries (including spatial queries). However, once a selection has been made, only a subset of the numeric identifiers that correspond to the selection set is saved. In future queries, attaining the resulting selection set can be achieved by using only the indices of the numeric identifier columns. For example, in the case where ten schools were previously selected, a simple SELECT query with a WHERE clause that uses the IN operator can be used to find the same set of schools without needing to calculate any spatial predicates (Query 4.1). In this example, ten school points (with the corresponding *school_ids* numbered 1-10) are selected from the school layer table *gis_schools* (Figure 4.4).

This method of maintaining sets proves to be highly efficient even in cases where the selection set is composed of hundreds of spatial objects and especially when compared to the equivalent fully-formed spatial query that would require that spatial predicates be computed on the same set of spatial objects.

```
SELECT the_geom
FROM gis_schools
WHERE school_id IN (1,2,3,4,5,6,7,8,9,10)                    (Query 4.1)
```

Numeric identifier columns are also used extensively in order to limit the number of spatial queries required for drilling down and rolling up. This is possible because the hierarchy of the education system is neatly nested. Thus every spatial object has exactly one parent. As such, it is possible to add one column to each map layer table for every layer that lies above it in the hierarchy. For example, the *gis_provinces* table has its own numeric identifier column (the primary key, *provinces_id)* but also has the identifier columns of the department and DRE that the province belongs to (the foreign keys, *department_id* and *dre_id*) (Figure 4.4). Using these additional columns, it is possible to avoid doing a spatial join between two layer tables when performing a drill down operation. For example, a drill down operation from a department with a *department_id* of 1 to the provinces that lie within it can be executed using only the province table (Query 4.2 and Query 4.3).

```
SELECT the_geom
FROM gis_provinces
WHERE department_id = 1                    (Query 4.2)
```

```
SELECT gis_provinces.the_geom
FROM gis_provinces prov JOIN gis_departments dept
        ON prov.the_geom && dept.the_geom
WHERE within(prov.the_geom, dept.the_geom)
AND dept.department_id = 1                    (Query 4.3)
```

Rolling up operations can be performed in a similar fashion. The complete set of parent spatial objects can be obtained by performing a SELECT DISTINCT query on the foreign key pertaining to the desired roll-up value. It is possible to find the encompassing set of provinces for the 10 schools of Query 4.1 by selecting the *province_id* of the *gis_schools* table and then eliminating duplicates the resulting set (Query 4.4). A second select can then be performed on the roll-up level table (either as a second query or as a nested query) using the resulting set of identifiers in the WHERE condition.

```
SELECT DISTINCT province_id
```

```
FROM gis_schools
WHERE school_id IN (1,2,3,4,5,6,7,8,9,10)                    (Query 4.4)
```

It should be noted, however, that EduCal provides the option, in every instance, of performing these operations in a purely spatial way using the *within* or *contains* spatial predicates (Query 4.3). Providing this option ensures that the drilling and roll-up operations programmed into the tool will work even in the presence of an education system with a hierarchy that is not neatly nested or with the inclusion of an external spatial data set that has not been coded with a numeric identifier to distinguish individual instances of spatial objects within a map layer.

The application logic flow of EduCal also places high demands on queries that involve both spatial and indicator data at the same level of spatial aggregation. By keying all of the tables at each level with the same numeric identifier, it is possible make highly efficient joins between two corresponding tables. For example, the *gis_provinces* and the *tbl_ind_provicnes_edu* tables are both keyed on a *province_id*, as are all the pairs of tables that are vertically aligned in Figure 4.6. In fact, two-table queries such as these (Query 4.5) form a large portion of the analysis-related queries in EduCal, as they provide users with the indicator data combined with the spatial data needed for the map interface. In the cases where both education and contextual data are selected for display or analysis, queries require a three table join, again based on the primary key of each of the tables (Query 4.6).

```
SELECT g.the_geom, i.indicator1, i.indicator2
FROM gis_provinces g JOIN tbl_ind_provinces i
      ON g.province_id = i.province_id                       (Query 4.5)
```

```
SELECT g.the_geom, i.indicator1, i.indicator2
FROM (gis_ provinces g JOIN tbl_ind_provinces_edu i
      ON g.province_id = i.province_id) JOIN tbl_ind_provinces_con i2
      ON g.province_id = i2.province_id
WHERE i.year = 2002 AND i2.year = 2004                       (Query 4.6)
```

The separation of education data relative to contextual data was chosen based on the lack of accordance in the time dimension between these two very distinct data sources. While, in both cases, the time dimension has only one level of spatial aggregation (annual), the set of years available for each source essentially comprise two parallel dimensions. Moreover, in a large number of cases, users will maintain a separation between contextual layers and education layers. The separation of the indicator data into two sets of tables results in smaller tables being used in these frequent cases. It was possible to accommodate efficiently the separation of indicator data largely because the time dimension, as it pertains to EduCal, does not have its own multi-dimensionality requiring its own set of temporal tables. The time dimension was incorporated into the data schema through the addition of an indexed and keyed *year* column to the indicator tables.

By not having any tables pertaining to time as part of the data schema, moving a selection *across* time can be achieved by simply selecting a different year in the WHERE condition of the query. However, this approach also forces comparisons across the time dimension to take a different form than comparisons across either of the other dimensions. Comparisons across both the space and the indicator dimensions are accomplished by changing the selected columns (and/or expressions) without a need to change the currently active table(s). However, comparisons across time require two instances of the currently active indicator table to be joined to the map layer table based on the primary key, but with a different year chosen for each instance (Query 4.7). Although computationally more intensive than other *across* operations, the resulting query is no more complex than the equivalent query under a traditional star or snowflake schema. Under any of these schemas, joins are performed between two instances of the indicator table and the appropriate time dimension table.

```
SELECT g.the_geom, i.indicator1, i.indicator2
FROM (gis_geometry g JOIN tbl_ind_provinces i
      ON g.province_id = i.province_id) JOIN tbl_ind_provinces i2
      ON g.province_id = i2.province_id
WHERE i.year = 2002 AND i2.year = 2004                    (Query 4.7)
```

As a result of the chosen schema, selections across the time dimension are the only ones that require the computation of an additional JOIN operation. However, the use of indices on these tables and limiting the number of rows being joined by a geographic selection translates into fast response times even for these more complex queries.

Helping in the automatic formulation of all of these queries is the *map_layers* metadata table. The *map_layers* table contains the names of all of the map layer tables as well as the names of the indicator tables associated with any given map layer. The relationship (i.e. the JOINs) between the map layer table and the indicator tables is explicitly defined in this table so that it can be extracted and used from within the code in order to generate the entire SQL statement dynamically. This table also contains the hierarchical relationship between the different map layers so that it is possible to drill down or roll-up one level without making the user specify the target level. Other important characteristic information, such as the name of the *id* field and the type of feature (POINT, LINE, or POLYGON), is also stored within each layer's row.

Just as there is a *map_layers* table to hold metadata for the map layers, there is a *column_info* table to hold metadata regarding the indicator tables. Unlike the first, this second metadata table is used for controlling the interface and not for the formulation of queries. Three key bits of information are required in order to present the indicators through the graphical user interface in an organized and intelligible manner. The first is the categorization of the indicators under the five education components outlined in Chapter 2. To organize the indicators further, the internal hierarchy of the indicator dimension is also stored within this table. Second, this table stores information concerning whether an indicator is a cost or a benefit. This information is very important at the time of performing a target-based analysis of education quality, as explained in the following section. Finally, the third type of information stored in this table pertains to the indicator names, both the field names in the database and their descriptions in both Spanish and English. All of the metadata contained within this table and the *map_layers* table described previously is heavily used in the provision of all the aspects of the EduCal user interface that interact with the rest of the database. All of these aspects are described in the following section along with the rest of the interface components.

4.6 Graphical User Interface

The EduCal graphical user interface (GUI) provides an intuitive way to access to all of the data contained within the database. Moreover, the interface provides the display, selection, and analysis tools that are required for data exploration and knowledge discovery. The interface takes into account the norms of both standard GIS packages and of the Web. At the same time, every effort is made to ensure that EduCal is presented in a manner that is both intuitive and that guides the user through the analysis process. A key accomplishment of the interface design is the ability to provide these tools within a map-centred interface. A summary of the most important characteristics follows, with detailed descriptions of the tool available in each of three EduCal Tutorial Books written in both Spanish and English during the EduCal project (Hall, Engler, Alperin, & Leahy, 2005).

4.6.1 Map Interface

The choice of a map-centric GUI was based on the experiences learned from the first implementation of EduCal version 1 that only presented a map at the end of the analysis process. Version 1 of EduCal forced users to perform analysis through a linearly structured process involving area selection, indicator selection, optional weighing of indicators, and finally displays in either map or table form. Each of these steps has an equivalent in the application logic flow (Figure 4.3) built into the current version of EduCal. The first two of these steps correspond to refinements along the space and indicator dimensions respectively and the second two form part of the results toolkit. One of the fundamental differences between Versions 1 and 2 is that the map-centric interface, by showing the current selection persistently, allows for different actions to be intermixed. This allows the analyst to perform iterations in their analysis without the need to return to the first step every time. In order to achieve this added flexibility, the EduCal map interface provides all the necessary tools simultaneously.

An EduCal user is presented with a Web interface that is not unlike that of many desktop GIS packages (

Figure 4.8). The look and feel of the traditional elements of the main form (e.g., buttons, drop-down menus) is changed through the use of a cascading style sheet (CSS) and button images, and non-traditional elements are added to enhance the interface (e.g. tabs to maximize the use of the available space on the screen). Furthermore, as explained in Section 4.3, in order to mimic best the response times and interactivity of desktop (non-Internet-based) applications, all the GUI elements are delivered through the use of the AJAX application model. Finally, these elements are arranged in the browser window in an organized fashion so that the user can easily find the appropriate tool.

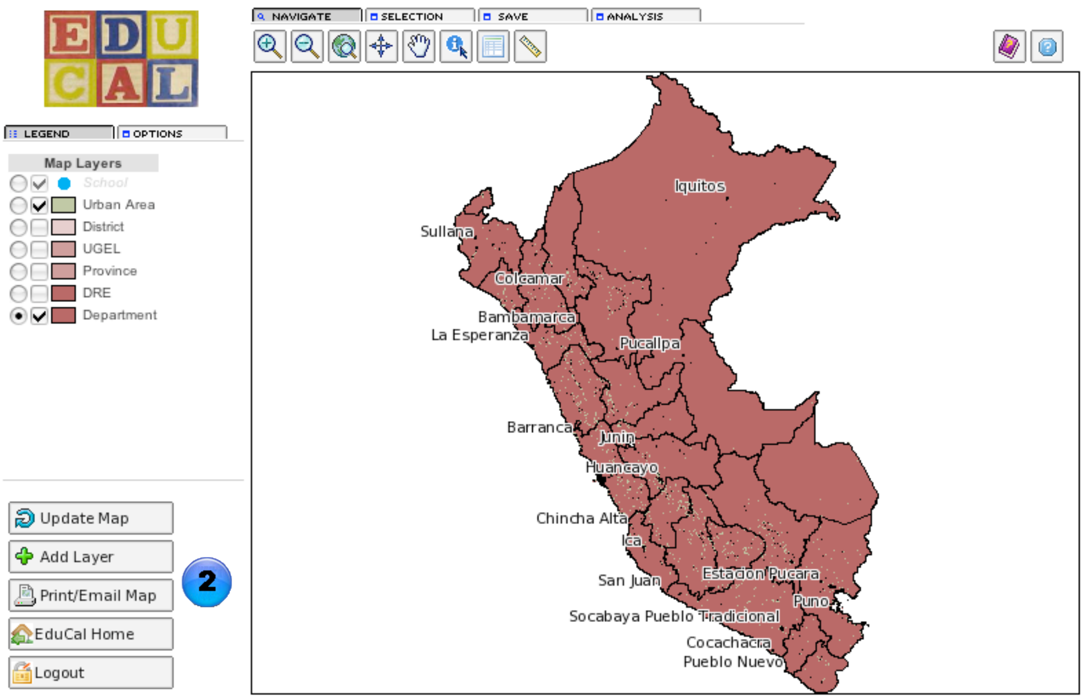


Figure 4.8 – Map Interface components

The EduCal map interface is organized around four key areas, which are described below:

1. The *Legend and Options* area of the page is used to control the visible and active layers, as well as interactions with the map.

2. The *Task Buttons* area of the page contains functions for adding new layers (the start point for most analyses of education quality), as well as printing and emailing the map.

3. The *Toolbars* contain all the tools that are used to navigate the map, select areas of interest, open and save scenarios, as well as create and develop an analysis of education quality.

4. The *Map* is used to make selections and view the results of the education quality analysis.

In traditional cartography, the legend simply explains the symbology of the current map. In EduCal, the legend not only retains this traditional role, but it also becomes a centre of control for what is displayed on the current map, as is the case with most interactive maps. The legend allows a user to control what is shown in the map in two ways. First the map can be controlled from the legend itself, where it is possible to turn layers on (show) or off (hide), select an active layer, and rearrange the layer order (Figure 4.9). At the same time, the legend provides an interface for editing a layer's content (activating the indicator selection screen described in Section 4.6.2), changing the symbology and labels, and deleting a layer entirely. This functionality (activated by double-clicking a layer name) gathers all layer-specific information and changes in the most logical part of the interface.

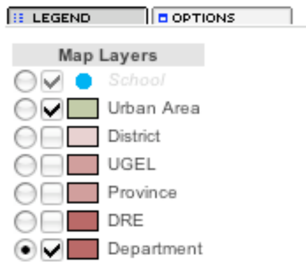


Figure 4.9 – Legend

Located next to the *legend* tab is the *options* tab (visible in Figure 4.9). This tab contains options related to map visualization that are generic and not tied to one specific layer. For example, depending on what is most appropriate for the size and resolution of the available monitor, the map viewing area can be changed using a drop-down menu. Figure 4.10 shows another drop-down menu used to control the layer used for map tips. The map tips tool is used to display, for a selected layer, information about the features that are found below the cursor when it hovers over an object in the active layer (Figure 4.10). Map tips are extremely useful for quickly and easily identifying map features that are of interest as well as identifying constituent information for the object from its data table.

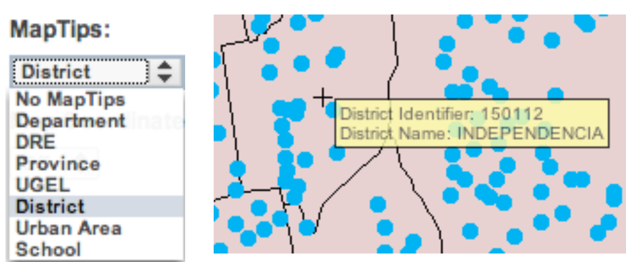


Figure 4.10 – Map tips

The second part of the map interface, *task buttons*, contains the *Update Map*, *Add Layer*, *Email/Print*, *EduCal Home*, and *Logout* buttons. Clicking the *Update Map* button will cause changes in the visibility of layers to take effect. The *Add Layer* button will generate a new layer and prompt the same interface as editing a layer from the legend. The *Email/Print* button is used either to print the current map for inclusion in a report or email the map to someone for discussion. Finally, the *EduCal Home* and the *Logout* buttons direct the user away from use of the EduCal tool. All of these buttons share the characteristic that they should not be hidden behind tabs or windows and need to be visible at all times.

The third section of the EduCal map interface contains the four tabs that appear in the top *Toolbars*. These tabs contain all the tools needed to navigate the map, select areas of interest, open and save scenarios, as well as create and develop an analysis layer. Respectively, the four tabs in the toolbar are *Navigation*, *Selection*, *Save*, and *Analysis*. The *Navigation* tab is comprised of the traditional array of map tools, such as *zoom in/out*, *pan*, and *recentre*. The navigation tools also include an *information* tool that returns information about any feature (or set of features) in the active map layer by clicking (or dragging) with the mouse.

A key element in the navigation toolbar is the *table view* button. This button launches a second window (Figure 4.11) that contains a table of information about the currently active map layer. A table view may reveal different kinds of information than the map view despite being populated with the same data. For example, by using the sorting feature on the top right of Figure 4.11, it is possible to determine at a glance which features have the lowest and highest values. The table view window also allows the user to switch to any other map layer without the need to return to the map by using the drop-down box on the top left. However, one of the most important features available from this view is the capability to download the currently displayed table.

The *Download* button, located in the lower portion of the window, allows the user to download either a Microsoft Excel (.xls) file or a comma-separated file (.csv) with the table information. A Shapefile (.shp) of the geographic features that accompany the table is also generated and downloaded at the same time. The availability of both the geographic and the indicator data in common formats allows an analyst to import EduCal data into any other application for further analysis (e.g. statistical analysis with SPSS or some other statistical software package) or to generate other displays (e.g. charts with Excel and maps with ArcGIS or another GIS that can read ESRI shapfiles). These downloadable files can also be shared with other analysts.

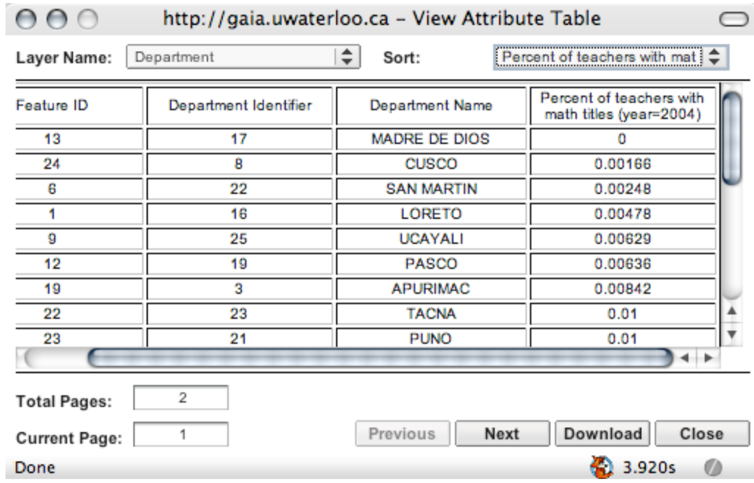


Figure 4.11 –Table View

Of the remaining three tabs in the Toolbars section, *Map Selection* is discussed in detailed in Section 4.6.3 and the *Analysis Tools* interface in Section 4.6.4. Finally, the *Save* tab holds buttons for opening, saving and deleting maps and selections. In this context, a map is understood to be all of the layers, their related data, and the symbology and a selection is the layer comprised only of the spatial features under current consideration. In essence, the save tab provides the tools from the *current selection* to the *saved selection* in the application logic flow of Figure 4.3.

Before continuing onto the last two tabs in the toolbar, the add/edit layer dialog mentioned previously is discussed in detail. This dialog is particularly useful for selecting or making changes along the indicator and time dimensions and as such it is fundamental to the education quality analysis presented in this thesis.

4.6.2 Indicator Selection

The dialog in Figure 4.12 can be invoked either by double clicking on a layer name in the legend or by pressing the *Add Layer* button. The main part of this screen (1) contains a tree view of the available indicators categorized by the five-components of education quality outlined previously in this thesis. Sub-categories (the internal hierarchy of the indicator dimension) can be found in each sub-branch of this tree. However, the entire tree is not loaded at once. Instead, the tree is dynamically loaded so that sub-branches are sent to the browser on demand, thereby decreasing the initial load time of the dialog. To ease the burden on the user, the number of indicators displayed at once can also be limited using the radio buttons along the top of the dialog (2). Indicators are classified as either beginner, intermediate, or advanced depending on the likelihood they are of being used. Indicators that are very specific to one type of analysis or that are not generally included in the evaluation of education quality will only appear when the intermediate or advanced option is selected.

Indicator selection is achieved by expanding the branch that contains the indicator of interest and subsequently placing a check in the box below its name. Indicators for which there exist multiple years of data will display a checkbox for each year, allowing for any number of years to be included in the layer. If the analyst is interested in a single year only, the *Year Filter* dropdown menu near the top left (2) can be used to eliminate all other years from the interface thereby *slicing* along the time dimension. The drop-down box can also be used to perform a *drill-across* operation along the time dimension. For example, a layer currently using only 2002 data can be switched to the 2004 values of the same indicators by a simple switch in this dropdown. The default value for this menu is *All*, which allows the selection of multiple year instances of the same indicator.

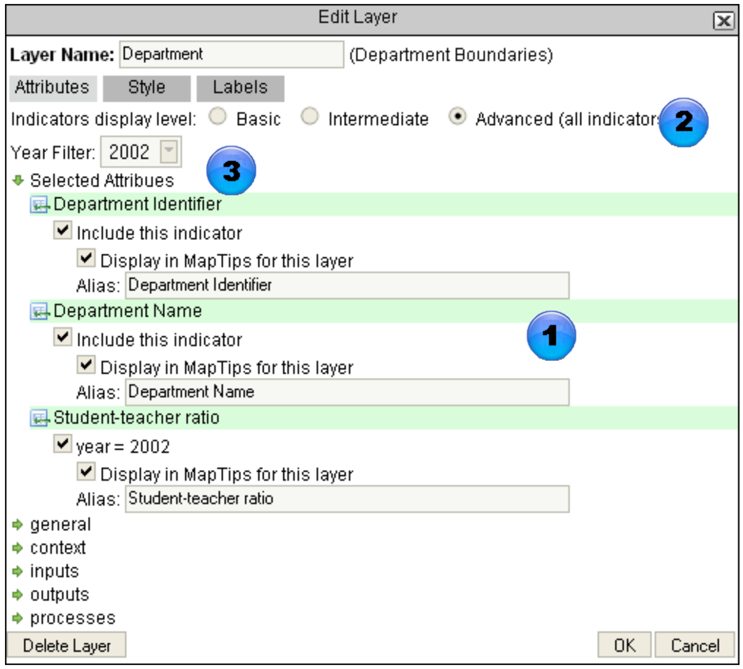


Figure 4.12 – Edit layer dialog

The selection along the indicator dimension and the time dimension are intricately tied together in this portion of the interface. The integration of selections along both dimensions is possible in large part because of the simplicity of the time dimension in the EduCal data cube. A hierarchical time dimension would warrant a set of time selection menus. Conversely, the spatial dimension is sufficiently complex to warrant an entire toolbar of map selection tools as discussed in the following section.

4.6.3 Map Selection Tools

The selection toolbar is available in the *Toolbars* area and contains all the tools necessary for the creation of a selection set of spatial features. Selection sets are created by selecting a desired mode, style, and shape tool and then using the mouse on the map to create a shape. All features from the active layer (shown in Figure 4.9) that coincide with a shape drawn by the user on the screen (based on your selection settings) is added to (or in some cases, removed from), the current selection set. The modes, styles and shapes are summarized in Appendix B.

The map drilling tools are also found in the *Selection* toolbar. As described in the application logic flow (Figure 4.3), drilling down/rolling up are distinct actions that enable the current selection to be changed. Unlike the above-described tools, the drilling drop-down box (shown and described in Appendix C) create a new selection based on the geometric properties of the previous selection.

The combination of the map selection tools with the drilling selection tools enables every type of spatial selection possible within EduCal, with the exception of the *Query Builder*. However, the *Query Builder* is grouped with the analysis tools since the type of selection actions performed with this tool are more exploratory in nature than those summarized above. The *Query Builder* is further explained in the following section, along with two analysis tools: *Basic Statistics* and *Target Based Weights Analysis*.

4.6.4 Analysis Tools

Three important analysis tools are accessed through the *Analysis Tools* tab. The tools found within this fourth and final aspect of the EduCal toolbox facilitate major aspects of education quality analysis by providing a statistical view of layers, the creation of new analysis layers, and advanced spatio-temporal selection for data exploration.

The first of these tools is the *Basic Statistics* tool shown in Figure 4.13. This tool can be used to view a statistical summary for data in any numeric column of any layer that is visible on the map. The summary includes the minimum, maximum, mean, and standard deviation as well as the number of null and zero values in the column of data.

By default, the dialog is preloaded with the currently active map layer and includes statistics for the current selection. However, the drop-down menus on the top allow the layer and indicator to be changed. Similarly, the drop-down menu on the bottom provides the option of viewing basic statistics for the entire layer or only for the current selection. Changing any of these options automatically updates the statistics without the need to open a new window, thereby making the navigation of different statistical values quick and seamless.

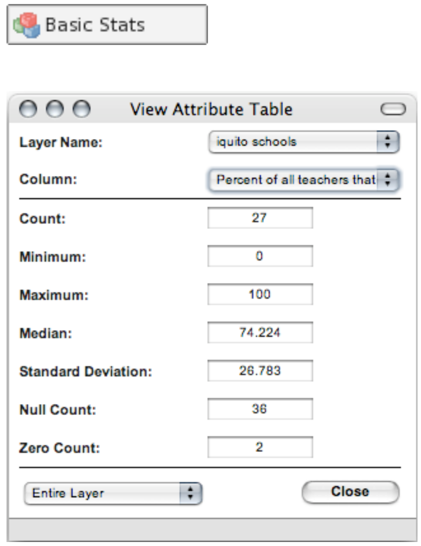


Figure 4.13 – Basic Statistics

All of the tools presented to date work only with the existing map layers and indicators. The *Target-Based Weights Analysis (TBWA)* tool (Figure 4.14) eliminates this restriction, thereby enabling a more advanced analysis of education data and allowing for the creation of new analysis layers in which indicators can be combined to form a composite index of education quality. This composite index is calculated by setting a valid range of values for each selected indicator (minimum and maximum values) and then choosing a desired value for each indicator (target value). Indicators are then combined using user defined weights, where the user assigns higher weights to indictors that are of more importance to him/her so that the higher weighted indicators affect the overall outcome of the analysis more than other indicators. The purpose behind this type of analysis is to normalize and combine the data values from multiple indicators into a single value representing the quality of education. This tool, its use, and the value of this type of analysis are all fully explained in the case study in the following chapter.

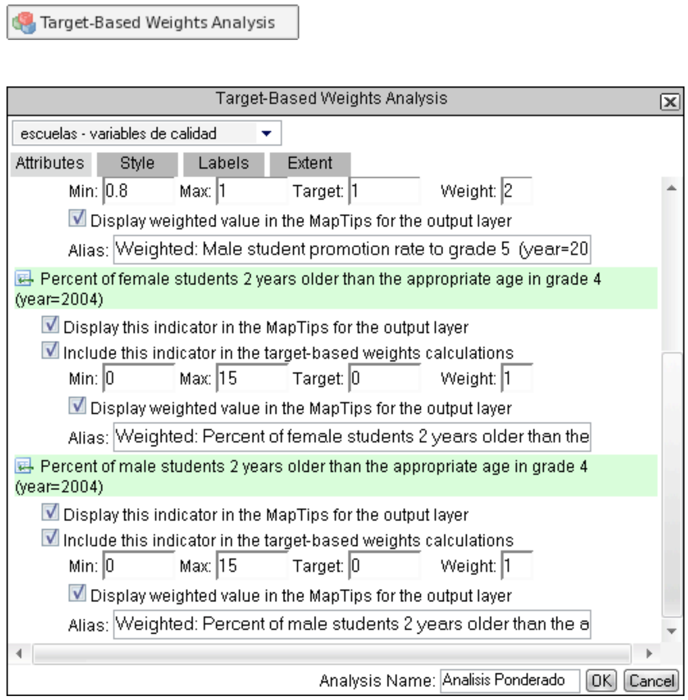


Figure 4.14 – Target-Based Weights Analysis

The last of the analysis tools is the *Query Builder* (Figure 4.15). The *Query Builder* enables a more complete form of feature selection than those presented earlier. By combining numerical, spatial, and temporal operators into a single tool, a user can select features from any layer that match the specific characteristics of interest. By navigating the interface shown in Figure 4.14, users can create a natural language query (a query statement written in plain English or Spanish rather than using the actual SQL syntax required to execute the query in the database) that corresponds with the aspect of the data they wish to explore. The tool translates the query, behind the scene, into the syntax required its execution. Conditions are specified by selecting operators and operands from the lists provided and by inputting a value into a text box when necessary. Once the condition reads as desired, from left to right, the *Add to Query* button can be used to move it to the top of the window and a new condition can be added. Conditions can be combined with either a logical AND or a logical OR as desired, and conditions can be removed if a mistake was made by the use of the *Remove* button to the right of each condition. Once all of the desired conditions are found in the query pane, the *OK* button causes the query window to disappear and the resulting selection set will appear in the map.

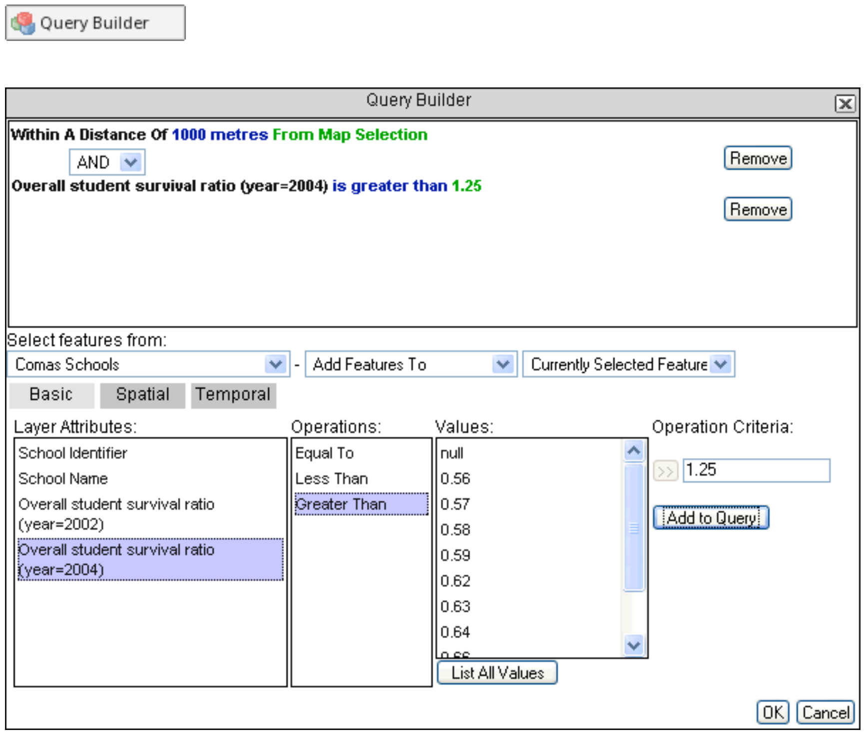


Figure 4.15 – Query Builder

The simplicity of the *Query Builder* layout hides the complexity of the set of operations possible through its interface. These operations can be found within the tab of each of the three sets, including *Basic*, *Spatial*, and *Temporal*. The basic operators (Table 4.1) require that users select an indicator, an operator, and a value. In order to help the user, a sample list of values is populated upon selecting an indicator and a full list of available values can also be loaded on-demand.

Operator	Type of Indicator	Description
equal to	alpha-numeric	Returns all of the features from the selected layer that have a chosen indicator with values exactly equal to some specified amount
greater than	numeric	Returns all of the features from the selected layer that have a chosen numeric indicator with values greater than some specified amount
less than	numeric	Returns all of the features from the selected layer that have a chosen numeric indicator with values less than some specified amount

Table 4.1 – Basic Operators available in Query Builder

The second tab contains a set of spatial operators (Table 4.2) that vary depending on whether or not the selected layer is a point or a polygon.

Operator	Type of Features	Description
adjacent to (beside)	polygon to polygon	returns all polygons that share at least one edge with the polygons of a previous selection or of another layer
discontinuous from	polygon to polygon	returns all polygons that do not share even one edge with the polygons of a previous selection or of another layer
within a distance of	polygon to polygon; point to point; point to polygon; polygon to point	returns all shapes (points or polygons) that are within a straight line distance of a user specified amount from any shape in a previous selection or another layer
at least a distance of (beyond)	polygon to polygon; point to point; point to polygon; polygon to point	returns all shapes (points or polygons) that are beyond a straight line distance of a user specified amount from every shape in a previous selection or another layer

Table 4.2 – Spatial Operators available in Query Builder

Certain operators, such as *adjacent to*, do not apply to a point layer since the concept of adjacency is not well-defined for point sets. Unlike the basic operators, the spatial operators do not require the user to select an indicator, as they operate on the entire layer. Spatial operations require users to select an operator from the available list and a layer to which comparisons are made. All of the available layers, as well as the current selection, are listed for the user. For the operators *within a distance of* and *at least a distance of*, the user is also required to input a numerical distance.

The third tab contains a set of temporal operators (Table 4.3) that can be used to operate on any indicators for which there are multiple years of data selected. Only those indicators for which there are multiple years of data appear in the attribute list. Temporal conditions are based on an attribute, an operator from the list in Table 4.3, and a base year for comparison. For the operators *increased by* and *decreased by*, users also have to input either an amount or a percentage to use as the basis for the comparison.

Operator	Description
increased	returns features that have increased by any amount since any other year

increased by	returns features that have increased by a user-specified amount or percentage
decreased	returns features that have decreased by any amount since any other year
decreased by	returns features that have decreased by a user-specified amount or percentage

Table 4.3 – Temporal Operators available in Query Builder

By adding basic, spatial, and temporal conditions through the *Query Builder*, users can explore the data in the multitude of ways described only conceptually earlier in this thesis. The *Query Builder* concretely implements, through a clickable interface, the ability to formulate very specific and potentially very complex queries that are spatio-temporal in nature. In this way, queries can be built using natural language constructs by the user which are automatically translated into efficient SQL statements using the techniques explained in Section 4.5. Accordingly, the *Query Builder* is a *de facto* key part of education analysis with EduCal. Going back and forth between the *Query Builder* and the map-based results becomes a simple, yet crucial, part of the analysis process. This process is complemented by all of the previously mentioned tools, and combined they give users information that can be used to find schools and regions of interest.

As can be seen by the features highlighted in this section, the EduCal interface is both robust and thorough. It allows for very complex manipulations of spatio-temporal education data in simple and intuitive ways. Moreover, the GUI does not restrict which facet of the data is analyzed at any step, giving users complete control and permitting multiple paths to reach the same map result. In this sense, the EduCal interface described above satisfies the requirements set out by the desired application logic flow described in Section 4.4 and shown in Figure 4.3. In fact, the only GUI elements highlighted are those that relate to the type of data interactions mentioned throughout this thesis. Further details can be found in (Hall, Engler, Alperin, & Leahy., 2005). However, despite the high degree of success with which the EduCal GUI meets the requirements set out previously, there exists a small set of features that are desirable but have not yet been added. These are summarized in the following section.

4.6.5 Future Developments

Although the current version of EduCal aptly meets all of the requirements outlined throughout this thesis, there are a few areas where further development of the interface would significantly enhance the user experience by providing greater flexibility in the analysis process.

It should be noted that all of the enhancements listed in the remainder of this chapter could be implemented using the architecture and database schema employed in EduCal.

In the current workflow, drilling across the time dimension is done through the layer-editing interface. This interface includes the available years for each indicator and a filter to slice the temporal dimension to a single year. However, it would be beneficial to add a single button that enables drilling across time (by going to the next/previous year available) keeping all other aspects unchanged. Such a feature would operate on the currently active layer and would allow the analyst to visualize changes without the need to leave the map view. A similar feature could also be employed to perform the drill up and down operations over space.

Although a drilling tool is built into the map selection toolbar, this tool only operates on the selection layer and cannot be employed to change the drill level of any other map layer. In the current EduCal workflow, users must perform the drill operation on their selection and then add a new layer based on the latest selection. This process can be cumbersome if users wish to do many comparisons between layers of aggregation. A single button tool that operates on the active layer would greatly facilitate multi-levelled analysis scenarios.

Lastly, there exists a limitation in terms of how indicators are combined both in the feature selection process and in the visualization process. In order to augment the current feature selection tools already in EduCal, an expression builder could be added into the *Query Builder* in order to allow selections based on combinations of variables (e.g. selecting all schools where the enrolment rate is less than the desertion rate). Similarly, users may wish to map expressions as a way of creating their own indicators. In its current form, EduCal only allows for the new creation of indicators through the TBWA tool, but this tool is restrictive in terms of the combinations that are possible. A new analysis tools that allows for the creation of indicators based on user-built expressions would greatly enhance the analysis capabilities available in the software.

As with all software, there are an infinite number of features, modifications, and visual enhancements that could be implemented to improve the tool. The refinement process of software such as EduCal can be perpetually extended in order to adjust for individual work practices and to incorporate newer user interfaces and techniques. However, in its current form, EduCal is considered to be a final release version that is sufficiently complete and user-friendly to conduct meaningful analysis of the education quality being received by primary school aged children in Peru.

4.7 Summary

This chapter has presented the open source software for the evaluation of education quality, EduCal. The Chapter began with an overview of the software and the open source philosophy that was followed since its conception. The open source components employed and their interactions were subsequently identified during an overview of the software architecture. Notably, this software architecture is general enough to be employed in a wide range of OSG applications (Hall et al., 2007). The section that followed described how the software architecture was implemented using the AJAX application model to improve the interface response times.

The remainder of the chapter explained in greater detail the implementation of the EduCal tool. First, the application logic flow was described. In the case of EduCal, the logic flow is the designed interaction between the user and the application so that the type of SOLAP operations described in the preceding chapter can be carried out with ease. Second, the details of the database schema were explained along with the reasons for the chosen design. Finally, details of the key parts of the user interface were highlighted. This description of the interface included a brief explanation of the use of each component of the tool.

The following chapter takes the context of the preceding chapters and uses the EduCal tool described in this chapter, plus additional statistical analysis with derived data, to conduct an assessment of education quality using the multi-dimensional data cube and the time-space manifold that is central to the content of this thesis.

Chapter 5
Assessment of Education Quality with EduCal

The importance of education and the need for tools that assist in the management and planning of education systems was laid out at the onset of this thesis. While the intervening chapters described a data model and the implementation details of a st-DSS for this purpose, the chapter that follows applies the ideas, model and tool to assess the state of the education system in the department of Cusco. The analysis that follows takes the perspective of the regional and local management units of the Peruvian Ministry of Education and uses the regional education plan presented for Cusco in 2006 as its basis.

5.1 Background

The department of Cusco is located in the south-eastern part of Peru and covers an area of approximately 71,891 km² (Peisa, 2003). The landscape includes a large Andean (mountainous) region with several peaks reaching more than 6,000 meters above sea level as well as Amazonian (tropical) regions. A total of 1,171,503 people inhabit the region although almost one third live in the capital city of Cusco. The settlement of the region was influenced strongly by the Urubamba-Vilcanota and Apuríac rivers, which form the numerous valleys in which the majority of population centres can be found. This distribution of population has lead to an almost perfect divide between urban and rural areas (51.77%-48.23% respectively) (INEI, 2005). Notably, the male-female ratio is almost exactly one to one (INEI, 2005).

The largest urban settlement, the regional capital of Cusco, acts as the central hub for the economic activity in the department. The majority of the manufacturing industry (21% of the region's GDP (INEI, 2007)) takes places in Cusco. The *Cusqueña* brewery stands out as one of the largest manufacturers, although soft drinks, leather products, and fertilizers are also among the region's manufactured goods. However, the majority of the manufacturing sector is comprised of family businesses that operate with traditional methods of production and with low levels of capital. This trend towards small-scale production is also present in the agriculture sector (14% of GDP (INEI, 2007)) and, in both cases, prevents real competition with both national and international markets. This, in turn, leads to a low level of prosperity in the region.

In recent years the largest source of growth in employment and prosperity has been from the service industry. The service industry has grown to 25% of the region's GDP (INEI, 2007) and is comprised primarily of the financial, tourism and transportation sectors. The growth is spread throughout the region in the form of hotels, restaurants, travel agencies, private tour operators, etc. However, the city of Cusco itself has a large concentration of related employment opportunities since it is a prime tourist attraction. As the oldest settlement in the Americas, the city was recognized in 1978 as a World Heritage Site and inducted into the list of World Heritage Properties of UNESCO in 1983. The city is used also as a starting point for other tourist expeditions, such as the hike along the Inca Trail to the famous Machu Picchu site and trips to the ruins at Ollantaytambo and Sacsayhuaman.

The agriculture and mining sectors comprise 14% and 11% of the region's GDP respectively (INEI, 2007). Both are extremely influential in the social dynamics of the region for different reasons. The expansion of mining since the 1970s infringes on areas traditionally devoted to agriculture, severely affecting peasant communities. There have been continuous conflicts between the mining companies and the populations living in the affected areas (55% of the agricultural areas nationwide) (de Echave, 2005). Unlike the situation in the 1970s and 1980s, when the conflicts were primarily labour-related, the local rural communities increasingly are concerned over the control and management of their natural resources (de Echave, 2001). These conflicts affect a large portion of the population, most of which are employed in the agricultural sector.

Although agriculture does not comprise a large portion of the GDP, it remains the sector that employs the largest number of people (INEI, 2007). In contrast with the other sectors, agricultural production is traditionally organized around consumption or bartering. Another influential characteristic of the agricultural sector is that work units tend to be comprised primarily of families. The prevalence of family work units creates severe implications for the topic of this thesis, namely education, in two ways. First, it implies that children must, in the best of cases, share their time between family work-related expectations and school and, in the worst cases, they are forced by the family's economic situation to abandon school in favour of work. Second, a family work unit implies that entire families are reliant on the same income source, which is subject to fluctuations in commodity prices, weather, and availability and control of natural resources.

The characteristics of all industries result in a relatively small economy, with production and consumption primarily at the local scale, and with a high percentage of the population living in poverty. A commonly used indicator of poverty is the basic market basket, which is comprised of food and basic needs items. A staggering 75.3% of Cusco's population cannot afford the basic market basket (20.5% higher than the national average) (INEI, 2005). Moreover, 68.1% cannot even afford the basic food items in the basket. The effects of this degree of systemic poverty on childhood education and on the education system as a whole cannot be overstated. One specific consequence of poverty is malnutrition, which has a negative effect on education (Powell, Walker, Chang, & Grantham-McGregor, 1998; Gail et al., 2005; Alderman et al., 2006).

Numerous studies have also shown that the socio-economic characteristics of students and their families have an important influence on the likelihood that students enrol in school and that they participates in the classroom (Duraisamy et al., 1998; Kadzamira and Rose, 2003; UNESCO, 2004a). Hence, it is reasonable to assume that children who grow up in higher-income homes with encouragement from educated parents are likely to perform better than those living in homes with lower incomes and less educated parents who are unable to help due to lack of time or education. Unfortunately, the number of children growing up in homes with uneducated parents in the department of Cusco is relatively high. The literacy rate in the department is 81.32% (INEI, 2007), which is comprised of approximately two thirds of the rural population and one third urban residents. This statistic points to a large group of children whose chances of becoming literate are greatly diminished.

The differences between the urban and rural education contexts are not limited to literacy rates. Among other things, the case study presented in this chapter analyzes some of the differences found between urban and rural schools in order to provide valuable information for the formulation of a regional education plan (REP) for Cusco. The following section discusses the most recent regional education plan for the department.

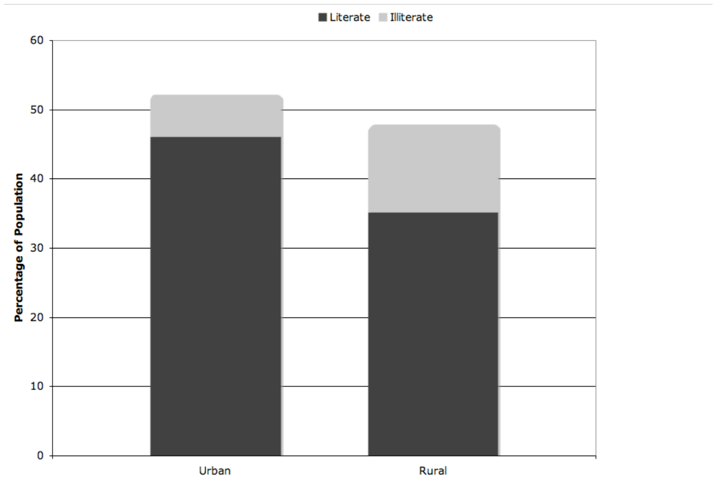


Figure 5.1 – Literacy Levels in Cusco

5.2 Regional Education Plan

In December 2006, the Participative Regional Education Council of Cusco (*Consejo Participativo Regional de Educación*, COPARE) formulated a thorough REP. The COPARE is organized and supervised by the Cusco DRE and comprises representatives from the following groups, namely teachers, university professors, the local education community, and public and private institutions (MINEDU, 2005). According to the official mandate, the COPARE is responsible for analyzing student access to education, factors that affect education quality, critical aspects of the functioning of the education system, and the available inventory of public and private resources in the region (MINEDU, 2005). The result of this thorough evaluation of the educational situation of the region is a document, the REP, which outlines both the current situation as well as concrete plans that should be carried out in order to improve education quality and delivery. Furthermore, the COPARE is also responsible for the monitoring and evaluation of the REP.

In the Cusco REP for 2006, the COPARE identifies a series of concrete problems present in the department of Cusco. Rather than attempt to summarize the entire contents of the REP, a few key points are highlighted here in conjunction with some of the informational limitations identified within the document.

It stands to reason that the document makes explicit mention of some of the ideas put forward in this thesis. For example, in the introductory statements of the REP, it reads:

> *The process at the regional level was developed based on the educational diagnostic that each province was able to carry out given the inputs they were able to compile.* (COPARE, 2006; p27, unofficial translation)

Clearly, the COPARE recognizes that the quality of the REP is directly proportional to the quality of informational inputs gathered by the individuals performing the assessment of the region. Thus, it can be suggested, perhaps optimistically, that the members of the COPARE are both keen and willing to use tools such as EduCal in order to enhance their knowledge of the educational circumstances present in Cusco.

Furthermore, the core concept of the importance geographic space in education quality assessment permeates throughout the document both implicitly and explicitly. In the discussion of virtually every indicator, the REP analyzes the differences between the urban and rural contexts. This emphasizes the importance placed by the COPARE in understanding the urban-rural dynamic of the region. Furthermore, the concern over the lack of geographic data is explicitly expressed:

> *The region does not have at its disposition geographic files or geo-referenced systems that can provide information regarding educational interventions in the different areas or, better yet, that can identify regions that require immediate attention or certain assistance.* (COPARE, 2006; p 100, unofficial translation)

This comment suggests that there is recognition of the importance of spatial data, on the part of the education planners, in the knowledge discovery process. At the same time, it highlights the emphasis placed on understanding regional differences within the department itself.

In addition, the REP identifies a series of issues and themes that are or particular importance for the department of Cusco. These key issues and themes are explored further and subsequently analyzed using EduCal, as discussed in the following section.

As mentioned previously, the COPARE clearly identifies the importance of the urban/rural split in the assessment of education provision. Within this theme, the demographic composition of the teacher population is highlighted as a cause for concern. For example, while the population is distributed nearly equally between the urban and rural areas, 68.7% of the teachers in the region work in urban settings. In addition, the REP identifies a lack of organization in the provision of culturally relevant teacher training, caused by a concentration of teacher-training centres in the capital city. The differences in education quality between the urban and rural contexts has lead to an out-migration from rural areas in an attempt to find education of higher quality (COPARE, 2006). It is not clear from the diagnostic presented in the REP whether the quality of education experienced by students in the urban areas is measurably better or whether this difference is simply one that is perceived subjectively to be better than in rural areas.

Another aspect of education that is highlighted in the REP is the high prevalence of students at each grade level that exceed the appropriate age (COPARE, 2006; p71). It is interesting that this measure of student advancement through the education system is included in the analysis, while promotion rates are not mentioned explicitly. However, it is clear from the REP that it is important to have some measure of student progress through the system as a means of assessing both the internal efficiency of the education system and rate of success of the students.

A final component of the education diagnostic carried out by the COPARE studied the levels of available physical infrastructure. The COPARE describes the lack of investment in infrastructure in recent years, for both physical structures and pedagogical tools, as "deplorable" (COPARE, 2006; 108). While the REP outlines the need to invest in order to provide the necessary conditions to optimize learning, the COPARE states that there is a lack of objective information of the infrastructure needs at the school level.

In summary, the REP developed by the COPARE in Cusco has identified a number of themes and a set of indicators that are considered to be of particular importance from their point of view. Of these, student promotion rates are recognized as an important measure of the success and efficiency of the education system from the point of view of the COPARE. For this measure, the appropriateness of the student's ages is identified as a valuable indicator because it provides some measure of the number of students who have repeated a grade multiple times or students who have simply entered the education system much later than they should have. Finally, the lack of investment in infrastructure in recent years is a cause for concern and a potential reason for inadequate education provision.

The following section takes the themes and indicators mentioned above and combine them with the limitations identified by the COPARE. The following sections present only one example of the type of analysis that can be carried out with the assistance of EduCal. In this case study, the analysis aims to evaluate the internal efficiency of the education system and the conditions that affect this efficiency. Additionally, an analysis of the differences between the urban and the rural areas in the department of Cusco is carried out, using the indicators identified in the REP, in a way in which the COPARE could not. Through this process, aided by the use of EduCal and statistical analysis, it is possible to identify areas that require immediate attention, as highlighted by the COPARE.

In order to achieve these objectives, the following tasks will be carried out:

i. An education quality index is created as a basis for comparison using EduCal.
ii. The distribution (both statistical and spatial) of the index is explored and described.
iii. A cluster analysis is performed in order to identify the relationship between explanatory variables and the education quality index using the statistical software Statistica 6.0.
iv. Using advanced spatial querying in EduCal, a comparison of the urban vs. rural conditions is explored.

5.3 Education Quality Analysis

To be able to compare regions and schools, it is necessary to create an education quality index that takes into account the perspective of the COPARE. There are several methods that can be employed in order to develop this index. For example, a method that has been employed previously with interesting results for the assessment of education quality is the target-based weighted averages (TBWA) method (Hall & Bowerman, 1995; Peters, 2002; Leahy, 2005; Engler 2007). As described in the previous chapter, a TBWA analysis tool is included in the EduCal toolbox (Figure 5.2). This allows for the creation of a wide range of indices that are responsive to various perspectives that allow individuals to establish which aspects of education they feel influence education quality, and also to determine the degree of this influence. With the TBWA analysis tool, an analyst selects which indicators to include, rates them in terms of their relative importance to each other, and finally specifies the minimum and maximum bounds of acceptable values and a target value for each indicator.

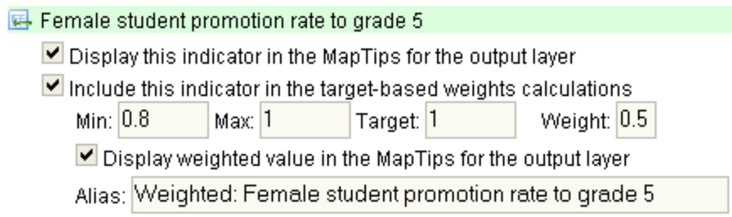


Figure 5.2 – TBWA analysis tool

The specification of minimum, maximum and target values for each indicator defines whether each indicator is a cost (target equals the lower bound), a benefit (target equals the upper bound), or neutral (target is between the minimum and the maximum bounds) (Peters, 2002). These three classifications for the indicators are used to create an additive education quality index, calculated so that higher scores subjectively correspond to a higher level of education quality. Thus, each indicator is first evaluated using the respective function described in Figure 5.3, which results in a TBWA quality score ranging between zero and the weight assigned to that indicator (w_i). If a school or district's indicator value falls within the acceptable bounds defined by the analyst, the score will increase (up to a maximum of w_i) in proportion to how closely it approximates the target value. As is depicted in the leftmost graph of Figure 5.3, if the value of a cost indicator falls below the minimum, the quality score assigned to that location will be equal to w_i while the quality score for locations whose indicator value is greater than the maximum will be zero. The inverse is true for benefit indicators. Finally, in the case of neutral indicators, all school or districts whose indicator value falls outside of the acceptable bounds will receive a quality score of zero.

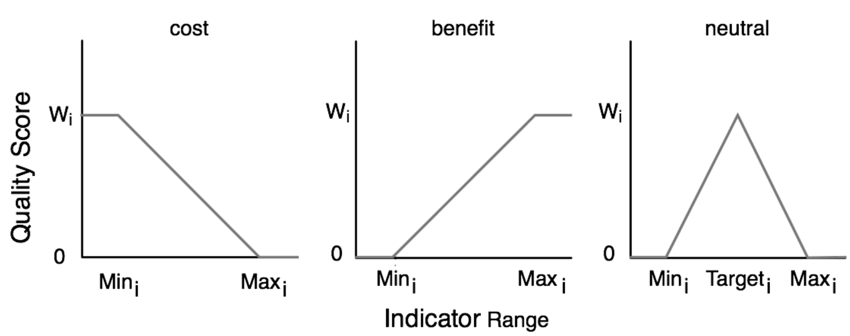


Figure 5.3 – TBWA analysis weight functions

The scores for each indicator at each location are summed to produce an overall index of education quality that reflects the importance of each indicator (by the use of weights) and the ideal values (by the use of a minimum, maximum, and target value) of the perspective taken for the analysis. Optionally, this indicator value can be divided by the sum of all of the weights and multiplied by 100 in order to convert the indicator to a percentage value. The percentage score is a reflection of the degree to which a selection (single or multiple schools, single or multiple districts, UGELs, DREs, provinces, departments etc) met the priorities of the analyst.

In this way, the TBWA method allows for normalization of the various scales and measures that compose an analyst's perspective of what defines an education of high quality. At the same time, this methodology allows analysts from a wide range of perspectives (MINEDU planners, to school administrators, through to parents groups) to capture accurately the degree to which each school or district has achieved their particular perspective of high quality education. Of course, the theoretical basis on which the definition of the indicators, and their respective minimum, maximum, target, and weight values is based on, determines the degree to which the resulting index values can be interpreted. In the analysis that follows, the indicators and values chosen are based on the interpretation of the REP outlined at the onset of this chapter.

As explained previously, the REP defines promotion rates as an important indicator of an efficient education system. In this sense, the COPARE defines education quality to be that which takes place in a system which efficiently advances students. Thus, a set of indicators was chosen from the library of available indicators in EduCal to capture this perspective adequately. Specifically, to conduct this analysis EduCal requires an indicator that explicitly measures promotion rates (broken down by both sex and grade) to be included as a benefit indicator.

Furthermore, the REP recognizes that the presence of a large number of students who are above the appropriate age for a given grade is a reflection of ongoing failure by the system to impart the skills necessary for student promotion or that students are not entering the school system at the mandated age. There exist indicators that measure the percentage of students who are two or three or more years above the appropriate age for each grade and sex, since student age is recorded as part of the school census. These indicators are included in the analysis as cost indicators.

Quality Index	Min	Max	Objective	Weight
Female student promotion rate to grade 5	0.8	1	1	0.5
Male student promotion rate to grade 5	0.8	1	1	0.5

Percent of female students 2 years older than the appropriate age in grade 4	0	25	0	0.25
Percent of female students 3 years or more older than the appropriate age in grade 4	0	25	0	0.25
Percent of male students 2 years older than the appropriate age in grade 4	0	25	0	0.25
Percent of male students 3 years or more older than the appropriate age in grade 4	0	25	0	0.25
Ratio of girls to boys in grade 4	0.9	1.1	1	0.5
Female student desertion rate in grade 4	0	0.1	0	0.5
Male student desertion rate in grade 4	0	0.1	0	0.5

Table 5.1 – Quality Index Indicators and TBWA parameters

While promotion rates measure the number of students who have successfully completed a grade, the additional inclusion of desertion rates as a cost indicator allows the index to reveal whether or not students are failing classes or abandoning school altogether. Finally, since the REP also identifies the importance of ensuring gender parity, the male to female ratio is also included, although with a lower overall weight, as a neutral indicator since concerns over gender parity are addressed by the inclusion of all indicators disaggregated by gender. Thus, differences between the sexes in any of the indicators will play a role in the resultant overall quality index value. Although no such recommendation was made in the REP, after verbal consultations with Alternativa, a Lima-based NGO, grade four was chosen as the basis on which to compose the education quality index. Section 5.9 discussed the reasons for the selection of grade four. The indicators chosen and their respective values are summarized in Table 5.1.

5.4 Descriptive Statistics of Quality Index

With EduCal, it is easy to classify and map the index at various levels of spatial aggregation by building the education quality index defined by Table 5.1. At first glance, the map of the education quality index at the DRE level shows that Cusco is in the lowest quintile of all DREs in Peru, but it is not the lowest (Figure 5.4). To understand better the situation in the Cusco DRE, the map can be zoomed in on Cusco and drilled down to be viewed at the UGEL level (Figure 5.5). This map reveals that there is a high degree of variability within the DRE of Cusco itself, even at the UGEL level. Three UGELs (80009, 80004, and 80012) have values of less than

10.2% of the targets chosen for the quality index. As noted earlier, the map produced in EduCal and shown in Figure 5.5 could be emailed to the directors of each of the three UGELs along with an inquiry on why they are underperforming in relation to the rest of the department, since the UGELs are sub-management units respectively of the MINEDU.

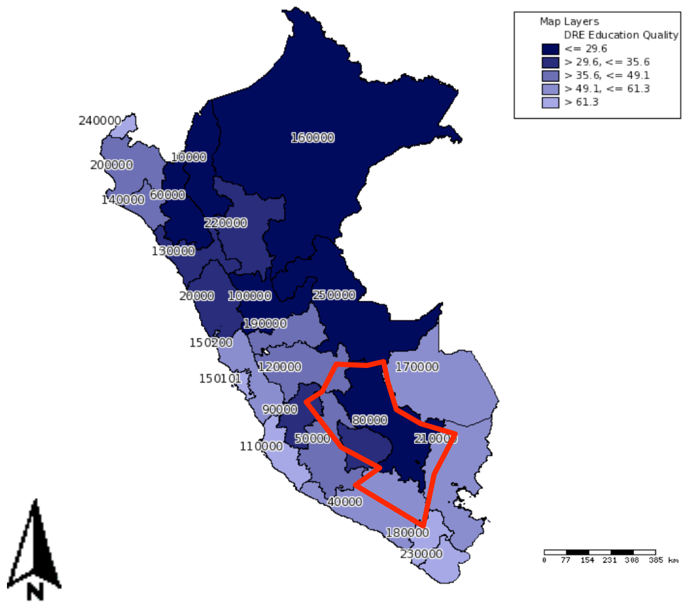


Figure 5.4 – Education Quality Index for all DREs in Peru

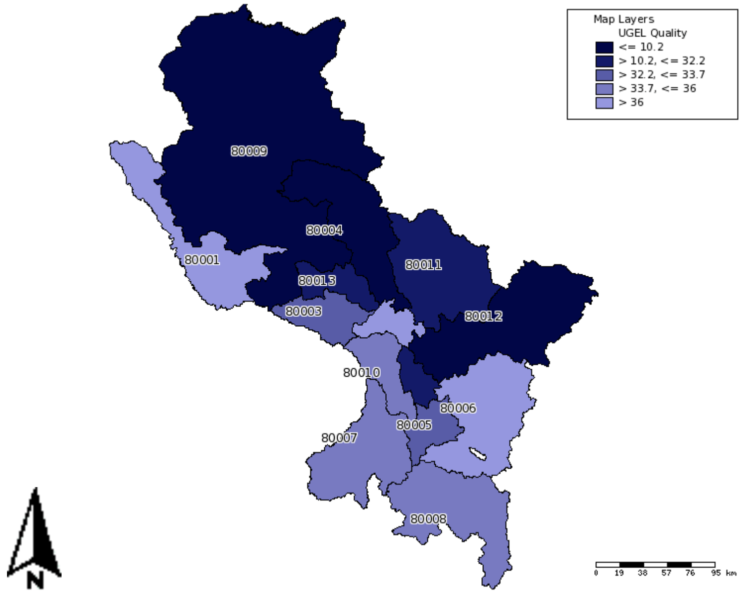


Figure 5.5 – Quality Index for all UGELs in the Department of Cusco

The creation and view of the education quality index at high levels of aggregation allows troubled areas to be identified easily. Based upon the discussion in earlier chapters, it is possible with the EduCal tool to continue exploring variations in the education quality index at more local levels to gain a better understanding of the root causes of the problems being observed. For example, the disaggregation of the index to the district level reveals two regions that are a cause for concern (shown in the north and the mid-east portions of Figure 5.6). Furthermore, basic statistics can be calculated in order to understand better the distribution of the index values. The basic statistics tool reveals that there are 108 districts in Cusco and whose quality index values range between 1.97% and 67.45%. The median for the group is 28.105 with a standard deviation of 15.107. Importing the district level values into the SPSS 15.0 statistical package allows for the creation of the histogram shown in Figure 5.7 that shows that the index is slightly right skewed in its distribution among the districts. The export of the district analysis layer also exports the quality score of each of the individual indicators, which can then be used to produce the set of histograms shown in Figure 5.8.

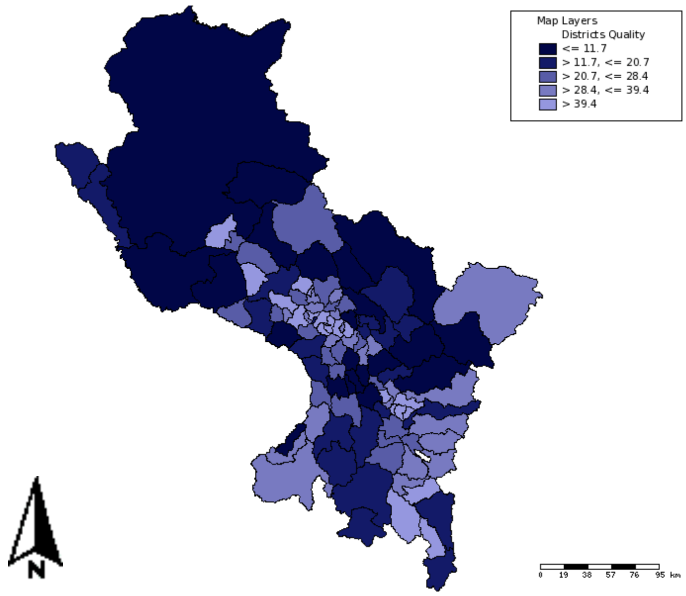


Figure 5.6 – Quality Index for all districts in the department of Cusco

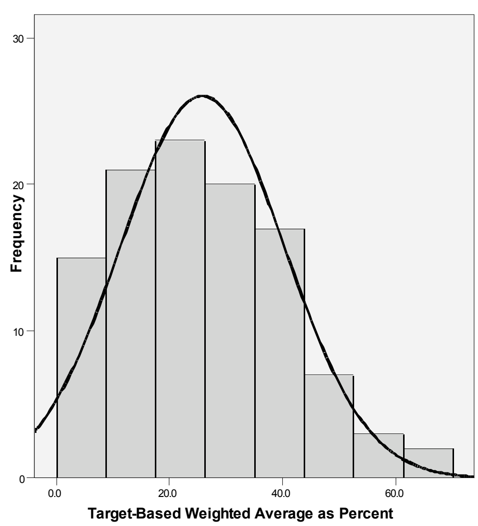


Figure 5.7 – Histogram of district-level TBWA values

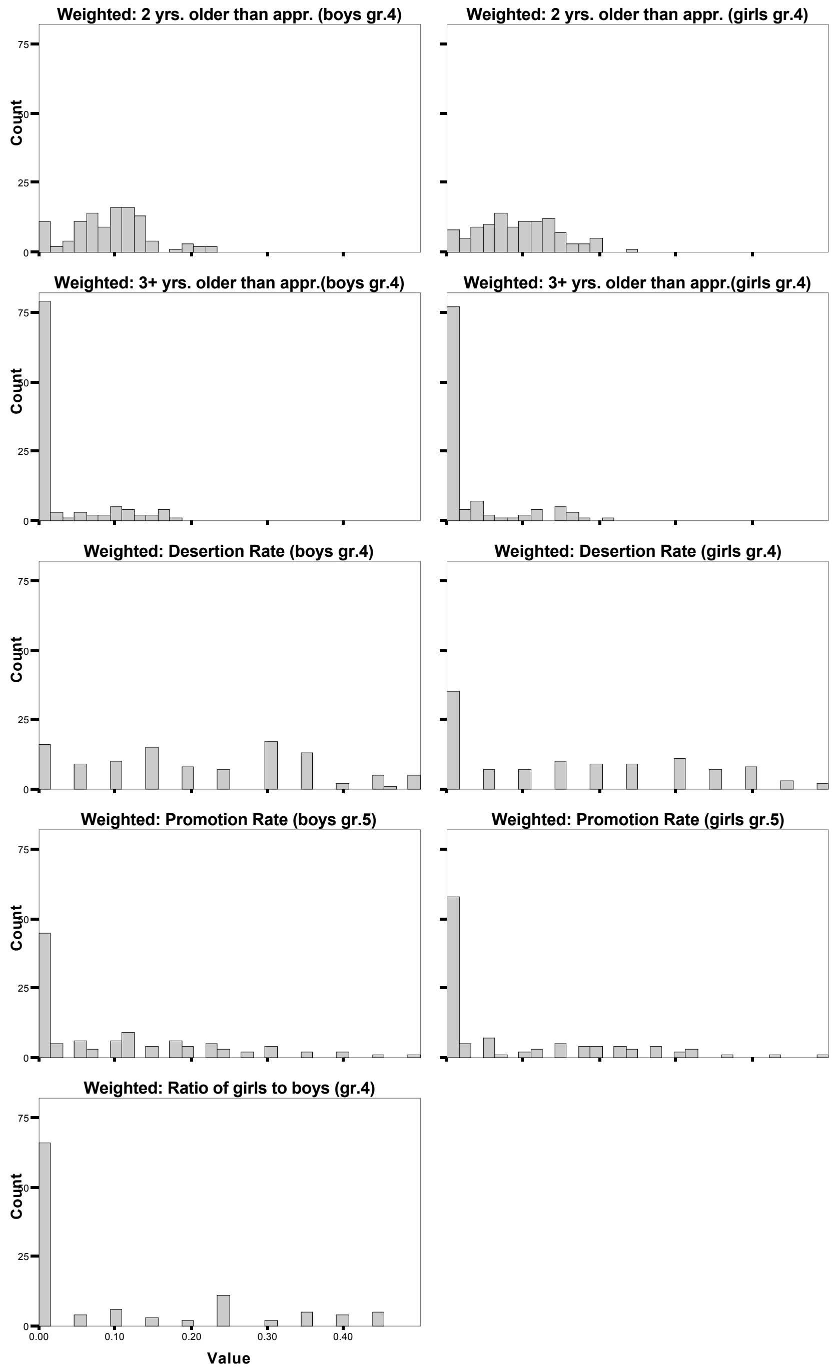


Figure 5.8 – Distribution of Weighted Values for Components of Quality Index

The histograms in Figure 5.8 explain which of the components used in the index contribute most to the maximum value of 65.3%. More than 50 districts exceed the minimum and maximum values for the 3 years or older than the appropriate age (both boys and girls), the ratio of girls to boys, and the promotion rate of girls. This observation is of particular relevance when taking into account that extra age was identified in the REP as the most important measure of poor promotion rates. Although these values could have been adjusted in order to allow for greater variability in the quality index, the *a priori* values were maintained since these values reflect the desired education outcomes for an education planner. By extension, a maximum value of 65.3% indicates that even in the best performing districts some schools are only meeting two thirds of the ideal values.

5.5 Index Validation

The same analysis can be performed by replicating the TBWA analysis for each primary grade. Using EduCal to add another five analysis layers, one for each grade, the quality indices for the first 5 grades of primary school were exported to comma delimited files. Grade 6 was not included in the analysis since the quality index used would require the promotion rate for children to secondary school and these data were not available. In addition, the five sets of values can be combined into a single data set using SPSS (or Excel) and a one way analysis of variance (ANOVA) can be used to determine whether there are statistically significant differences between their means. However, in order for an ANOVA to be valid, the five sets of values must be shown to have equal variances. Visual inspection of the box and whisker plot, shown in Figure 5.9, suggests that the grade 1 dataset has a different variance than the other grades, despite all five grades having seemingly equivalent means. This result can be confirmed using Levene's test for homogeneity of variance, which tests the null hypothesis that the variances are equal. The result of this test applied to the five datasets is a p-value of 0.000 (Levene-statistic=10.699; df1=4, df2=535). Thus, the null hypothesis of equal variances is rejected and it can be stated with greater than 99.99% confidence that there are significant differences between the variances of the five grades.

Since it appears, from visual inspection of Figure 5.9, that only grade 1 has a different variance, the test can be repeated for grades 2 through 5 to determine if the means found in those grades is statistically similar. Further analysis would be required to understand the reasons why the variance for grade 1 differs significantly, but this may be the result of a set of homogeneous characteristics amongst the population entering the first primary grade. The conditions appear to alter significantly by grade 2 and then remain fairly consistent for the remainder of primary school. A second Levene test that excludes the grade 1 data resulted in a p-value of 0.356 (Levene-statistic=1.083; df1=3, df2=428). Hence, the null hypothesis of homogeneity in variances cannot be rejected at the 95% (p=0.05) confidence interval. Given that homogeneity of variances exist, the results of a one-way ANOVA (presented in Table 5.2) are statistically valid. This result (p=0.336) indicates that the null hypothesis of difference of means cannot be rejected at the p=0.05 confidence interval.

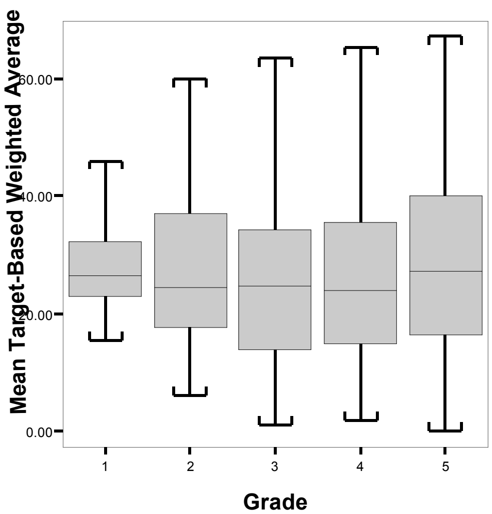


Figure 5.9 – Box and Whisker Plot of TBWA Mean TBWA values by grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	718.345	3	239.448	1.130	.336
Within Groups	90656.746	428	211.815		
Total	91375.091	431			

Table 5.2 – Summary of ANOVA for TBWA values by grade

This combination of results suggests that that the quality index based on grade 4 data can be used as a representative year for all of the other grades, with the exception of grades 1 and 6. This corroborates the opinions expressed by various UGEL members with which Alternativa staff have been working in Cusco (Santiago Kerrigan, Pers. Comm.). Moreover, there are significant policy implications for an education quality index where, in the best case, a district only meets 65% of the criteria laid out at the beginning of this chapter. This is especially significant because it appears that the issues raised by the COPARE appear to span all grades and to be of particular importance in terms of the overall quality of education of children in the department of Cusco.

Having established that the proposed education quality index is statistically representative for other grades, the next step in the analysis is to ascertain the causes of the low index values found for all grades. The following section explores the indicators chosen to assess this as well as their potential explanatory relevance.

5.6 Explanatory Variables

In the REP, the COPARE identifies infrastructure and the availability of trained teachers as two likely causes for poor promotion rates. Although there are a wide range of factors that affect education quality, the MINEDU has direct control over influencing those that fall under these two categories. Using these two criteria as the basis for the selection, a set of five indicators were identified from those available in the EduCal database. A list of indicators is itemized in the correlation matrix shown in Table 5.3.

While the quality index was comprised of data from the most recent year available (2004), the data for the explanatory variables are from 2002 because some of the indicators do not exist for 2004. However, this multi-year analysis should not be regarded as a limitation since, as noted earlier, education is a temporal process and a single snapshot in time cannot fully measure education quality. In other words, the conditions that existed in the school two years prior to the assessment year (i.e. in 2002) are considered to have influenced every student attending the school in the assessment year (2004). In fact, the students who have repeated a grade twice (e.g. students who are two years above the appropriate age) would have experienced the conditions present in 2002 when they first entered grade 4.

		Average number of libraries per school	Student-teacher ratio	Percent of all teachers that have titles	Student-class section ratio for grade 4	Shortage of seats as percent of total
Average number of libraries per school	Pearson Correlation	1				
	Sig. (2-tailed)					
Student-teacher ratio	Pearson Correlation	-.002	1			
	Sig. (2-tailed)	.988				
Percent of all teachers that have titles	Pearson Correlation	.212(*)	.103	1		
	Sig. (2-tailed)	.028	.287			
Student-class section ratio for grade 4	Pearson Correlation	.419(**)	.250(**)	.217(*)	1	
	Sig. (2-tailed)	.000	.009	.024		
Shortage of seats as percent of total	Pearson Correlation	.244(*)	.133	.017	.080	1
	Sig. (2-tailed)	.011	.171	.858	.412	

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 5.3 – Correlation Table for Explanatory Variables

A linear regression model can be created to determine the relationship between the education quality index composed previously and this set of potential explanatory variables. However, the statistical validity of a linear regression model is relative to the orthogonality of the independent variables. As can be seen from Table 5.3, there are statistically significant correlations between the five chosen indicators and thus only a subset of mutually orthogonal independent variables are considered for the model (percent of teachers that have titles, student teacher ratio, and shortage of seats as percent of total). To build the model, *step-wise regression* was used through the addition and removal of independent variables in an iterative fashion. This process requires a minimum statistical significance with which to reject or accept each variable (in this case the 95% confidence level, p=0.05 to enter and 90% confidence level, p=0.10 to exit). Table 5.4 shows the results of the complete step-wise regression at each iteration.

Model	Variable Entered	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	Student-teacher ratio	.287(a)	.082	.074	13.9271
2	Percent of all teachers that have titles	.391(b)	.153	.137	13.4446

a Predictors: (Constant), Student-teacher ratio
b Predictors: (Constant), Student-teacher ratio, Percent of all teachers that have titles

Table 5.4 – Regression models produced by a step-wise regression

The result of the step-wise regression produced a maximum adjusted R-squared of only .137 and with a standard error of 13.44. These values indicate that the model has very limited predictive utility and thus any relationship between the education quality index and the set of explanatory variables is either not evident or non-linear in the effect of the independent variables on education quality. Although it is possible that the R-squared value influenced by the potential spatial factors, the complete lack of non-spatial correlation suggests that a geographically weighted regression would not yield a robust predictive model. This seems contrary to the ideas put forth by the group of professionals that comprise the COPARE.

An alternative approach is, instead of a single linear combination of potential explanatory variables, to use a combination of conditions (or various different combinations) to explain better the dependent variable. This relationship can be examined by performing a cluster analysis. Cluster analysis (a term first used by Tryon, 1939) is an exploratory analysis technique used to organize data into meaningful structures (i.e. to determine taxonomies) (Zar, 1984). Specifically, using cluster analysis it is possible to identify groups (clusters) of districts that possess similar characteristics. The clusters can then be compared to the dependent variable in order to determine if the conditions that define each cluster have a positive or negative impact on the dependent variable, which is, in this case, the education quality index.

5.7 Cluster Analysis

There are various clustering algorithms, such as joining (tree clustering), two-way joining, and K-means that can be used to identify clustered of observations on a specific variable. For the purposes of this analysis, the tree clustering algorithm was chosen because it forms successively larger clusters using some measure of similarity or distance. These distances are based on either a single dimension or multiple dimensions. At the first step of the process, when each object represents its own cluster, the distances between the objects is defined by a chosen distance measure. However, once several objects have been linked together, linkage rules (also known as amalgamation rules) are used to determine when two clusters are sufficiently similar to be linked together.

In the case presented in this chapter, the objective of the cluster analysis is to identify groups of districts that differentiate themselves from the rest by at least one of the measures. That is to say, the clusters formed should have a combination of the input variables that is different from all the other groups in at least one variable. The objective is to identify which combination of explanatory variables are highly correlated to high values in the education quality index.

In order to achieve these properties in the groups, a Chebychev distance was chosen, as it defines two objects (cases) as "different" if they are different among any one of the variables (Zar, 1984). Equation 5.1 describes this definition. In this particular case, other typical measures of distance, such as the Euclidian distance, may have lead to deriving groups that did not possess distinguishing characteristics because they measure distances based on the linear combination of all the dimensions and not on a dimension-by-dimension basis.

$$\text{distance}(x,y) = \text{maximum}|x_i - y_i| \qquad \text{(Equation 5.1)}$$

The second defining characteristic of a cluster analysis is the linkage rule. Several methods can be used to link clusters together. For example, clusters can be linked when any two objects in each cluster are closer together than their respective linkage distance. Put another way, the nearest neighbour of any two clusters is used to determine the distance between clusters. This method is called single linkage. However, in order to try to minimize variance within each group and maximize it between groups, a method known as Ward's method is used. Ward's method (Ward, 1963) uses ANOVA to minimize the sum of squares between two hypothetical clusters at each step.

The result of the cluster analysis using Ward's method and the Chebychev distance metric on the 108 districts in Cusco and the five explanatory variables mentioned earlier is shown in the dendogram in Figure 5.10. This diagram was produced using the software package Statistica v.6. Every step in the algorithm creates larger clusters such that every cluster is entirely contained by the cluster to the right of it. Typically, a cut-off point is chosen, at a point where the analyst feels the clusters are sufficiently large enough to yield significant results. A cut-off point of 80 was chosen by looking at the dendogram in Figure 5.10. This point demarcates five distinct groups of districts.

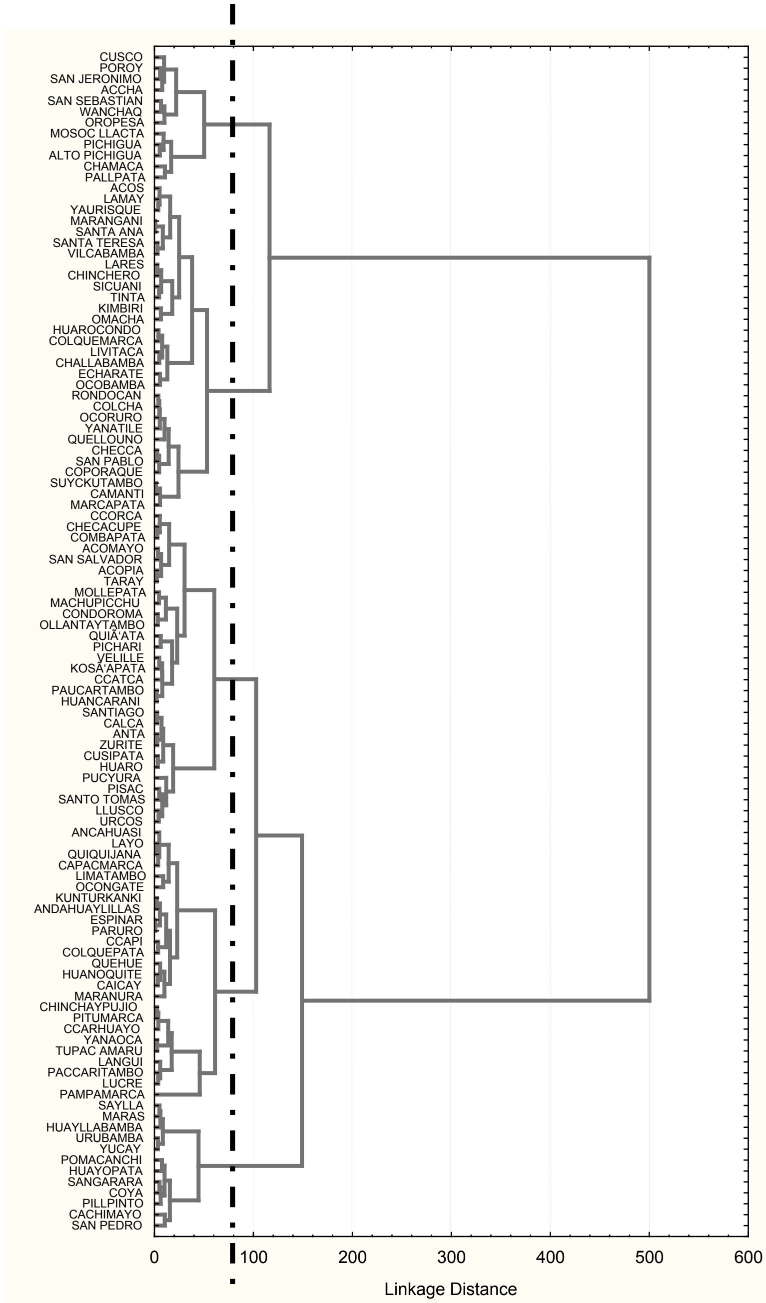


Figure 5.10 – Dendogram for Cusco districts using Ward's method and Chebychev distances

The members of each of the groups are identified by studying the amalgamation schedule produced by Statistica and a cluster number can subsequently be assigned to each district. In order to visualize the clusters easily, a shapefile of the districts was exported from EduCal and the cluster number was appended to the DBF file containing the attribute data. The advantage of exporting the data to a geographic format is that the clusters can be mapped using any GIS package (in the case of Figure 5.11 it was qGIS, http://qgis.org). As shown in Figure 5.11, there does not appear to be any clear spatial pattern in the distribution of the clusters comprising of between 14 and 30 districts each.

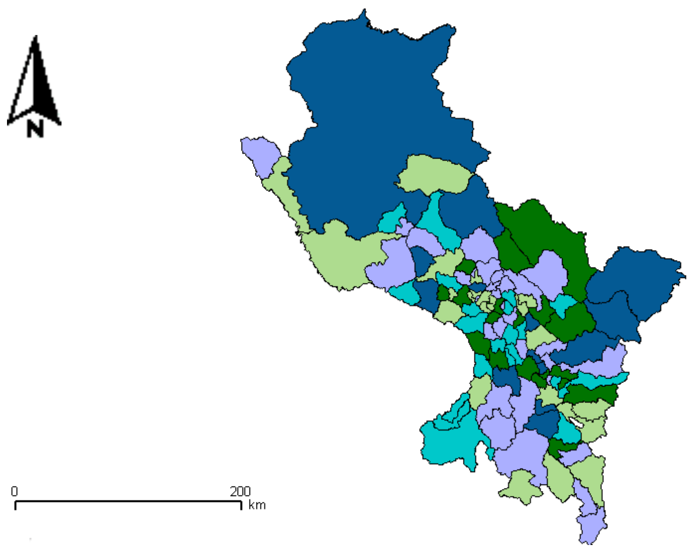


Figure 5.11 – Five Clusters defined at linkage distance of 80

The overall averages of each of the explanatory variables and of the TBWA quality index can be viewed by cluster in order to understand better the conditions that exist within each cluster. The Chebychev distance metric and Ward's method for performing linkages guarantee that each cluster differs in at least one of the five variables that comprise the cluster. Table 5.5 summarizes these averages, as well as the TBWA index and the totals for each cluster so that the identifying characteristics of each cluster can be discerned.

	Cluster					
Data	**1**	**2**	**3**	**4**	**5**	**Grand Total**
Average of Average number of libraries per school	0.063	0.037	0.048	0.036	0.128	0.060
Average of Student-teacher ratio	28.617	31.667	33.838	30.511	31.161	31.421
Average of Percent of all teachers that have titles	95.567	83.655	95.673	90.330	94.355	91.075

Average of Student-class section ratio for grade 4	20.984	18.196	16.700	13.250	22.826	18.223
Average of Shortage of seats as percent of total	12.821	29.566	27.945	42.989	52.882	34.070
Average of TBWA as a Percent	32.389	25.095	22.219	18.935	32.530	25.733

Table 5.5 – Averages of explanatory variables and quality index per cluster

Qualitatively, the values of Table 5.5 show that the clusters have the following characteristics:

Cluster 1
- Lowest shortage of seats as percent of total

Cluster 2
- Lowest percent of all teachers that have titles
- 50% lower than the average of the average number of libraries per school

Cluster 3
- Highest percent of all teachers that have titles
- Slightly below average shortage of seats as percent of total

Cluster 4
- 50% lower than the average of the average number of libraries per school
- Lowest student-class section ratio
- High shortage of seats

Cluster 5
- More than double the average number of libraries per school
- Highest shortage of seats as percent of total

The average index value also varies between clusters. It is possible to determine, by the use of an ANOVA if the differences in this index value are significantly different to establish a statistical basis for defining conditions that influence education quality, as measured by the proposed index. Visually, differences in means and similarities in variances can be detected using the box and whisker plot shown in Figure 5.12. These observations can then be corroborated with a Levene's test of homogeneity of variance and an ANOVA. The Levene's test shows that the null hypothesis of equal variances cannot be rejected (Levene-statistic=1.525; df1=4, df2=103; p=.200) and thus the results of the ANOVA summarized in Table 5.6 can be considered statistically valid. These results show that the null hypothesis that the means of the five clusters are equal can be rejected beyond the 95% confidence interval. Thus, the conditions found in each of the clusters have lead to the conclusion that a statistically significant differences in the education quality index values.

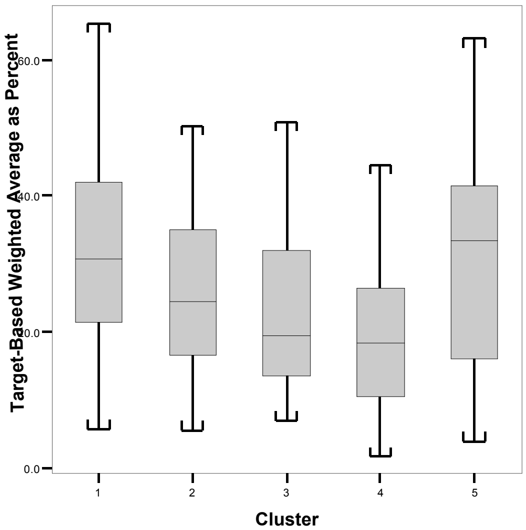


Figure 5.12 – Box and whisker plot for TBWA values per cluster

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2810.791	4	702.698	3.694	.007
Within Groups	19592.495	103	190.218		
Total	22403.286	107			

Table 5.6 – Summary of ANOVA for TBWA values by Cluster

Post-hoc ANOVA tests can be run once difference in means have been established in order to identify homogeneous subsets. That is, the *post-hoc* tests can be used identify two subsets of clusters whose means are homogeneous within the group, but differ between each other. One such test is the Tukey-test (also known as "honestly significant difference test" or "wholly significant difference test"), which is used to test the null hypothesis that every two pairs of means are equal ($H_0=\mu_1=\mu_2$). A second test, known as the Newman-Keuls (or Student Newman-Keuls), is less conservative, but tests the same null hypothesis. Some authors believe that the Tukey-test is unnecessarily conservative (for example, Miller, 1966; pp88), while others think the Student Newman-Keuls test is not strict enough (Einot and Gabriel, 1975; Ramsey, 1978). The results of both tests are presented in Table 5.7. Both tests generate the same two homogeneous subsets (although with different levels of significance). Hence, it can be stated with at least 95% confidence (p=0.05) that cluster 4 is significantly different from clusters 1 and 5, while clusters 2 and 3 cannot be distinguished from any other cluster.

Test	Cluster	N	Subset for alpha = .05	
			1	2
Student-Newman-Keuls	4	20	18.935	
	3	23	22.219	22.219
	2	30	25.095	25.095
	1	14		32.389
	5	21		32.530
	Sig.		.332	.087
Tukey HSD	4	20	18.935	
	3	23	22.219	22.219
	2	30	25.095	25.095
	1	14		32.389
	5	21		32.530
	Sig.		.614	.128

Table 5.7 – Summary of Post-hoc tests

The implications of these results are that the distinguishing characteristics of cluster 4 lead to statistically significant lower values in the education quality index. Cluster 4 has, on average, approximately 50% less libraries per school, the lowest student-class section ratio, and a very high shortage of seats as a percent of the total number of seats. This cluster information can now be used as inputs into the cycle of knowledge discovery provided by EduCal.

5.8 Results Integration with EduCal

There are two ways to use the cluster information within EduCal. One approach is to input the cluster number as a variable into EduCal and further analyze the cluster characteristics and spatial layout of the clusters. As an alternative, the EduCal Query Builder can be used to identify all of the districts that score poorly in all three characteristics. This approach identifies a set of districts where it is beneficial to invest additional resources with a greater likelihood of increasing education quality as defined by the index. A query using the cluster mean values for each of the three indicators as the cut-off point is depicted in Figure 5.13 below and the result is shown in Figure 5.14. This cut-off and the use of AND join condition is very restrictive and will only yield the districts with the lowest values (Figure 5.14). However, similar queries can be carried out with the use of the Query Builder in order to explore both the locations of the clusters as well as their characteristics.

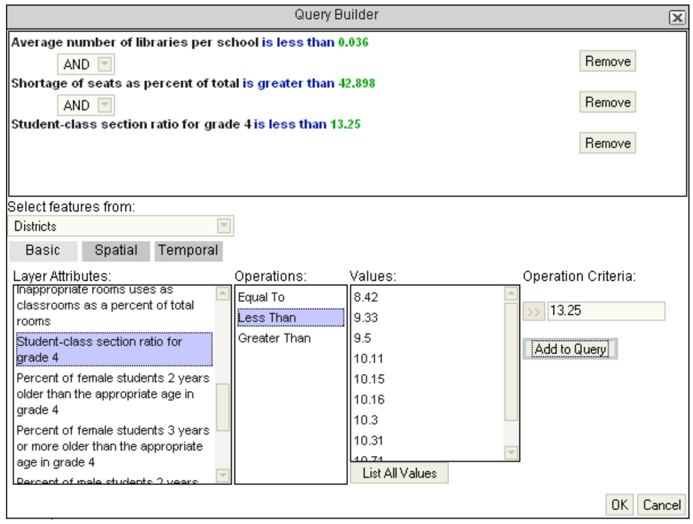


Figure 5.13 – Query Selecting the Characteristics of Cluster 4

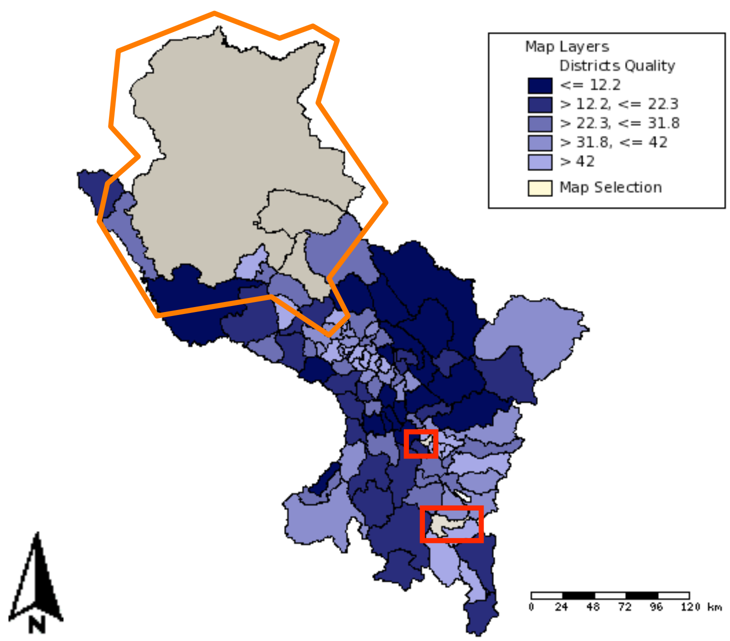


Figure 5.14 – Query Result for Cluster 4 characteristics

The basic statistics tool for the current map selection shows an average TBWA index value of 18.874%. It is possible to sub-select features from the current selection to identify those which fall below the average. This query separates the results into the two higher value districts (marked in red in Figure 5.14) and the three districts in the north of Cusco (marked in brown in Figure 5.14). The average TBWA index value for these three districts is 8.427%. It is notable that the three districts identified by this analysis are all adjacent to each other. The significance of this spatial pattern can be evaluated by examining the surrounding districts in two ways. First, all of the adjacent districts can be selected using the query builder (Figure 5.15). The selection can be exported and a t-test performed to compare the mean TBWA value of the three selected districts to those districts that are directly adjacent to them. Second, the drill selection tool can be used to drill up to the UGELs that contain the three districts and subsequently to drill down to all the districts within these two UGELs (080009 and 080004). The resulting selection can then be exported t-tests can be performed to compare the mean TBWA values of the three districts in comparison to the others in the two UGELs.

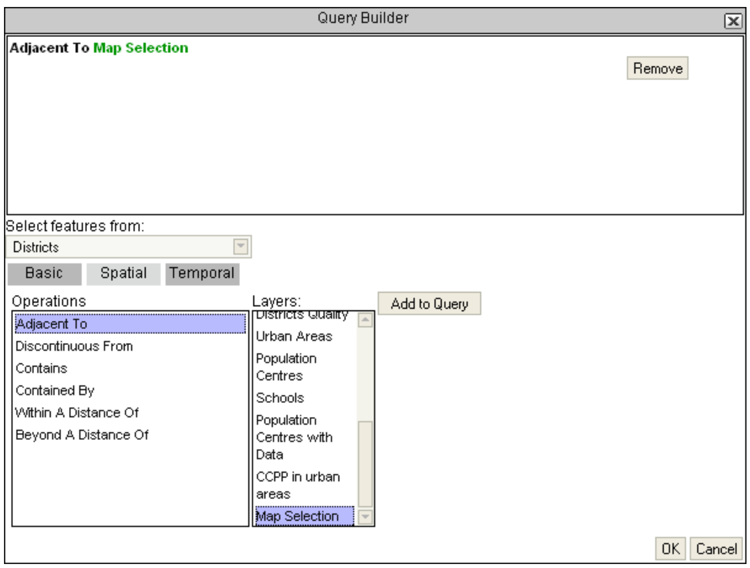


Figure 5.15 – Query Builder for Adjancent Districts

Both t-tests result in a p-value less than 0.05, which indicates that the null hypothesis that the mean value for the three selected districts is not the same as the mean value of the districts around it can be rejected with greater than 95% confidence. The results of the t-tests are summarized in Table 5.8. It can be concluded from both of these tests that these three districts in the north of the Department are significantly underperforming compared to the rest of the department, their neighbouring districts, and the districts in their UGELs. A final query using the query builder (Figure 5.16) reveals that all three districts have decreased in education quality in relation to the calculated education quality index. This indicates further that there is a need for investment in these three districts.

	N	Mean	Std. Dev.	Sig. (1-tailed)	Mean Difference
Selection	3	8.4267	3.37793		
Adjacent Districts	8	22.4338	12.59070	.0491	-14.00708
Districts in same UGELs	9	25.4156	11.48217	.017	-16.98889

Table 5.8 – Summary of t-tests for neighbouring districts

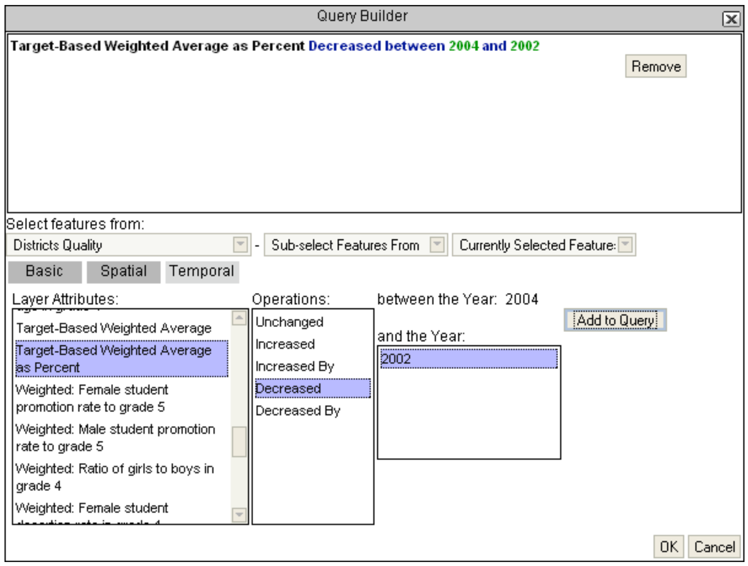


Figure 5.16 – Query Builder for temporal query

According to the analysis carried out previously, investing in any one of the three characteristics that define cluster 4 would likely improve the education quality index in this region of the department of Cusco. In this case, it is recommended that investment should be focused on the district level, since the three selected districts are statistically below the mean of all other districts in their respective UGELs. Further drilling and querying would identify those population centres within the three districts that have lowered the education quality index average, in the case that strategies targeting population centres were to be used. However, it is also possible to draw more general conclusions of the pattern of education quality by examining more aggregate trends within the department. To illustrate this, the following section presents an analysis of the differences between the urban and rural population centres found within the department as part of the analysis at the population centre level.

5.9 Analysis of Urban versus Rural Situations

It is interesting to note that the only two urban population centres in the three problem districts have education quality index values in the lowest two quintiles. Initially, this can be identified by turning on the population centre and urban area layers within EduCal and visually inspecting the map. However, in order to conduct a more comprehensive study of the urban/rural variation in the department of Cusco, a few additional steps must be taken to separate the population centre points into two distinct groups. For example, a new selection can be made with the Query Builder to identify all of the points within a five-kilometre radius of an urban area and with positive promotion rate data (Figure 5.17).

The *greater than 0* clauses eliminate the population centres with null values, which could indicate cases of 'not applicable', 'no data' or 'zero'. Although there is no documentation explaining the use of null values in the database, the large number of null values found in population centre points is likely a result of schools that have not reported data to the MINEDU. While EduCal gracefully handles null values, the use of the *greater than* 0 query clauses results in more manageable layers (with less points). It is important to note that these null values might not be missing at random, and may be a result of a particular group of schools that is reporting with less regularity than average. A separate query could be used to identify all the schools with null values in order to attempt to identify if values are missing at random or not.

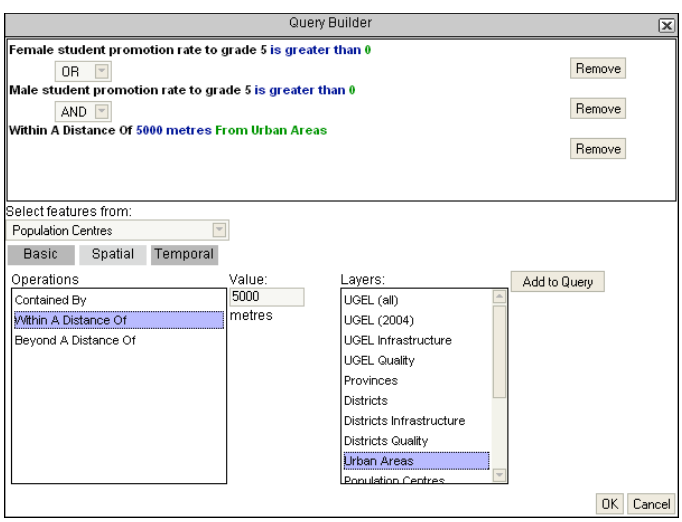


Figure 5.17 – Query Builder to select urban areas with positive promotion rate data

A new layer can be added to the set of all urban population centres with data selected on the map using the *Add Layer* tool's Extent option (Figure 5.18). This results in a new layer (CCPP urban) of all of the population centre points within five linear kilometres of the urban area polygons. Similarly, another new layer can be added after inverting the selection using the query in Figure 5.19, or alternatively by making a new selection using the *Beyond A Distance Of* operator. Both approaches lead to two layers (CCPP urban and CCPP rural) that can be used as the basis for two new TBWA layers using the index defined previously in this chapter.

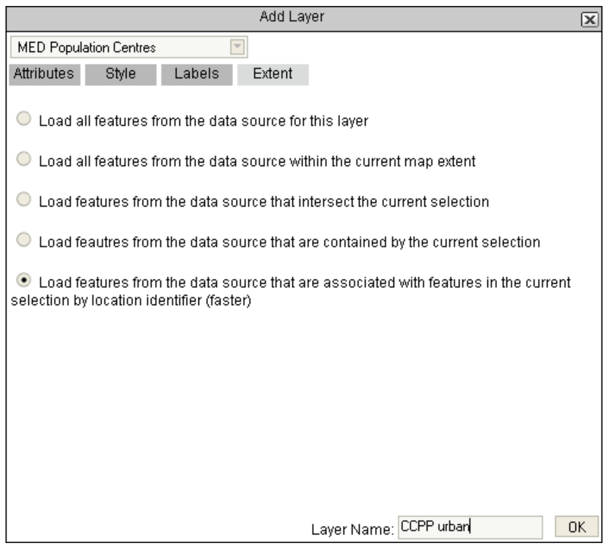


Figure 5.18 – Add Layer tool applied to map selection

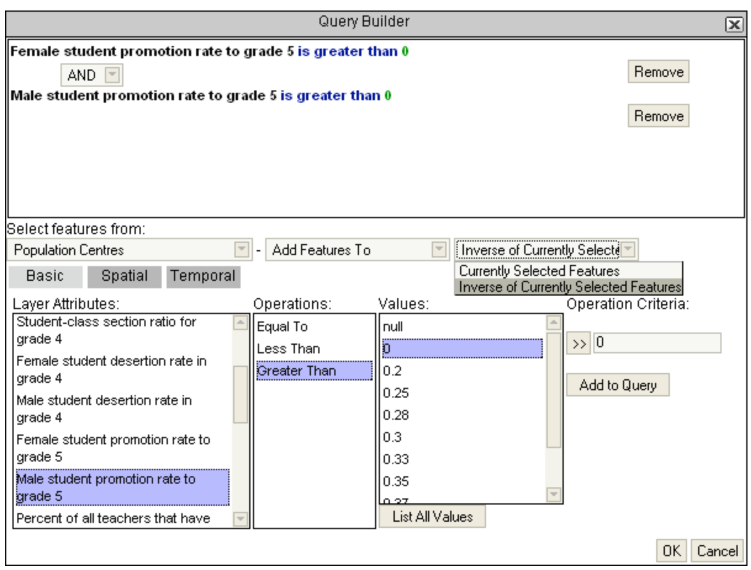


Figure 5.19 – Query Builder to select rural areas

Both analysis layers can be exported from EduCal and a t-test performed to verify if there is a significant difference between the urban and rural TBWA values. A summary of the TBWA values for the urban and rural sets is shown in Table 5.9. This reveals that the null hypothesis of no difference in means cannot be rejected beyond the 95% confidence interval (t=-1.838, p=0.068, df=136.689). Therefore, it would appear that there is no significant difference between the urban and rural areas in relation to the education quality index.

	N	Mean
Rural	637	38.4995
Urban	92	41.6649

Table 5.9 – Mean values of Index at Urban and Rural Population Centres

This result may seem surprising given the concerns expressed in the REP over the differences between urban and rural areas. However, this lack of statistically significant differences between the two groups shows that this facet of relatively poor education quality exists in both contexts. This does not imply that there are no measurable differences whatsoever between the two groups. It appears that students are being promoted indiscriminately throughout the department, but not necessarily while receiving education of equivalent quality. This reinforces the concerns expressed by the EFA Monitoring Report 2005 (UNESCO, 2005), which emphasized that quality should be examined as well as quantity.

It is also possible to identify changes in the patterns of student promotion with EduCal. With the temporal query builder, population centres can be identified where student promotion rates have either increased or decreased from 2002 to 2004. Figure 5.20 shows two layers added on top of the district level map shown previously. At first glance, there does not appear to be a clear spatial pattern, although it does appear that there is a slightly higher concentration of increased (green) points near the centre (where the city of Cusco is located).

Utilizing the urban and rural layers created earlier, the urban-rural divide in education quality can be superimposed on the temporal layers. These results are shown in Figure 5.21, zoomed in to the area surrounding the city of Cusco. The creation of this type of consistency map in EduCal requires that four layers be added (two increased/decreased layers and two urban/rural TBWA analysis layers), but the results allow for a single image to display all of the possible scenarios of urban/rural, high values/low values, and increased/decreased values. The small study area shown in Figure 5.21 has examples of all eight possibilities.

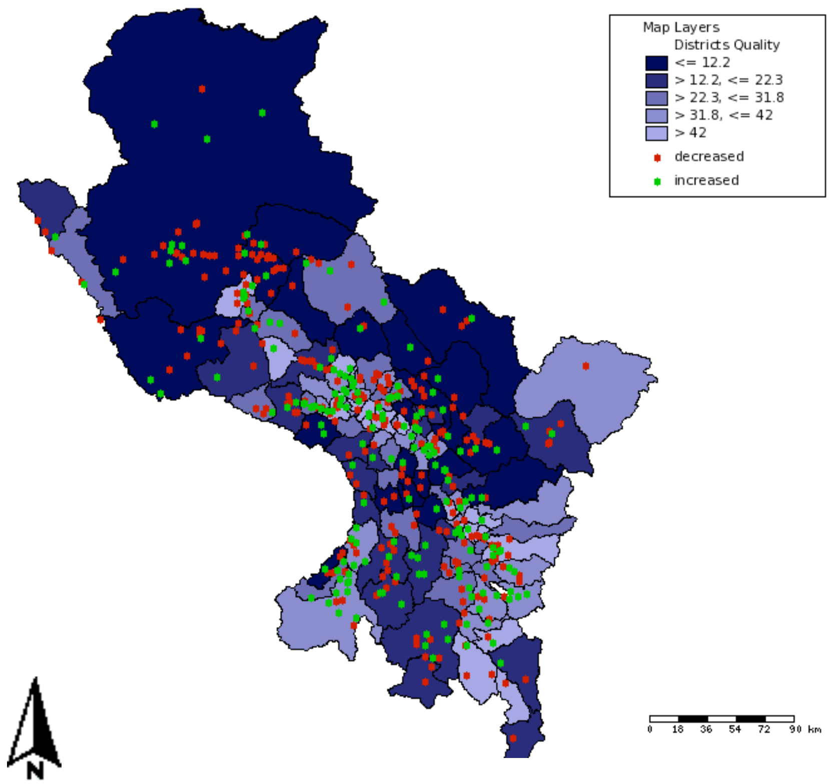


Figure 5.20 – Changes in Quality between 2002 and 2004 for Cusco

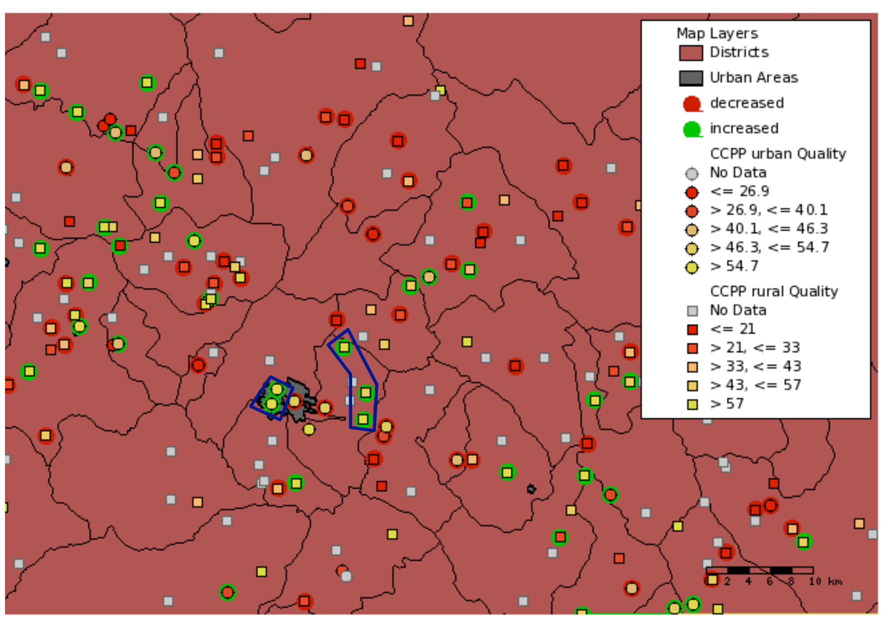


Figure 5.21 – Consistency Map of Changes between 2002 and 2004 near Cusco

Within the Cusco city boundaries, all four population centre points have high index values, indicating that the promotion rates in the city are high and desertion rates are low. The next five nearest centres to the urban area (with data) are also in the highest quintile (shown in blue). This suggests that the city of Cusco, by virtue of being the largest urban area and the departmental capital, has different conditions than the other urban areas. However, also notable in the city of Cusco is that the two easternmost population centres have decreased in values and the two westernmost have increased in values. Other urban areas scattered throughout this region of the department show an irregular distribution of high and low quality index values while having increases and decreased over the previous two years. Thus, on a broad scale, it cannot be said that promotion problems appear to improve or worsen with regards to the degree of rurality of a population centre. However, inspection at local scale can identify those centres in special need of attention in both urban and a rural contexts.

5.10 Summary

This chapter has shown how the EduCal tool can be used to perform an education quality assessment of a study region in Peru using the concept of space-time and principles of SOLAP-based data mining. The analysis presented drew upon the Regional Education Plan (REP) for the department of Cusco proposed by the Peruvian educational advisory committee known as the COPARE. At the start of the chapter, the socio-economic context of Cusco was discussed to establish a sense of the context of the study area. Subsequently, the primary findings of the REP, along with the limitations found in its preparation by the COPARE, were described. The target-based weighted average technique was used to create an education quality index using the indicators identified as a cause for concern within the REP.

The following sections used a combination of EduCal and statistical analysis software to analyze the distribution of this index and to identify the more problematic areas in the department of Cusco. To begin, the index was examined at every level of spatial aggregation and studied statistically. This index was then compared against a group of indicators, also identified in the REP, as possible explanatory variables. Using cluster analysis, the Cusco districts were then divided into five groups. These groups, each with distinct characteristics, were tested for statistical differences in mean quality index values and one of the groups was found to have statistically lower values than the rest. Subsequently, the EduCal query builder was used to select the districts with characteristics similar to this group and three problem districts in the north of the department were identified. Finally, using advanced spatial and temporal querying within EduCal, the urban and rural areas of the central region surrounding the city of Cusco were examined through the use of consistency maps.

Overall, this chapter showed how the EduCal tool satisfies the analysis needs identified by the group in charge of the preparation of the REP in Cusco. The analysis was aimed at answering the questions left unanswered in the REP while, at the same time, highlighting the spatio-temporal decision support capabilities of the software.

Chapter 6
Conclusions

The unprecedented growth of the Internet, both in its global spread and use, is a testament to the depth to which ICTs have become entrenched in modern societies. The growth of ICTs over the last ten or fifteen years has played a pivotal role in the development of these societies in a wide variety of ways and, although the changes that have occurred have not been caused by the technologies themselves, these changes would not have been possible without the existence of foundation ICTs. The propagation of the Network Society paradigm (Castells, 2000) has lead to the incorporation of ICTs into daily lives and organizational functions, which has resulted in the largest collection of data and information resources to date in the history of humanity. However, despite the promises of improved efficiency and of other benefits from the increased collection of data, there continue to exist problem domains in which these promises have gone substantially unfulfilled. In particular, problem domains that involve both the spatial and the temporal dimensions, such as education planning and education quality assessment, fall within this category.

In the context of education, there continues to be a shortage of appropriate tools that can turn collected data into valuable information and *knowledge*. The emphasis on knowledge, as opposed to data, is an important distinction for all of the processes that inform decision-making, given that planners and decision makers are constrained by the quantity and quality of *useful* information that is available to them. In this sense, it is evident that there is a need for decision support systems (DSS) that can facilitate the extraction of spatio-temporal knowledge from large databases. Therefore, the tools currently known as DSS or spatial DSS must be extended to include a new class of tools, defined here as spatio-temporal decision support systems (st-DSS).

As stated above, this need is particularly true when the decision-making context revolves around education. Education is both a spatial and a temporal process. Education systems have schools distributed throughout geographic areas and typically are organized around hierarchical organizational and administrative structures that can be mapped onto nested spatial boundaries. Such is the case for the Peruvian education system, which was used as a case study in this thesis. At the same time, education itself is a process that takes place over time. That is, the education received by a student is a direct result of the cumulative learning experiences obtained both inside and outside of the education system. From this point of view, any assessment of an education system would be incomplete without the inclusion of *both* the spatial and temporal dimensions as part of the analysis. Thus, it is essential that ICT tools be designed to include both dimensions. For this purpose, the unified concept of space-time can be used as a guide in the design of st-DSS to ensure that an analyst can access both dimensions *simultaneously*. This spatio-temporal nature of education suggests that this new class of tools are extremely appropriate.

At the same time, it is also true that the need for the development of such tools is urgent, especially if the goals proposed by the Education For All (EFA) initiative are to be fulfilled. This is particularly true in light of the EFA Monitoring Report for 2005, which emphasized that to reach the EFA goals, it is imperative that the education received by all be of good *quality*. To measure the progress towards reaching these goals at national and sub-national levels, as is suggested by the Dakar Framework (DF), there need to be tools that allow for the assessment of education quality. Given that the original millennium goals proposed at the first World Conference on Education for All were not achieved, there is an urgency in improving the global state of education in order to not miss the target year proposed in the DF for attaining the EFA goals by 2015.

Since the concept of education quality does not have a single interpretation, the tools designed must be dynamic and complete enough to allow for the evaluation of quality from a broad range of perspectives. These perspectives must include those members of society who are in charge of making decisions regarding the education system at local, regional, and national levels. For example, an improvement in the allocation of resources (physical, monetary, and human) within the education system is one of the key ways in which the EFA goals will be achieved. This is especially true in developing countries where there exist tight constraints on resources. Thus, education planners are particularly well positioned to use st-DSS in order to make changes and improvements to the education systems and to enhance the quality of education received by children.

This thesis proposed that with an appropriate data model and knowledge-generating tools, education planners can take advantage of the large volumes of education-related data that are routinely collected, but often not well used, within their own countries. Furthermore, through the use of these carefully designed tools in tandem with the ubiquity of the Internet, it is possible to involve more sectors of society in the education enhancement process than ever before. The remainder of this chapter shows how, with the current data infrastructure in Peru, it was possible to design a data model, build an st-DSS, and most importantly, use it in order to perform an assessment of the department of Cusco.

6.1 Thesis Objectives

This thesis used the case study of Peru to provide a concrete example of a st-DSS that advances the global primary education agenda. A set of specific objectives were outlined at the beginning of the thesis to assess the extent to which the st-DSS has satisfied this goal. The first objective was *to specify an operational data model that was both powerful and intuitive for the extraction of spatio-temporal knowledge from the type of large educational databases made available by various ministries of education.*

This objective was achieved with the proposition of a 3D data cube that uses space, time, and indicators as its three main axes. Under this model, each of the dimensions has its own dimensionality, conceptually creating sub-cubes (or cuboids). This is the same type of multi-dimensional approach typically used in OLAP systems, but it is suggested in the thesis that the model be enhanced by the addition of Spatial OLAP (SOLAP) functions and characteristics so that the spatial dimension can be fully exploited. The SOLAP functions thus developed provide all the necessary operations to manipulate the spatial, temporal, and indicator dimensions of the cube in ways that are well suited for the extraction of knowledge from spatio-temporal systems, such as education. Intrinsic in this model is the space-time paradigm, as each cube and cuboid represents a region of space-time.

The second objective of this thesis was *to take the conceptual data model described above and to build a functional instance using the data published by the MINEDU.* Although the data used in the thesis were by-and-large freely available on the Internet, their quality and completeness were poor and a wide range of obstacles were found with the organization, interpretation, and standardization of the datasets. Despite these obstacles, multiple datasets from the MINEDU and the INEI were compiled into a coherent data warehouse with a total of 724 indicators from 1993, 2002, and 2004. The data warehouse was designed to facilitate efficient execution of the spatio-temporal queries that were proposed throughout this thesis. Specifically, the database schema presented in the thesis showed how the organization of tables and the appropriate use of indices can reduce even the most complex spatio-temporal queries to less than ten seconds processing time, which is well within the human cognitive band (Newell, 1990).

Having developed both an operational data model and its implementation, the third objective of the thesis was *to build a Web-mapping interface to access this data warehouse*. This objective was achieved through the creation of the EduCal software, currently in its second version (http://gaia.uwaterloo.ca/educal2). The use of the Web to deliver the application stems from the DF strategy to engage civil society on the constantly increasing importance of networks of communication (Castells, 2000), and on the principles of public participation GIS (Hall et al., 2007). The second emphasis was on the *mapping* aspect of the interface, thereby entrenching the spatial aspect of education systems through the analysis process and making the interactions with the user more intuitive. Spatial tools such as *zoom, pan, map classification,* and an interactive *legend* were coupled with spatial selection tools, such as *select by point, circle, and polygon,* for identifying a region of interest.

Although the above-mentioned tools provide access to the education data warehouse, the fourth objective of the thesis was *to enhance the EduCal interface with a set of easy-to-use SOLAP tools in order to provide full access to the spatial and temporal dimensions of the model*. Moreover, the intention was for the SOLAP tools to provide access to both dimensions *simultaneously*, in line with the space-time paradigm used throughout the thesis. These functions included *drill down* (disaggregate), *roll-up* (aggregate), *slice* (section model), as well as a *query builder* to facilitate the selection of features with any combination of characteristics. This objective was achieved through the inclusion of various UI elements, often requiring only a few mouse clicks, and through the programming of a query builder that removes the need for non-expert users to learn a query language to use the tool fully. These elements, combined with the traditional GIS toolkit, enable an iterative logic flow that leads to a process of knowledge discovery.

In order to show the type of knowledge-discovery that is possible through tools such as EduCal, the final objective of the thesis was *to use the Web-mapping tool and develop a comprehensive assessment of the current educational conditions in a study area of Peru*. To accomplish this, the final chapter of the thesis used the department of Cusco as a case study. Using the regional education plan published by the COPARE as its basis, and using the themes and limitations identified therein, the analysis, conducted at multiple layers of spatial aggregation and using multiple years of data, was able to identify problem regions within the department.

Combined, all of the above developments have satisfied the overarching goal to fill the gap in education planning tools that exist currently and to move the global agendas closer toward providing a quality educaiton. Although the data model and software tool specified in the thesis are specific to Peru, the underlying principles and design choices are general enough to be used anywhere. Moreover, it is proposed here that the data model and the software design could be applied in other spatio-temporal problem domains than the educational domain. This possibility, as well as other areas and directions for further research are outlined in the following section.

6.2 Directions for Future Research

The research presented in this thesis can lead to a variety of further research avenues. In order for the benefits of tools such as EduCal to be realized, approaches for the dissemination of the tool to local and regional branches of the MINEDU must be examined. Moreover, diffusion of the tool to other countries should be considered.

It is possible that an approach that includes parent groups and non-government organizations (NGOs) help to increase the positive impact of the new st-DSS could be a desirable route for local acceptance and use of the software. The current research into education planning does not yet answer questions on how to create a culture around the use of information and ICTs as part of the decision making process. While it is noted throughout the thesis that tools such as EduCal are imperative in providing analysts with the type of knowledge discovery that is necessary for the formulation of comprehensive education plans, there needs to be further research into how to facilitate the uptake of such tools within already existing education planning infrastructures.

In relation to the study of the dissemination of the tool, further research using control groups should be undertaken in order to examine the impact that the tool can have on education planning. As more years of data become available, the EduCal tool can be used to measure changes in education quality, as measured by the available indicators, at any level of spatial aggregation. This type of meta-analysis is part of the intended use of the tool, as education planners will certainly be interested in knowing the results of their interventions. Colleagues at the Peruvian NGOs Alternativa (http://www.alter.org.pe) and GRADE (http://www.grade.org.pe) are currently undertaking such a study on a small scale, but more research is needed in a wider variety of contexts.

In general, there is a lack of proper information management practice across Peru that is likely mirrored across other developing countries. This was made evident when importing the freely available education data provided by the MINEDU and when attempting serious analysis with these data, as discussed previously. Until the quality of the source data that are produced by governments improves, the use of st-DSS tools will continue to have limited benefits or, if blindly used without any quality assurance or control, may even have detrimental effects. There needs to be case studies conducted in large institutions, such as the MINEDU, to understand the reasons for the lack of standardised information management. It is suggested here a lack of training is likely to be a large part of the problem, but this is not the full explanation. It is possible that the poor information management practices are also caused by internal organizational politics and individual and departmental power struggles. These types of barriers, and others yet unknown, to free-flowing information transmission must be understood if they are to be corrected.

From a software-design perspective, there needs to be usability studies that examine the effectiveness of the current user interface. It would also be of great value to study the portability of the EduCal tool to other developing countries as well as the versatility of the software to other spatio-temporal problem domains, such as health. Both of these aspects can be tested through the creation of new software solutions using the same overall design.

The overall design of EduCal, while flexible, is formed by a combination of many different components interacting together. This complexity in design emphasizes that the st-DSS technologies are still in their infancy, even though the individual components used in the design of EduCal are mature. More work must also be done to explore which other components would be well suited for inclusion into the design of future st-DSS. For example, advanced spatio-temporal database techniques and languages have been developed but have not been integrated into any decision support software to date. The use of automated spatio-temporal data mining modules within st-DSS software, such as EduCal, also holds promise for problems that are not easily discovered by manual data exploration. Finally, the use of spatial statistical techniques and modules within general st-DSS may lead advanced users to produce more thorough analyses than would be possible without such tools. None of these three components outlined here were included in EduCal, but they each add potential to its future evolution.

Appendix A

Indicators Available in EduCal

General			
	District Identifier		
	Block Identifier		
	School Identifier		
	Province Identifier		
	Department Identifier		
	Regional Education Directorate Identifier		
	Local Education Management Unit Identifier		
	Population Centre Identifier		
	Population Centre Identifier		
	Population Centre Name		
	Population Centre Name		
	Department Name		
	District Name		
	Regional Education Directorate Name		
	Province Name		
	School Name		
	Name of Urban Area		
Inputs			
	Total number of students		benefit
	Number of schools		benefit
	Percent of students receiving assistance		
		Percent of students assisted by the glass of milk program	
		Percent of male students in grade 1 assisted by the glass of milk program	benefit
		Percent of female students in grade 1 assisted by the glass of milk program	benefit
		Percent of male students in grade 2 assisted by the glass of milk program	benefit
		Percent of female students in grade 2 assisted by the glass of milk program	benefit
		Percent of male students in grade 3 assisted by the glass of milk program	benefit
		Percent of female students in grade 3 assisted by the glass of milk program	benefit
		Percent of male students in grade 4 assisted by the glass of milk program	benefit
		Percent of female students in grade 4 assisted by the glass of milk program	benefit

Percent of male students in grade 5 assisted by the glass of milk program	benefit
Percent of female students in grade 5 assisted by the glass of milk program	benefit
Percent of male students in grade 6 assisted by the glass of milk program	benefit
Percent of female students in grade 6 assisted by the glass of milk program	benefit
Percent of total students assisted by the glass of milk program	benefit
Percent of students assisted by breakfast program	
Percent of male students in grade 1 assisted by breakfast program	benefit
Percent of female students in grade 1 assisted by breakfast program	benefit
Percent of male students in grade 2 assisted by breakfast program	benefit
Percent of female students in grade 2 assisted by breakfast program	benefit
Percent of male students in grade 3 assisted by breakfast program	benefit
Percent of female students in grade 3 assisted by breakfast program	benefit
Percent of male students in grade 4 assisted by breakfast program	benefit
Percent of female students in grade 4 assisted by breakfast program	benefit
Percent of male students in grade 5 assisted by breakfast program	benefit
Percent of female students in grade 5 assisted by breakfast program	benefit
Percent of male students in grade 6 assisted by breakfast program	benefit
Percent of female students in grade 6 assisted by breakfast program	benefit
Percent of total students assisted by breakfast program	benefit
Percent of male students in grade 6 involved in free assistance program	benefit
Percent of students assisted by food program	
Percent of male students in grade 1 assisted by food program	benefit
Percent of female students in grade 1 assisted by food program	benefit
Percent of male students in grade 2 assisted by food program	benefit
Percent of female students in grade 2 assisted by food program	benefit
Percent of male students in grade 3 assisted by food program	benefit
Percent of female students in grade 3 assisted by food program	benefit
Percent of male students in grade 4 assisted by food program	benefit
Percent of female students in grade 4 assisted by food program	benefit
Percent of male students in grade 5 assisted by food program	benefit
Percent of female students in grade 5 assisted by food program	benefit
Percent of male students in grade 6 assisted by food program	benefit
Percent of female students in grade 6 assisted by food program	benefit
Percent of total students assisted by food program	benefit
Percent of students involved in free assistance program	
Percent of male students in grade 1 involved in free assistance program	benefit
Percent of female students in grade 1 involved in free assistance program	benefit

Percent of male students in grade 2 involved in free assistance program	benefit
Percent of female students in grade 2 involved in free assistance program	benefit
Percent of male students in grade 3 involved in free assistance program	benefit
Percent of female students in grade 3 involved in free assistance program	benefit
Percent of male students in grade 4 involved in free assistance program	benefit
Percent of female students in grade 4 involved in free assistance program	benefit
Percent of male students in grade 5 involved in free assistance program	benefit
Percent of female students in grade 5 involved in free assistance program	benefit
Percent of male students in grade 6 involved in free assistance program	benefit
Percent of total students involved in free assistance program	benefit
Percent of students assisted by other programs	
Percent of male students in grade 1 assisted by other programs	benefit
Percent of female students in grade 1 assisted by other programs	benefit
Percent of female students in grade 2 assisted by other programs	benefit
Percent of male students in grade 3 assisted by other programs	benefit
Percent of female students in grade 3 assisted by other programs	benefit
Percent of male students in grade 4 assisted by other programs	benefit
Percent of female students in grade 4 assisted by other programs	benefit
Percent of male students in grade 5 assisted by other programs	benefit
Percent of female students in grade 5 assisted by other programs	benefit
Percent of male students in grade 6 assisted by other programs	benefit
Percent of female students in grade 6 assisted by other programs	benefit
Percent of total students assisted by other programs	benefit
Percent of male students in grade 2 assisted by other programs	benefit
Student Health	
Average student nutrition level per school	benefit
Average male student height in grade 1 per school	benefit
Average female student height in grade 1 per school	benefit
Average male student height in grade 2 per school	benefit
Average female student height in grade 2 per school	benefit
Average male student height in grade 3 per school	benefit
Average female student height in grade 3 per school	benefit
Average male student height in grade 4 per school	benefit
Average female student height in grade 4 per school	benefit
Average student height per school	benefit
Average Male Student Nutrition Levels by Grade	benefit

Average Female Student Nutrition Levels by Grade	benefit
Average Male Student Nutrition Levels by Grade	
Average nutrition level for male students in grade 1 per school	benefit
Average nutrition level for male students in grade 2 per school	benefit
Average nutrition level for male students in grade 3 per school	benefit
Average nutrition level for male students in grade 4 per school	benefit
Average Female Student Nutrition Levels by Grade	
Average nutrition level for female students in grade 1 per school	benefit
Average nutrition level for female students in grade 2 per school	benefit
Average nutrition level for female students in grade 3 per school	benefit
Average nutrition level for female students in grade 4 per school	benefit
Percentage of Students at each Nutrition Level	
Percentage of Students with a high nutrition level	benefit
Percentage of Students with a normal nutrition level	benefit
Percentage of Students with a slightly low nutrition level	cost
Percentage of Students with a moderately low nutrition level	cost
Percentage of Students with a severely low nutrition level	cost
School facilities	
Shortage of seats as percent of total	cost
Average number of libraries per school	benefit
Shortage of tables as percent of total	cost
Average number of administrative offices per school	benefit
Average number of teachers' lounges per school	benefit
School workshops	benefit
Technology	benefit
School facilities used as classrooms	benefit
School workshops	
Average number of biology labs per school	benefit
Average number of computer labs per school	benefit
Average number of physics labs per school	benefit
Average number of language labs per school	benefit
Average number of other laboratories per school	benefit
Average number of chemistry labs per school	benefit
Average number of labs per school	benefit
Average number of computer rooms per school	benefit
Average number of art and graphics workshops per school	benefit
Average number of carpentry workshops per school	benefit
Average number of cooking workshops per school	benefit
Average number of electricity workshops per school	benefit
Average number of electronics workshops per school	benefit
Average number of mechanic workshops per school	benefit
Average number of other workshops per school	benefit
Average number of workshops per school	benefit
Technology	

Average number of computers 3 years old or newer per school	benefit
Average number of computers used for administrative purposes per school	benefit
Average number of computers used for student access per school	benefit
Ratio students per computer used for student access	cost
Ratio of operating computers to total computers	benefit
Proportion of schools with internet connections	benefit
Proportion of schools with local network connections	benefit
School facilities used as classrooms	
Average number of appropriate rooms used as classrooms per school	benefit
Average number of inappropriate rooms used as classrooms per school	cost
Inappropriate rooms uses as classrooms as a percent of total rooms	cost
Average number of libraries not used as classrooms per school	benefit
Average number of administrative offices not used as classrooms per school	benefit
Average number of computer rooms not used as classrooms per school	benefit
Average number of teachers' lounges not used as classrooms per school	benefit
Net intake	
Net intake average	
Average net intake rate of male students per school in grade 1	benefit
Average net intake rate of female students per school in grade 1	benefit
Average net intake rate of male students per school in grade 2	benefit
Average net intake rate of female students per school in grade 2	benefit
Average net intake rate of male students per school in grade 3	benefit
Average net intake rate of female students per school in grade 3	benefit
Average net intake rate of male students per school in grade 4	benefit
Average net intake rate of female students per school in grade 4	benefit
Average net intake rate of male students per school in grade 5	benefit
Average net intake rate of female students per school in grade 5	benefit
Average net intake rate of male students per school in grade 6	benefit
Average net intake rate of female students per school in grade 6	benefit
Net intake as percent	
Net intake as percent of total male students in grade 1	benefit
Net intake as percent of total female students in grade 1	benefit
Net intake as percent of total male students in grade 2	benefit
Net intake as percent of total female students in grade 2	benefit
Net intake as percent of total male students in grade 3	benefit
Net intake as percent of total female students in grade 3	benefit
Net intake as percent of total male students in grade 4	benefit
Net intake as percent of total female students in grade 4	benefit
Net intake as percent of total male students in grade 5	benefit

Net intake as percent of total female students in grade 5	benefit
Net intake as percent of total male students in grade 6	benefit
Net intake as percent of total female students in grade 6	benefit
Apparent intake	
Apparent intake average	
Average apparent intake rate of male students per school in grade 1	benefit
Average apparent intake rate of female students per school in grade 1	benefit
Average apparent intake rate of male students per school in grade 2	benefit
Average apparent intake rate of female students per school in grade 2	benefit
Average apparent intake rate of male students per school in grade 3	benefit
Average apparent intake rate of female students per school in grade 3	benefit
Average apparent intake rate of male students per school in grade 4	benefit
Average apparent intake rate of female students per school in grade 4	benefit
Average apparent intake rate of male students per school in grade 5	benefit
Average apparent intake rate of female students per school in grade 5	benefit
Average apparent intake rate of male students per school in grade 6	benefit
Average apparent intake rate of female students per school in grade 6	benefit
Apparent intake as percent	
Apparent intake as percent of total female students in grade 1	benefit
Apparent intake as percent of total male students in grade 1	benefit
Apparent intake as percent of total male students in grade 2	benefit
Apparent intake as percent of total female students in grade 2	benefit
Apparent intake as percent of total male students in grade 3	benefit
Apparent intake as percent of total female students in grade 3	benefit
Apparent intake as percent of total male students in grade 4	benefit
Apparent intake as percent of total female students in grade 4	benefit
Apparent intake as percent of total male students in grade 5	benefit
Apparent intake as percent of total female students in grade 5	benefit
Apparent intake as percent of total male students in grade 6	benefit
Apparent intake as percent of total female students in grade 6	benefit
School personnel	
Average number of librarians per school	benefit
Average number of chauffeurs per school	benefit
Average number of cooks or guards per school	benefit
Average number of laboratory workers per school	benefit
Average number of office workers per school	benefit
Average number of professional accountants per school	benefit
Average number of professional administrators (undefined role) per school	benefit
Average number of education auxiliaries per school	benefit
Average number of coordinators per school	benefit
Average number of directors and sub-directors per school	benefit
Average number of department heads per school	benefit
Average number of laboratory heads per school	benefit

	Average number of workshop heads per school	benefit
	Average number of advisors per school	benefit
	Teachers	
	Average number of female teachers per school in grade 6	benefit
	Percent of all teachers that have titles	benefit
	Average number of teachers per school	benefit
	Average number of male teachers per school in grade 3	benefit
	Average number of female teachers per school in grade 3	benefit
	Average number of male teachers per school in grade 4	benefit
	Average number of female teachers per school in grade 4	benefit
	Average number of male teachers per school in grade 5	benefit
	Average number of female teachers per school in grade 5	benefit
	Average number of male teachers per school in grade 6	benefit
	Average number of male teachers per school in grade 1	benefit
	Average number of female teachers per school in grade 1	benefit
	Average number of male teachers per school in grade 2	benefit
	Average number of female teachers per school in grade 2	benefit
Net enrolment		
	Net enrolment average	
	Average net enrolment rate of male students per school in grade 1	benefit
	Average net enrolment rate of female students per school in grade 1	benefit
	Average net enrolment rate of male students per school in grade 2	benefit
	Average net enrolment rate of female students per school in grade 2	benefit
	Average net enrolment rate of male students per school in grade 3	benefit
	Average net enrolment rate of female students per school in grade 3	benefit
	Average net enrolment rate of male students per school in grade 4	benefit
	Average net enrolment rate of female students per school in grade 4	benefit
	Average net enrolment rate of male students per school in grade 5	benefit
	Average net enrolment rate of female students per school in grade 5	benefit
	Average net enrolment rate of male students per school in grade 6	benefit
	Average net enrolment rate of female students per school in grade 6	benefit
	Net enrolment as percent	
	Net enrolment as percent of total male students in grade 1	benefit
	Net enrolment as percent of total female students in grade 1	benefit
	Net enrolment as percent of total male students in grade 2	benefit
	Net enrolment as percent of total female students in grade 2	benefit
	Net enrolment as percent of total male students in grade 3	benefit
	Net enrolment as percent of total female students in grade 3	benefit
	Net enrolment as percent of total male students in grade 4	benefit
	Net enrolment as percent of total female students in grade 4	benefit
	Net enrolment as percent of total male students in grade 5	benefit
	Net enrolment as percent of total female students in grade 5	benefit
	Net enrolment as percent of total male students in grade 6	benefit
	Net enrolment as percent of total female students in grade 6	benefit

Percent of teachers with titles	
Percent of teachers with agriculture titles	benefit
Percent of teachers with artistry titles	benefit
Percent of teachers with biology and chemistry titles	benefit
Percent of teachers with carpentry titles	benefit
Percent of teachers with social sciences and philosophy titles	benefit
Percent of teachers with natural sciences titles	benefit
Percent of teachers with electricity titles	benefit
Percent of teachers with electronics titles	benefit
Percent of teachers with special education titles	benefit
Percent of teachers with family studies titles	benefit
Percent of teachers with physics titles	benefit
Percent of teachers with history and geography titles	benefit
Percent of teachers with English (and possibly other language) titles	benefit
Percent of teachers with pre-school titles	benefit
Percent of teachers with language and literature titles	benefit
Percent of teachers with math and physics titles	benefit
Percent of teachers with math titles	benefit
Percent of teachers with mechanic titles	benefit
Percent of teachers with other titles	benefit
Percent of teachers with primary school titles	benefit
Percent of teachers with religion titles	benefit
Average Number of enrolled students per school	
Average total students (based on the recorded approved/disapproved students at the start of the previous school year)	target
Average male students per school in grade 1	benefit
Average female students per school in grade 1	benefit
Average male students per school in grade 2	benefit
Average female students per school in grade 2	benefit
Average male students per school in grade 3	benefit
Average female students per school in grade 3	benefit
Average male students per school in grade 4	benefit
Average female students per school in grade 4	benefit
Average male students per school in grade 5	benefit
Average female students per school in grade 5	benefit
Average male students per school in grade 6	benefit
Average female students per school in grade 6	benefit
Average number of students per school	benefit
Class sections	
Average number of class sections per school in grade 1	cost
Average number of class sections per school in grade 2	cost
Average number of class sections per school in grade 3	cost
Average number of class sections per school in grade 4	cost

		Average number of class sections per school in grade 5	cost
		Average number of class sections per school in grade 6	cost
		Average number of class sections per school	cost
Processes			
	Per-student/teacher ratios		
		Student-seat ratio	cost
		Student-classroom ratio	cost
		Student-teacher ratio	cost
		Student-class section ratio for grade 1	cost
		Student-class section ratio for grade 2	cost
		Student-class section ratio for grade 3	cost
		Student-class section ratio for grade 4	cost
		Student-class section ratio for grade 5	cost
		Student-class section ratio for grade 6	cost
		Overall student-class section ratio	cost
		Hours-student ratio	benefit
	Student-teacher ratio for teachers with titles		
		Student-teacher ratio for teachers with agriculture titles	cost
		Student-teacher ratio for teachers with artistry titles	cost
		Student-teacher ratio for teachers with biology and chemistry titles	cost
		Student-teacher ratio for teachers with carpentry titles	cost
		Student-teacher ratio for teachers with social science and philosophy titles	cost
		Student-teacher ratio for teachers with natural science titles	cost
		Student-teacher ratio for teachers with electricity titles	cost
		Student-teacher ratio for teachers with electronic titles	cost
		Student-teacher ratio for teachers with special education titles	cost
		Student-teacher ratio for teachers with family studies titles	cost
		Student-teacher ratio for teachers with physical education titles	cost
		Student-teacher ratio for teachers with history and geography titles	cost
		Student-teacher ratio for teachers with English (and possibly other language) titles	cost
		Student-teacher ratio for teachers with pre-school titles	cost
		Student-teacher ratio for teachers with language and literature titles	cost
		Student-teacher ratio for teachers with physics and math titles	cost
		Student-teacher ratio for teachers with math titles	cost
		Student-teacher ratio for teachers with mechanic titles	cost
		Student-teacher ratio for teachers with other titles	cost
		Student-teacher ratio for teachers with primary school titles	cost
		Student-teacher ratio for teachers with religion titles	cost
	Average age of students		
		Average age of male students in grade 1	target
		Average age of female students in grade 1	target
		Average age of male students in grade 2	target

Average age of female students in grade 2	target
Average age of male students in grade 3	target
Average age of female students in grade 3	target
Average age of male students in grade 4	target
Average age of female students in grade 4	target
Average age of male students in grade 5	target
Average age of female students in grade 5	target
Average age of male students in grade 6	target
Average age of female students in grade 6	target
Average student age overall	target
Percent of students older than the appropriate ages	
Percent of male students 1 year older than the appropriate age in grade 1	cost
Percent of female students 1 year older than the appropriate age in grade 1	cost
Percent of male students 1 year older than the appropriate age in grade 2	cost
Percent of female students 1 year older than the appropriate age in grade 2	cost
Percent of male students 1 year older than the appropriate age in grade 3	cost
Percent of female students 1 year older than the appropriate age in grade 3	cost
Percent of male students 1 year older than the appropriate age in grade 4	cost
Percent of female students 1 year older than the appropriate age in grade 4	cost
Percent of male students 1 year older than the appropriate age in grade 5	cost
Percent of female students 1 year older than the appropriate age in grade 5	cost
Percent of male students 1 year older than the appropriate age in grade 6	cost
Percent of female students 1 year older than the appropriate age in grade 6	cost
Percent of male students 2 years older than the appropriate age in grade 1	cost
Percent of female students 2 years older than the appropriate age in grade 1	cost
Percent of male students 2 years older than the appropriate age in grade 2	cost
Percent of female students 2 years older than the appropriate age in grade 2	cost
Percent of male students 2 years older than the appropriate age in grade 3	cost

Percent of female students 2 years older than the appropriate age in grade 3	cost
Percent of male students 2 years older than the appropriate age in grade 4	cost
Percent of female students 2 years older than the appropriate age in grade 4	cost
Percent of male students 2 years older than the appropriate age in grade 5	cost
Percent of female students 2 years older than the appropriate age in grade 5	cost
Percent of male students 2 years older than the appropriate age in grade 6	cost
Percent of female students 2 years older than the appropriate age in grade 6	cost
Percent of male students 3 years or more older than the appropriate age in grade 1	cost
Percent of female students 3 years or more older than the appropriate age in grade 1	cost
Percent of male students 3 years or more older than the appropriate age in grade 2	cost
Percent of female students 3 years or more older than the appropriate age in grade 2	cost
Percent of male students 3 years or more older than the appropriate age in grade 3	cost
Percent of female students 3 years or more older than the appropriate age in grade 3	cost
Percent of male students 3 years or more older than the appropriate age in grade 4	cost
Percent of female students 3 years or more older than the appropriate age in grade 4	cot
Percent of male students 3 years or more older than the appropriate age in grade 5	cost
Percent of female students 3 years or more older than the appropriate age in grade 5	cost
Percent of male students 3 years or more older than the appropriate age in grade 6	cost
Percent of female students 3 years or more older than the appropriate age in grade 6	cost
Percent of male students equal to appropriate age in grade 1	benefit
Percent of female students equal to appropriate age in grade 1	benefit
Percent of male students equal to appropriate age in grade 2	benefit
Percent of female students equal to appropriate age in grade 2	benefit
Percent of male students equal to appropriate age in grade 3	benefit
Percent of female students equal to appropriate age in grade 3	benefit
Percent of male students equal to appropriate age in grade 4	benefit

Percent of female students equal to appropriate age in grade 4	benefit
Percent of male students equal to appropriate age in grade 5	benefit
Percent of female students equal to appropriate age in grade 5	benefit
Percent of male students equal to appropriate age in grade 6	benefit
Percent of female students equal to appropriate age in grade 6	benefit
Student desertion	
Male student desertion rate in grade 1	cost
Female student desertion rate in grade 1	cost
Male student desertion rate in grade 2	cost
Female student desertion rate in grade 2	cost
Male student desertion rate in grade 3	cost
Female student desertion rate in grade 3	cost
Male student desertion rate in grade 4	cost
Female student desertion rate in grade 4	cost
Male student desertion rate in grade 5	cost
Female student desertion rate in grade 5	cost
Male student desertion rate in grade 6	cost
Female student desertion rate in grade 6	cost
Net desertion	
Net male student desertion rate in grade 1	cost
Net female student desertion rate in grade 1	cost
Net male student desertion rate in grade 2	cost
Net female student desertion rate in grade 2	cost
Net male student desertion rate in grade 3	cost
Net female student desertion rate in grade 3	cost
Net male student desertion rate in grade 4	cost
Net female student desertion rate in grade 4	cost
Net male student desertion rate in grade 5	cost
Net female student desertion rate in grade 5	cost
Net male student desertion rate in grade 6	cost
Net female student desertion rate in grade 6	cost
Apparent desertion	
Apparent male student desertion rate in grade 1	cost
Apparent female student desertion rate in grade 1	cost
Apparent male student desertion rate in grade 2	cost
Apparent female student desertion rate in grade 2	cost
Apparent male student desertion rate in grade 3	cost
Apparent female student desertion rate in grade 3	cost
Apparent male student desertion rate in grade 4	cost
Apparent male student desertion rate in grade 5	cost
Apparent female student desertion rate in grade 5	cost
Apparent male student desertion rate in grade 6	cost
Apparent female student desertion rate in grade 6	cost
Apparent female student desertion rate in grade 4	cost

Absenteeism	
Total Absenteeism	cost
Non-health Absenteeism	
Total Non-health Related Absenteeism	cost
Non-health Absenteeism for Grade 1 Males	cost
Non-health Absenteeism for Grade 2 Males	cost
Non-health Absenteeism for Grade 3 Males	cost
Non-health Absenteeism for Grade 4 Males	cost
Non-health Absenteeism for Grade 5 Males	cost
Non-health Absenteeism for Grade 6 Males	cost
Non-health Absenteeism for Grade 1 Females	cost
Non-health Absenteeism for Grade 2 Females	cost
Non-health Absenteeism for Grade 3 Females	cost
Non-health Absenteeism for Grade 4 Females	cost
Non-health Absenteeism for Grade 5 Females	cost
Non-health Absenteeism for Grade 6 Females	cost
Health-related Absenteeism	
Total Health-related Absenteeism	cost
Health-related Absenteeism for Grade 1 Males	cost
Health-related Absenteeism for Grade 2 Males	cost
Health-related Absenteeism for Grade 3 Males	cost
Health-related Absenteeism for Grade 4 Males	cost
Health-related Absenteeism for Grade 5 Males	cost
Health-related Absenteeism for Grade 6 Males	cost
Health-related Absenteeism for Grade 1 Females	cost
Health-related Absenteeism for Grade 2 Females	cost
Health-related Absenteeism for Grade 3 Females	cost
Health-related Absenteeism for Grade 4 Females	cost
Health-related Absenteeism for Grade 5 Females	cost
Health-related Absenteeism for Grade 6 Females	cost
Absenteeism by Gender	
Total Male Absenteeism	cost
Total Female Absenteeism	cost
Absenteeism for Males in Grade 1	cost
Absenteeism for Males in Grade 2	cost
Absenteeism for Males in Grade 3	cost
Absenteeism for Males in Grade 4	cost
Absenteeism for Males in Grade 5	cost
Absenteeism for Males in Grade 6	cost
Absenteeism for Females in Grade 1	cost
Absenteeism for Females in Grade 2	cost
Absenteeism for Females in Grade 3	cost
Absenteeism for Females in Grade 4	cost
Absenteeism for Females in Grade 5	cost

	Absenteeism for Females in Grade 6	cost
Ratio of girls to boys		
	Ratio of girls to boys in grade 1	neutral
	Ratio of girls to boys in grade 2	neutral
	Ratio of girls to boys in grade 3	neutral
	Ratio of girls to boys in grade 4	neutral
	Ratio of girls to boys in grade 5	neutral
	Ratio of girls to boys in grade 6	neutral
Student repetition		
	Male student repetition rate (including students from other schools) in grade 1	cost
	Female student repetition rate (including students from other schools) in grade 1	cost
	Male student repetition rate (including students from other schools) in grade 2	cost
	Female student repetition rate (including students from other schools) in grade 2	cost
	Male student repetition rate (including students from other schools) in grade 3	cost
	Female student repetition rate (including students from other schools) in grade 3	cost
	Male student repetition rate (including students from other schools) in grade 4	cost
	Female student repetition rate (including students from other schools) in grade 4	cost
	Male student repetition rate (including students from other schools) in grade 5	cost
	Female student repetition rate (including students from other schools) in grade 5	cost
	Male student repetition rate (including students from other schools) in grade 6	cost
	Female student repetition rate (including students from other schools) in grade 6	cost
	Male student repetition rate in grade 1	cost
	Female student repetition rate in grade 1	cost
	Male student repetition rate in grade 2	cost
	Female student repetition rate in grade 2	cost
	Male student repetition rate in grade 3	cost
	Female student repetition rate in grade 3	cost
	Male student repetition rate in grade 4	cost
	Female student repetition rate in grade 4	cost
	Male student repetition rate in grade 5	cost
	Female student repetition rate in grade 5	cost
	Male student repetition rate in grade 6	cost
	Female student repetition rate in grade 6	cost

Percent of students that work outside of school		
	Percent of male students that work outside of school in grade 1	cost
	Percent of female students that work outside of school in grade 1	cost
	Percent of male students that work outside of school in grade 2	cost
	Percent of female students that work outside of school in grade 2	cost
	Percent of male students that work outside of school in grade 3	cost
	Percent of female students that work outside of school in grade 3	cost
	Percent of male students that work outside of school in grade 4	cost
	Percent of female students that work outside of school in grade 4	cost
	Percent of male students that work outside of school in grade 5	cost
	Percent of female students that work outside of school in grade 5	cost
	Percent of male students that work outside of school in grade 6	cost
	Percent of female students that work outside of school in grade 6	cost
Outputs		
Overall student survival ratio		benefit
Male student survival ratio		benefit
Female student survival ratio		benefit
Rate of male students continuing to secondary school after grade 6		benefit
Rate of Female students continuing to secondary school after grade 6		benefit
Overall rate of students continuing to secondary school after grade 6		benefit
Student promotion		
	Male student promotion rate to grade 1 (presumably from pre-school)	benefit
	Female student promotion rate to grade 1 (presumably from pre-school)	benefit
	Male student promotion rate to grade 2	benefit
	Female student promotion rate to grade 2	benefit
	Male student promotion rate to grade 3	benefit
	Female student promotion rate to grade 3	benefit
	Male student promotion rate to grade 4	benefit
	Female student promotion rate to grade 4	benefit
	Male student promotion rate to grade 5	benefit
	Female student promotion rate to grade 5	benefit
	Male student promotion rate to grade 6	benefit
	Female student promotion rate to grade 6	benefit
	Overall student promotion rate	benefit
Context		
Average wealth index per home		benefit
Percent of homes with a water connection		benefit
Percent of homes with automobiles		benefit
Percent of homes with a kitchen or shared kitchen		benefit
Percent of homes without electricity		benefit

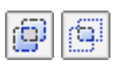
Percent of homes within improper forms of dwellings ('area', 'shack', 'improvised', 'non-humane', 'other', 'hotel', 'hospital', 'jail', 'asylum', 'collective-other', 'transient', 'homeless','boat','house-of-minors')	cost
Percent of children with adults that can read and write	benefit
Percent of children with parents that can read and write	benefit
Percent of children with parents that work	benefit
Total population within the homes	benefit
Total population recorded	benefit
Total number of homes	benefit
Children's home characteristics	
Percent of children in homes where income producing activities are performed	cost
Percent of children living in homes with at least one poorly constructed component	benefit
Percent of children living in homes with at least two poorly constructed components	benefit
Percent of children living in homes with all three components poorly constructed	benefit
Children in homes with a bathroom or shared bathroom	benefit
Percent of children in homes with a bathroom or shared bathroom	cost
Children with homes classified as either 'shack', 'improvised' or 'non-humane'	benefit
Percent of children with homes classified as either 'shack', 'improvised' or 'non-humane'	cost
Children in homes with a kitchen or shared kitchen	benefit
Percent of children in homes with a kitchen or shared kitchen	cost
Children in homes within improper forms of dwellings ('area', 'shack', 'improvised', 'non-humane', 'other', 'hotel', 'hospital', 'jail', 'asylum', 'collective-other', 'transient', 'homeless','boat','house-of-minors')	benefit
Percent of children in homes within improper forms of dwellings ('area', 'shack', 'improvised', 'non-humane', 'other', 'hotel', 'hospital', 'jail', 'asylum', 'collective-other', 'transient', 'homeless','boat','house-of-minors')	cost
Children in homes that are classified as 'invaded', or 'used without payment'	benefit
Percent of children in homes that are classified as 'invaded', or 'used without payment'	cost
Children with homes classified as either 'transient', 'homeless', or 'house-of-minors'	cost
Percent of children with homes classified as either 'transient', 'homeless', or 'house-of-minors'	cost
Children in homes that are fully paid-for	benefit
Percent of children in homes that are fully paid-for	benefit
Children in homes that are owned, but not fully paid-for	benefit

Percent of children in homes that are owned, but not fully paid-for	cost
Children in homes that are in rented dwellings	benefit
Percent of children in homes that are in rented dwellings	cost
Percent of children with parents that speak Spanish or Quechua	benefit
Percent of children with parents that have completed university education	benefit
Percent of children with parents that have incomplete education	benefit
Percent of children with parents that have at least one year of non-university post secondary education	benefit
Percent of children with parents that have completed primary education	benefit
Percent of children with parents that have at least one year of secondary school	benefit
Percent of children with parents that have any level of education	benefit
Percent of children with parents that have at least one year of university post secondary education	benefit
Percentage of children with parents that speak Spanish	benefit
Percent of children with parents that speak Quechua	benefit
Percent of children living in homes with poorly constructed walls	benefit
Percent of children living in homes with poorly constructed floors	benefit
Percent of children living in homes with poorly constructed ceilings	benefit
Percent of children living in homes with computers	benefit
Percent of children without adults that speak Spanish or Quechua	cost
Percent of children without adults that have any level of education	cost
Percent of children without adults that speak Spanish	cost
Percent of children without adults that can read and write	cost
Percent of children without adults that speak Quechua	cost
Percent of children without adults that work	cost
Percent of children without parents that speak Spanish or quechua	cost
Percent of children without parents that have any level of education	cost
Percent of children without parents that speak Spanish	cost
Percent of children without parents that can read and write	cost
Percent of children without parents that speak Quechua	cost
Percent of children without parents that work	cost
Percent of children with adults that work	benefit
Children in homes with more than two people per bedroom	benefit
Percent of children living in homes with a wealth index of 50 or higher	benefit
Population characteristics	
Population that speak Spanish for ages 5 and up	benefit
Population that can read	benefit
Total number of children under the age of 15	benefit
Population of age 15 and up	benefit
Population of age 15 and up as a percent of total	benefit
Population of age 5 to 14	benefit

	Population of age 5 to 14 as a percent of total	benefit
	Population of age 5 and up	benefit
	Population of 5 and up as a percent of total	benefit
	Population that have some level of education for ages 5 and up	benefit
	Percent of population that have some level of education for ages 5 and up	benefit
	Population that completed primary education for ages 15 and up	benefit
	Percent of population that completed primary education for ages 15 and up	benefit
	Percent of population that speak Spanish for ages 5 and up	benefit
	Population that have some kind of physical impediment	benefit
	Percent of population that have some kind of physical impediment	benefit
	Percent of population that can read	benefit
	Population that have a profession for ages 15 and up	benefit
	Population that have a profession in armed forces or policing for ages 15 and up	benefit
	Percent of population that have a profession in armed forces or policing for ages 15 and up	benefit
	Population that have a scientific profession for ages 15 and up	benefit
	Percent of population that have a scientific profession for ages 15 and up	benefit
	Population that have an official profession for ages 15 and up	benefit
	Percent of population that have an official profession for ages 15 and up	benefit
	Percent of population that have a profession for ages 15 and up	benefit
	Population that have a technical profession for ages 15 and up	benefit
	Percent of population that have a technical profession for ages 15 and up	benefit
	Population that speak Quechua for ages 5 and up	benefit
	Percent of population that speak Quechua for ages 5 and up	benefit
	Population that are underemployed for ages 15 and up	cost
	Percent of population that are underemployed for ages 15 and up	cost
	Population that are unemployed for ages 15 and up	cost
	Percent of population that are unemployed for ages 15 and up	cost
	Population that completed university education	benefit
	Percent of population that completed university education for ages 15 and up	benefit
	Population that employed under the age of 15	cost
	Percent of population that employed under the age of 15	cost
Household Characteristics		
	Total number of homes with automobiles	benefit
	Average population per home	benefit
	Number of habitants per bedroom.	cost
	Homes where income producing activities are performed	benefit
	Percent of homes where income producing activities are performed	benefit

Total number of homes with a water connection	benefit
Percentage of homes with poor construction material	benefit
Homes with a bathroom or shared bathroom	benefit
Percent of homes with a bathroom or shared bathroom	benefit
Homes classified as either 'shack', 'improvised' or 'non-humane'	cost
Percent of homes classified as either 'shack', 'improvised' or 'non-humane'	cost
Homes with a kitchen or shared kitchen	benefit
Total number of homes without electricity	benefit
Homes within improper forms of dwellings ('area', 'shack', 'improvised', 'non-humane', 'other', 'hotel', 'hospital', 'jail', 'asylum', 'collective-other', 'transient', 'homeless','boat','house-of-minors')	cost
Homes that are classified as 'invaded', or 'used without payment'	cost
Percent of homes that are classified as 'invaded', or 'used without payment'	cost
Homes classified as either 'transient', 'homeless', or 'house-of-minors'	cost
Percent of homes classified as either 'transient', 'homeless', or 'house-of-minors'	cost
Homes that are fully paid-for	benefit
Percent of homes that are fully paid-for	benefit
Homes that are owned, but not fully paid-for	cost
Percent of homes that are owned, but not fully paid-for	cost
Homes that are in rented dwellings	cost
Percent of homes that are in rented dwellings	cost
Total number of homes without sanitation	benefit
Percent of homes without sanitation	benefit
Percentages of homes with poor walls	cost
Percentages of homes with poor floors	cost
Average bedrooms per household	benefit
Percentages of homes with poor ceilings	cost
Total bedrooms per household	benefit

Appendix B
Selection Tools

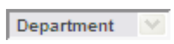
Tool	Icon	Description
Mode		There are three mode options: *New*, *Add To*, *Remove*. The *New* option ignores previous selections and only includes new selections in the selection set. The *Add To* option retains the current selection and adds the new selections to form a larger selection set. The *Remove* option retains the current selection, except for the features selected, to form a smaller selection set.
Style		There are two style options: *Pass-through* and *Within*. The *Pass-through* option accepts features that have some or all of their extent within the drawn selection. The *Within* option accepts only those features whose entire extent falls inside the drawn selection shape.
Point		This tool is used to select map features with a single click of the mouse in the map. Only those features from the active layer that fall directly under the point are selected.
Line		This tool can be used to select map features by drawing a line on the map. Clicking on the map, once for each endpoint, draws a line. Multi-segmented lines are drawn by continuing to drag and click. A double-click ends the process and those features from the active layer that touch the line are selected.

Tool	Icon	Description
Rectangle		This tool is used to select map features by drawing a rectangle on the map. A rectangle is drawn by clicking on the map, dragging the mouse to a new location, and clicking again.
Circle		This tool is used to select map features by drawing a circle on the map. A circle is drawn by clicking on the map to define the centre of the circle, dragging the mouse to define the size of the circle, and clicking a second time to complete the circle.
Polygon		This tool is used to select map features by drawing a polygon on the map. A polygon is drawn by clicking the mouse anywhere on the map, dragging the mouse to a new location and clicking again to define one edge of the polygon. This process can be continued until the polygon is complete and then a double click closes the polygon. The first and last points of the polygon do not need to be the same as EduCal automatically generates the last polygon edge and completes the polygon.
Remove		This tool is used to remove completely the current selection. The previous selection set is restored to the map. In this way, it is possible to scroll back through selection sets (i.e., undo).
Remove All		This tool is used to remove completely all previous and current map selections. This tool completely erases the selection history.

Source: Adapted from Hall et al. (2006)

Appendix C

Drilling Tools

Drill Level		The drill level drop-down menu is used to use the current selection in order to create a selection from another layer. By selecting any of the listed layers in the drop-down menu, the current selection will either roll-up or drill-down by using whichever of the drill styles is currently pressed. As the selection set moves from one layer to another it takes on a different size and shape, based on the boundaries of the features in the new layer.
Drill Style		There are two drill mode options: *By Identifier* and *By Geometry*. *Drill by Identifier* ensures that the unique identifiers associated with individual features (in the database schema of Figure 4.6) are used to drill from one level to another. This option gives faster results, but is only available for those layers that are linked by foreign keys. *Drill by Geometry* ensures that the actual geometry of the selected features is used to perform the drill. This option is available for all layers, including those added by users.

Source: Adapted from Hall et al. (2006)

Appendix D

Glossary of Acronyms

AJAX	Asynchronous Javascript and XML
API	Application Programming Interface
CCPP	Population Centres (Centros Poblados)
COPARE	Consejo Participativo Regional de Educación
DBMS	Database Management System
DF	Dakar Framework
DRE	Regional Education Management Unit (Dirección Regional de Educación)
DSS	Decision Support System
EFA	Education For All
ESRI	Environmental Systems Research Institute
ESU	Education Statistics Unit
GIS	Geographic Information Systems
GML	Geographic Markup Language
GUI	Graphical User Interface
HIV/AIDS	Human Immunodeficiency Virus / Acquired Immunodeficiency Syndrome
HTML	HyperText Markup Language
ICT	Information and Communication Technology
INEI	National Statistics and Information Institute (Instituto Nacional de Estadística e Informática)
MAUP	Modifiable Areal Unit Problem
MBR	Minimum Bounding Rectangle
MINEDU	Peruvian Ministry of Education (Ministerio de Educación)
NGO	Non-governmental Organisation
OLAP	On-Line Analytical Processing
OSG	Open Source GeoSpatial
OSS	Open Source Software
REP	Regional Education Plan
SDSS	Spatial Decision Support System
SOLAP	Spatial On-Line Analytical Processing
SQL	Structured Query Language
st-DSS	Spatio-Temporal Decision Support System
TBWA	Target Based Weighted Average
UGEL	District Education Management Unit (Unidad de Gestión Educativa Locale)
UMC	Unit of Education Quality Measurement
UN	United Nations
UNESCO	United Nations Educational Scientific and Cultural Organisation
XML	eXtensible Mark-up Language
WCEFA	World Conference on Education for All
WEF	World Education Forum
WMS	Web Mapping Service

Bibliography

"context." (2004). Merriam-Webster Online Dictionary. Retrieved October 9, 2005, from: http://www.merriam-webster.com.

"geography." (2004). Merriam-Webster Online Dictionary. Retrieved September 21, 2005, from: http://www.merriam-webster.com.

"plan." (2004). Merriam-Webster Online Dictionary. Retrieved October 9, 2005, from: http://www.merriam-webster.com.

"quality". (2005). Oxfored English Dictionary. [Electronic version].

Abraham T., & Roddick, J.F. (1999). Survey of Spatio-Temporal Databases. *GeoInformatica*. 3(1): 61-99.

AdaptivePath. (2005). Ajax a new approach to Web Applications, Adaptive Path. Retrieved on December 19, 2006, from: http://www.adaptivepath.com/publications/essays/archives/000385.php

Ahmad, I., Azhar, S., & Lukauskis, P. (2004). Development of a decision support system using data warehousing to assist builders/developers in site selection. *Automation in Construction*, 13(4), 525–542.

Alderman H., Hoddinott J., & Kinsey B. (2006). Long-term consequences of early childhood malnutrition. *Oxf Econ Pap*, 58, 450–74.

Andrienko, N.V., & Andrienko, G.L. (2001) Intelligent Support for Geographic Data Analysis and Decision Making in the Web, *Journal of Geographic Information and Decision Analysis*, 5(2), 115-128.

Armstrong, M. P., Densham, P. J. & Rushton, G. (1986) Architecture for a microcomputer based spatial decision support system, *Second International Symposium on Spatial Data Handling*, 120-131

BBC. (2007, October 22). Peruvians kept indoors for census. Retrieved October 23, 2007, from: http://news.bbc.co.uk/2/hi/americas/7056053.stm.

Beckmann, N., Kriegel, H.P., Schneider, R., & Seeger, B. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles. *Proc. of the ACM SIGMOD Conf.*, 322-331.

Bédard, Y. (2005). JMAP Spatial OLAP: On-Line Analytical Processing for Spatial Databases. *KHEOPS Technologies White Paper*

Bédard, Y., Gosselin, P., & Rivest, S. (2003). Integrating GIS components with knowledge discovery technology for environmental health decision support. *International Journal of Medical Informatics*, 70(1), 79-94.

Bédard, Y., Larrivée, S., Proulx, M.J., Caron P.Y. & Létourneau, F. (1997). Geospatial Data Warehousing: Positionnement technologique et stratégique. Rapport pour le Centre de recherche pour la défense de Valcartier (CRDV).

Bédard, Y., Merrett, T., & Han, J. Fundamentals of spatial data warehousing for geographic knowledge. In H. Miller, & H. Jiawei (Eds.), *Geographic data mining and knowledge discovery* (pp. 53-73). 2001. London: Taylor & Francis.

Bertrand, D. (2003). *Achieving the universal primary education goal of the millennium declaration: New challenges for development cooperation*. Geneva, CH: United Nations.

Bhargava, Hemant K., Power, Daniel J., Sun, Daewon. 2007. Progress in Web-based decision support technologies. *Decision Support Systems*. 43: 1083–109.

Billa, L., Shattri, M., Mahmud, A.R., & Ghazali, A.H. (2006). Comprehensive planning and the role of SDSS in flood disaster management in Malaysia. *Disaster Prevention and Management*, 15(2), 233-240.

Blalock, M. (1964). A comparison of aspects of the US and UK censuses of population. *Transactions in GIS,* 1, 49-60

Bojórquez-Tapia, L., Díaz-Mondragón, S. & Ezcurra, E.. (2001). GIS-based approach for participatory decision making and land suitability assessment. *International Journal of Geographical Information Systems,* 15(2), 129-151.

Bradford, M. (1991). School-performance indicators, the local residential environment, and parental choice. *Environment and Planning A*, 23(3), 319-332.

Byrne, D. & Rogers, T. (1996) Divided Spaces - Divided School: An Exploration of the Spatial Relations of Social Division. *Sociological Research Online*, 1(2).

Caron, P.-Y. (1998). Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle. (Masters, Faculté de foresterie et géomatique, Université Laval), 132.

Carver, S., Evans, A., Kingston, R., & Turton, I. (1998) Geographical Information Systems on the World Wide Web: improving public participation in environmental decision making. *Proceedings of Conference of the European Association for the Study of Science and Technology,* Lisbon, Portugal, September 30-October 3, 1998.

Carver, S., Evans, A., Kingston, R., & Turton, I. (2001) Public participation, GIS and cyberdemocracy: evaluating on-line spatial decision support systems. *Environment and Planning B: Planning and Design*, 28(6), 907-921.

Castells, M. (1998). Information Technology, Globalisation and Social Development. *UNRISDN Conference on Information Technologies and Social Development*, Palais des Nations, Geneva.

Castells, M. (2000). *The Rise of the Network Society* (2nd ed.). Oxford: Blackwell

Chakravarthy, S. & Kim, S.-K. (1994). Resolution of time concepts in temporal databases. *Information Sciences*, 80(1-2), 91-125.

Chaudhuri S., Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65-74.

Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, 13(6), 377-387.

Codd, E. F. (1971). A Data Base Sublanguage Founded on the Relational Calculus. *Proc. ACM-SIGFIDET Workshop on Data Description, Access, and Control*, San Diego.

Codd, E.F. (1995). *Providing OLAP to User Analysts: An IT Mandate*. California: Hyperion.

COPARE. (2006). Proyecto Educativo Regional Cusco. Retrieved on January 10, 2007, from: http://www.cne.gob.pe/descentralizacion/cusco.zip

Combi, C., & Montanari, A. Data Models with Multiple Temporal Dimensions: Completing the Picture (revised version). *Research Report 40/00, Dipartimento di Matematica ed Informatica, Universita'di Udine*, December 2000.

Couclelis, H. (1999). Space, time, geography. In Longley, MF Goodchild, DJ Maguire (eds.) *Geographical information systems: principles and techniques*, New York: Oxford University Press.

de Echave, J. (2001). *Construyendo un proceso de toma de decisiones frente a operaciones mineras*. Lima: Cooperación.

de Echave, J. (2005). Peruvian peasants confront the mining industry. *Socialism and Democracy*, 19(3), 117-127.

de Silva, N.F., Eglese, R.W. (2000). Integrating simulation modelling and GIS: spatial decision support systems for evacuation planning. *Journal of Operational Research Society*, 51(4), 423-430.

Densham, P. (1991). Spatial Decision Support Systems. In Maguire, D.J., Goodchild, M. & Rhind, D. W. *Geographical Information Systems: principles and applications*, Volume 1, United Kingdom: Longman Scientific and Technical.

DiNIECE. 2006. Promoción del uso de estadísticas educativas en investigación. Ministerio de Educación, Ciencia y Tecnología, Argentina. Retrieved on March 30, 2007, from: http://diniece.me.gov.ar/concurso2006/basededatos.html.

Duraisamy, P., James, E., Lane J., & Tan, J.-P. (1998). Is There a quantity-quality trade-off as pupilteacher ratios increase? evidence from Tamil Nadu, India. *International Journal of Educational Development*, 18(5), 367-383.

Einot, I. & Gabriel, KR. (1975). A survey of the powers of several methods of multiple comparisons. *J Amer Statist Assoc*, 70, 574-583.

Engler, N.J. (2007). Spatial Analysis of the Effect of Absenteeism on Education Quality in Maynas, Peru. (Masters, University of Waterloo).

Evans, A.J., Kingston, R., Carver, S. (2004). Democratic input into the nuclear waste disposal problem: The influence of geographical data on decision making examined through a Web-based GIS. *Journal of Geographical Systems*, 6, 117-132.

Flowerdew, R., & Pearce, J. (2001). Linking point and area data to model primary school performance indicators. *Geographical & Environmental Modelling*, 5(1), 23-41.

Fotheringham, A.S. Charlton, M.E., & Brunsd, C. (2001). Spatial Variations in School Performance: a Local Analysis Using Geographically Weighted Regression. *Geographical and Environmental Modelling*, 5(1), 43-66.

Frank, A. (1982). Mapquery-Database query language for retrieval of geometric data and its graphical representation. *ACM Comp. Graphics*, 16(3), 199-207.

Greene, B. (2005). *Fabric of the Cosmos*. XXX:Vintage.

Guenther, O. & Buchmann, A. Research Issues in Spatial Databases. *SIGMOD Record*, 19, 61-68.

Güting, R.H. (1994). An Introduction to Spatial Database Systems. *Very Large Database Journal*, 3(4), 537-399.

Guttman, A. (1984). R-Trees: A Dynamic Index Structure for Spatial Searching. *Proc of the ACM SIGMOD Conf*, 47-57.

Hall GB, Alperin JP, Léon Kerrigan S. El uso de Internet con software libre y fuentes abiertas espaciales para colaborar en la toma de decisiones. XI Conferencia Iberoamericana en Sistemas de Información Geográfica. Buenos Aires, Argentina. May 28-31, 2007.

Hall, G.B., & Bowerman, R.L.. (1996). Using GIS to evaluate the accessibility of family planning services in the central valley of Costa Rica, *ITC Journal*, 1, 38-48.

Hall, G.B., Engler, N.J., Alperin, J.P., & Leahy, M.G. (2005). EduCal Tutorial Book: Iquitos Edition. Waterloo, Canada: University of Waterloo.

Han, J. (1997). OLAP Mining: An Integration of OLAP with Data Mining. *Proc. IFIP Conf. Data Semantics*.

Hoch, C. (2007). Making Plans: Representation And Intention. *Planning Theory*, 6(1), 16-35.

Hopkins,L. (2000) *Urban Development: The Logic of Making Plans*. Washington,DC: Island Press.

Ibaugh, A., & Rushton, G. (2003). A spatial decision support system for improving the coordination and delivery of health care services. Retrieved on 21 June, 2007, from http://www.uiowa.edu/~geog/health/index3.html

INEI. (2007). Registro Nacional de Municipalidades. Retrieved on September 22, 2007, from: http://www.inei.gob.pe

Internet World Stats. (2007). World internet usage statistics. Retrieved May 13, 2007, from: http://www.internetworldstats.com/stats.htm

Jankowski, P., Robischon, S., Tuthill, D., Nyerges, T., & Ramsey, K. (2006) Design Considerations and Evaluation of a Collaborative, Spatio-Temporal Decision Support System. *Transactions in GIS*. 10(3): 335–354.

Johnson, M.P. (2001). A spatial decision support system prototype for housing mobility program planning. *Journal of Geographical Systems*, 3, 49-67.

Kadzamira, E. & Rose, P. (2003) Can free primary education meet the needs of the poor? Evidence from Malawi. *International Journal of Educational Development*, 23, 501-516.

Keenan, P. (1996). Using a GIS as a DSS generator. In: Darzentas, J., Darzentas, J.S. and Spyrou, T., (eds.). *Perspectives on DSS*, University of the Aegean, Greece, 33–40.

Kimball, R. (1996). *The Data Warehouse Toolkit*, New York: J. Wiley & Sons.

Kingston, R., Carver, S., Evans, A., & Turton, I. (2000) Web-based public participation geographical information systems: an aid to local environmental decision-making. *Computers, Environment and Urban Systems*, 24, 109-125.

Klinkenberg, B. (2003). The true cost of spatial data in Canada. *The Canadian Geographer*. 47, 37-49.

Kokkotos, S., Ioannidis, E.V., Panayiotopoulos, T., & Spyropoulos, C.D. (1995). On the Issue of Valid Time(s) in Temporal Databases. *SIGMOD Record*, 24(3), 40-43.

Koutsoukis N-S, Mitra G., &Lucas C. (1999). Adapting on-line analytical processing for decision modeling: the interaction of information and design technologies. *Decision Support Systems*, 26(1), 1-30.

Leahy, M.G. (2005). Spatial Analysis of Education Quality in Peru. (Masters, University of Waterloo).

Malczewski, J. (1999) GIS and Multicriteria Decision Analysis. New York: John Wiley & Sons.

Malczewski, J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*. 20:7, 703-726.

Marinoni, O. (2006). A discussion on the computational limitations of outranking methods for land-use suitability assessment. *International Journal of Geographical Information Science*, 20(1), 69-87.

Maxfield, D.W. (1972). Spatial Planning Of School Districts. *Annals of the Association of American Geographers*. 62(4), 582-590.

McGinn, N. & Welsh, T. (1999). *Decentralization of education: why, when, what and how?* Paris: IIEP.

McMahon, W. W. (1993). *An Efficiency-Based Management Information System*. Paris:UNESCO - International Institute for Education Planning.

Medeiros, C.B. & Pires, F. (1994). Databases for GIS. *SIGMOD Record*, 23, 107-115.

Miller, R.G. Jr. (1966). *Simultaneous Statistical Difference.* New York: McGraw-Hill.

MINEDU. (2005). Consejo Participativo Regional de Educación. Retrieved on November 11, 2007, from: www.minedu.gob.pe/oaae/xtras/folletoCOPARE.pdf

MINEDU. (2005). Estadística de Calidad Educativa. Retrieved on March 5, 2005, from: http://escale.minedu.gob.pe/

MINEDU Bolivia. Estadísticas escolares 1996 - 2004. Retrieved on June 22, 2007, from: http://www2.minedu.gov.bo/estads/dossie/default.html.

Mokbel, M.F., Ghanem, T.M., & Aref, W.G. (2003). Spatio-temporal Access Methods. *IEEE Data Engineering Bulletin*, 26(2).

Newell, A. (1990). *Unified theories of cognition.* Cambridge: Harvard University Press.

Ochoa, M., & Bonifaz, A. (2002) An analysis of disparities in education: the case of primary school completion rates in Bolivia. Retrieved on October 10, 2005, from: http://ideas.repec.org/p/wpa/ wuwphe/0302001.html

OLAP council. 1995. OLAP Research and Technology. Retrieved on June 3, 2006, from: http://www.olapcouncil.org/research/resrchly.htm

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Openshaw, S. (1991). Developing Appropriate Spatial Analysis Methods for GIS. In Maguire, D.J., Goodchild, & M.F., Rhind, D.W. (eds.) *Geographical Information Systems: principles and applications*. Vol. 1, Longman, London, 389-402.

Openshaw, S., & Taylor P. J. 1981. The modifiable areal unit problem. In Wrigley N., Bennet R.J. (eds) *Quantitative Geography: a British View*. London, Routledge: 60-70.

Peisa S.A. (2003). *Atlas departamental del Perú*. Lima: Peisa.

Peters, P.A. (2002). Assessing education quality in peru: The role of information in education planning and management. (Masters, University of Waterloo).

Peters, P.A., & Hall, G.B. (2004). Evaluation of education quality and neighbourhood well-being: A case study of independencia, peru. *International Journal of Education Development*, 24, 85-102.

Peuquet, D. (2001). Making Space for Time: Issues in Space-time Data Representation. *GeoInformatica*, 15(1), 11-32.

Powell, C.A., Walker, S.P., Chang, S.M., & Grantham-McGregor, S.M. (1998). Nutrition and education: A randomized trial of the effects of breakfast in rural primary school children. *Am J Clin Nutr*. 68, 873-9.

Power, D.J. A Brief History of Decision Support Systems, version 4.0. *DSSResources.com*, Retrieved March 10, 2007, from: http://DSSResources.COM/history/dsshistory.html

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., Townsend, T. (2000). An introduction to school effectiveness research. In: Reynolds, D., & Teddlie, C. (Eds.), *The International Handbook of School Effectiveness Research*. New York: Falmer Press.

Riddell, A.R. (1999). The need for a multidisciplinary framework for analyzing education reform in developing countries. *International Journal of Educational Development*, 19(3), 207-217.

Rinner, C. (2003). Web-based Spatial Decision Support: Status and Research Directions. *Journal of Geographic Information and Decision Analysis*. 7(1), 14-31.

Rivest, S., Bédard, Y., & Marchand, P. (2001). Towards better support for spatial decision-making: Defining the characteristics of spatial on-line analytical processing (SOLAP). *Journal of the Canadian Institute of Geomatics*, 55, 539-555.

Robinson, W.S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15, 351-7.

Scheerens, J. (2000). *Improving School Effectiveness*. Paris: UNESCO.

Schutzberg, A. (2005, June 22). Open Source Geospatial '05: Ten Ideas Worthy of Note. *Directions Magazine*. Retrieved Available at April 7, 2007, from: http://www.directionsmag.com/article.php?article_id=891.

Scott Morton, M.S. (1971). A Framework f o r Management Information Systems. *Sloan Management Review*, 13(1), 55-70.

SecuritySpace. (2006a). Web Survey Survey. Retrieved on December 19, 2006 from: http://www.securityspace.com/s_survey/data/200611/index.html

SecuritySpace. (2006b). Apache Module Report. Retrieved on December 19, 2006 from: http://www.securityspace.com/s_survey/data/man.200611/apachemods.html

Sellis, T., Roussopoulos, N., Faloutsos, C. (1987). R+-Tree: A Dynamic Index for Multi-Dimensional Objects. *Proc. of the 13th VLDB Conf.*, 507-518.

Sharifi, M.A. & v. Herweijnen, M. (2003). *Spatial decision Support Systems*. Enschede: ITC.

Shekhar, S., Coyle, M, Goyal, B., Liu, D.-R., & Sarkar, S. (1997). Data models in geographic information systems. *Communications of the ACM*. 40(4), 103-111.

Sriraj, P.S., Minor, M., Thakuriah, P. (2006). Spatial Decision Support System for Low-Income Families Relocation Tool for the Chicago, Illinois, Region. *Transportation Research Record*. 1956(1), 119-126. Retrieved on June 24, 2007, from: http://trb.metapress.com/content/924267705u867n27.

Statistics Canada. Education in Canada: Chronological index. Retrieved on June 22, 2007, from: http://www.statcan.ca/bsolc/english/bsolc?catno=81-229-X&CHROPG=1

The Register. (2005a). South America warms to Open Source. Retrieved December 22, 2006, from: http://www.theregister.co.uk/2005/02/10/south_america_open_source/

The Register. (2005b). Peru's parliament approves pro-open source bill. Retrieved December 22, 2006, from: http://www.theregister.co.uk/2005/09/29/peru_goes_open_source/

Thomsen, E. (2002). *OLAP solutions: Building multidimensional information systems* (Second ed.). New York: John Wiley & Sons, Inc.

Tryon, R.C. (1939). *Cluster Analysis*. Ann Arbor, MI: Edwards Brothers.

UNESCO. (1990). World Conference on Education For All: Framework for Action: Meeting Basic Needs

UNESCO. (1998). Wasted Opportunities: When schools fail. Repetition and drop-out in primary school. Education for All, Status and Trends 1998.

UNESCO. (2000a). The Dakar Framework for Action, Education for All: Meeting our collective commitments. Proceedings of the World Education Forum in Dakar, Senegal, April 26-28.

UNESCO. (2000b). Education for All 2000: Statistical Document. Paper presented at the World Education Forum in Dakar, Senegal. April 26-28.

UNESCO. (2002). Education For All: Is The World On Track. France: UNESCO.

UNESCO. (2004). Education For All Global Monitoring Report: Gender and Education for All – The Leap To Equality. UNESCO, Paris.

UNESCO. (2005a). Education For All Global Monitoring Report: The Quality Imperaive. UNESCO, Paris.

UNESCO. (2005b). Decentralization in education: National policies and practices. Education Policies and Strategies 7. Paris.

UNESCO. (2006). Education management information system. Retrieved January 28, 2008, from: http://portal.unesco.org/education/en/ev.php-URL_ID=10202&URL_DO=DO_TOPIC&URL_SECTION=201.html

Ward, J.H. Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *J Amer. Statist. Assoc*. 58(301), 236-244.

Zar, J.H. (1984). *Biostatistical Analysis*. New Jersey: Prentice Hall, Englewood Cliffs.