

## **Designing school-wide assessment systems: Challenges and opportunities**

Alexis A. López  
Universidad de los Andes  
allopez@uniandes.edu.co

### **Abstract**

In May 2009, the Colombian National Ministry of Education passed a new law, Decree 1290, which now requires all primary and secondary schools to design their own school-wide assessment system. Prior to this new law, the Ministry had always imposed a national assessment system to all schools. This national assessment system had always been subject to many criticisms from school administrators, teachers, parents and students because it was not formative in nature and because it forced schools to promote at least 95% of the students. The task of creating a new assessment system poses a great challenge for many schools in the sense that the members of the school community have no experience designing formative assessment systems. This study focuses on a case study of a public school in Bogota. Through interviews to teachers and school administrators, observations of planning meetings, and artifacts, the study documents the school's difficulties in trying to design a valid assessment system. The study also documents the opportunities that this new law provides to schools to align their assessment system to their curriculum, to align it to their needs and to include all the stakeholders in this process.

### **Key Words**

Assessment system, accountability, autonomy, democratic assessment, alignment

## **INTRODUCTION**

In April, 2009, the Colombian National Ministry of Education passed a new decree, Decree 1290, which regulates the assessment of learning and promotion of students in elementary and secondary schools. This decree requires all schools in Colombia to design institutional assessment systems. The purposes of these assessment systems are to measure student learning, to monitor progress, to provide information to help students who need support, to provide information to design improvement plans, and to determine the promotion of students. Another important point in this new law is the need for each school to define a rating scale to assess their students. Even though schools have autonomy to establish their own rating scale, this has to be aligned to a National Rating Scale (Superior, High, Basic and Low). This, in theory, will allow mobility from one school to another.

Prior to this new law, the Ministry had always imposed a national assessment system to all schools. This national assessment system had always been subject to many criticisms from school administrators, teachers, parents and students because it was not formative in nature and because it forced schools to promote at least 95% of the students, regardless of their performance. The task of creating a new assessment system poses a great challenge for many schools in the sense that the members of the school community have no experience designing formative assessment systems. The purpose of this study is to examine the challenges and opportunities that teachers and administrators at Simon Bolivar School (a pseudonym) have when designing a valid school-wide assessment system in light of Decree 1290. In particular, this study attempts to answer the following two research questions:

1. What challenges do teachers and administrators at Simon Bolivar School have to comply with Decree 1290?
2. What opportunities does Decree 1290 provide for teachers and administrators at Simon Bolivar School to design an appropriate institutional assessment system?

## **LITERATURE REVIEW**

### **Educational Reform, Assessment and Accountability**

From an educational perspective, reform may be sought as an “expression of concern with how well schools are functioning and the quality of educational outcomes and/or student learning” (Chalhoub-Deville, 2008, p. 12). Educational initiatives intend to achieve “the closing of gaps in achievement among racial/ ethnic groups and between economically disadvantaged students and their more affluent counterparts” (Linn, 2006, p. 23). Furthermore, educational reform might signal a deep belief that “education lies at the heart of economic development, international competitiveness, and social harmony” (Chalhoub-Deville, 2008, p. 12).

Linn (2006) explains test-based accountability as the “engine” of educational reform. Tests are created and used for various reasons: attractiveness to stakeholders such as the public, politicians, and policymakers, cost-effectiveness. Another reason tests are used concerns tangibility; this is to say that teachers and administrators can be held responsible for gains or losses described by test scores (Chalhoub-Deville, 2008; Linn, 2006).

In many countries, “standardized testing continues to play a prominent role in educational policy and in efforts to improve the quality of education” (Herman & Golan, 1993, p. 20). High-stakes standardized tests are often used a way to enforce educational reforms that are aimed at improving teaching and learning, and to hold teachers and schools accountable for student achievement (Linn, 2000; Thompson, 2001). Several educational agencies attach

consequences at the school level to test scores such as warnings, loss of accreditation and funding costs (Bond, Braskamp, & Roeber, 1996). Consequently, many schools are forced to take measures that guarantee improvement in test scores such as purchasing test preparation materials (Vogler & Kennedy, 2003) or teachers are forced to increase test preparation practices (Ligon, 2000). A lot of the instruction time is spent preparing students for the test, on focusing more on the content that is on going to be tested, and on teaching test-taking skills (Thompson, 2001). There is also a tendency for teachers to narrow the curriculum by only focusing on the content areas and skills that are assessed on the test (Falk, 2002; Shepard, 1990).

Accountability refers to the process of making decisions and applying consequences based on the information collected through assessment systems (Hill & DePascale, 2003). To be accountable is to take seriously the investments that different stakeholders have in any given testing process (Norton, 1997). But in order for accountability systems to be valid and reliable, assessment systems must also be valid and reliable (Hill & DePascale, 2003). In recent years, calls for accountability have focused on the consequences of assessment practices for test takers, who have been up to this point a relatively powerless group in (Norton Pierce & Stein, 1995; Raines, 1990; Shohamy, 1993). The recognition that testing practices should be accountable to test takers is indicative of the broader trend towards democratization of educational assessment in general (Darling-Hammond, 1994; Gipps, 1994). Part of this is due to the fact that there is an unequal relationship between test makers and test takers, resulting in greater accountability for test takers (Pierce, 1992).

As interest in accountability toward individual test takers increases, the demand that assessment practices address system-wide accountability is also growing. So there is a need to address the tensions between accountability to individuals and accountability to systems (Norton, 1997). The work of test developers can affect the future of many people that they many never see and who will never hold them accountable for their actions (Hamp-Lyons, 1997b). Stevenson (1981) claims that all the stakeholders should be held accountable for the test and the consequences brought upon them. Accountability should be a central priority for all stakeholders and should presuppose a priori obligations to all stakeholders.

Large-scale assessments are closely linked to accountability. It is believed that school achievement will improve if education systems identify what is to be learned and then assess to determine the effectiveness of instruction. The education reform movement in the United States brought about higher standards along with new assessments in order to help students who were not succeeding. Schools are under a lot of pressure to inform the public about what they are teaching and how effective they are (Brindley, 1998).

High-stake tests have become the accountability tool of choice in many states as policy makers struggle to find ways to increase student achievement and monitor progress (Gottlieb, 2003). Many schools resort to quick-fix strategies to increase test scores as they feel these competing pressures. Schools and teachers now have the pressure to demonstrate, through large-scale assessments, that ELLs are making improvements in their learning process. This is a clear example of how tests are used to determine accountability for student learning based on each state's academic content and achievement standards (Gottlieb, 2003).

## **METHOD**

### **Research Design**

This is a qualitative study that adheres to a constructivist paradigm (Mertens, 2005). The basic assumption guiding the constructivist paradigm are that knowledge is socially constructed by people active in the research process and that the researcher should attempt to understand a case

from a point of view of those who live it (Schwandt, 2000). This study uses an intrinsic case study design (Yin, 1994) and attempts to understand and describe, in depth, the challenges and opportunities that a school in Bogota, Colombia went through in designing a school-wide assessment system.

### **The Case**

The case for this study is Simon Bolivar School, a coeducational, low-income public school in the southeast part of Bogota, Colombia. I used purposeful sampling to select the school; according to Patton (1990), “the purpose of purposeful sampling is to select information-rich cases whose study will illuminate the questions under study” (p. 169). I established the following criteria to select the case:

1. I sought a public school in Bogota.
2. I sought a school that did not have an assessment system according to the new law.
3. I sought a school that allowed me to work with them in the design of the new assessment system.

Simon Bolivar School has three buildings (branches), two for primary school and one for secondary school. The school day is five hours long and each building hosts two sessions per day (6:30 – 11:30 am and 12:00 – 5:00 pm). The school has 82 teachers. All faculty members hold university degrees in education and/or their field of specialty. The school is led by a principal, a vice principal, and three building coordinators. The school has a population of 2561 students and the average class size is 45. Students come from poor, violent areas in the southeast part of Bogota. Most of them have many economic problems and come from single-parent families.

### **Participants**

Twenty teachers and five school administrators participated in this study. These were the stakeholders that led the team that designed the institutional assessment system in order to comply with Decree 1290. In addition, three teachers and two administrators were interviewed to provide more depth about the school’s challenges and opportunities in designing the assessment system. The teachers and the administrators were selected because they were the more vocal in the meetings and became the leaders of the process. The principal has been in the school for three years and has over 30 years of experience. The coordinator has been in the school for 10 years and works in the high school afternoon session. The three teachers work in primary, middle school and high school, respectively.

### **Data Collection**

Data were collected from a variety of sources to establish credibility and validity and it was collected over a period of 16 weeks, from September to November, 2009. I was able to collect rich data so it could be processed and analyzed continuously throughout the study. Below I describe each data collection instrument.

*Observations.* I employed qualitative observations as the primary means to describe and analyze what took place in the planning meetings. The school set up six planning meetings, two hours each, to design their institutional assessment system. A total of five school administrators (the principal, the vice-principal, and three school coordinators) and 20 teachers from all three campuses attended those meeting. I also participated in those meetings as an advisor. A research assistant sat in those meetings taking detailed field notes about everything that was said in those meetings. The meetings were audio recorded to complement the field notes.

*Interviews.* Two school administrators and three teachers were interviewed once, using a semi-structured interview protocol. The purpose of these interviews was to gather background information from each participant and to gather information about their perceptions on the new law, and the challenges and opportunities it gave the school to improve their assessment system. Each interview lasted approximately 30 minutes and were audio-recorded.

*Artifacts.* After each planning meeting, teachers were divided into groups to work on different tasks. The tasks were: 1) revise the current assessment system, 2) establish criteria for assessing students and scale, strategies to assess students, criteria for promotion of students, 3) establish actions to monitor student progress, strategies to support student learning and self-assessment criteria, 4) design reports to students, establish how the report works, and design communication strategies with parents, and 5) establish criteria to assess students with special needs and establish how the assessment system will be socialized to the school community. I collected all these tasks so they could be analyzed at a later time.

### **Data Analysis and Validity**

All the data were analyzed qualitatively, using a grounded theory approach (Strauss & Corbin, 1994). I chose this approach because I was not sure what to expect from the collected data. Next, the field notes, interviews and artifacts were transcribed, and I analyzed the transcriptions through a process of coding. The transcriptions were read and re-read to identify emergent categories and common themes (Bogdan & Biklen, 1998). The categories, concepts, and themes were not pre-set, but instead were derived directly from the data as issues and ideas that were important and relevant to the participants. To ensure that the data analysis was valid and reliable, I employed data triangulation and peer debriefing to improve the likelihood that our findings and interpretations were credible (Lincoln & Guba, 1985). I compared and contrasted findings from one data collection instrument to confirm or disconfirm my interpretations, and I also presented the data analysis to two different researchers to explore inquirer biases and to clarify the meanings and the biases for the interpretations that were made.

### **Ethical Considerations**

In order to guarantee that this study was ethical and responsible, I sought informed consent from the school and all the participants. I met with school officials prior to the beginning of the study to inform them about the purpose of the study, their involvement, and the voluntary nature of the study. Only the teachers and administrators that signed the consent letter participated in the study and were allowed to withdraw from it at any time without any negative consequences. All names were changed to protect the identity of the school and the participants.

## **FINDINGS AND DISCUSSION**

### **Challenges for the School**

Complying with Decree 1290 posed many challenges for Simon Bolivar School and its stakeholder. For example, I found that teachers and administrators lacked adequate training in assessment. Also, the school found it difficult to design an assessment system that met their needs and the needs of the students because their initiatives were in conflict with the national academic standards and with the national standardized tests. And finally, the school had difficulties setting proper promotion criteria. Below, I describe each of these challenges in more detail.

**Lack of Training in Assessment.** One of the biggest problems in using assessments to improve the quality of education is the lack of training in assessment by teachers (Lopez, 2009). One of

the coordinators explained that the school is not prepared to design their own assessment system because they are not trained to do so. Likewise, one of the teachers commented that “we have always used the same instruments to assess our students and we don’t know what other alternatives we have.” Many teachers in the school use only multiple-choice questions because these are the types of questions that are used in the national standardized tests. In the planning meetings, several teachers wanted the school to mandate multiple-choice tests at the end of each academic quarter to help students do well in standardized tests. In the end, the school decided to use multiple-choice tests at the end of each quarter. This decision shows the lack of training that teachers and administrators have in assessment and it was opportunity for the school to move towards using more authentic assessment instruments – portfolios, hand-on activities, essays, and others. Despite numerous meetings on these topics, the teachers and administrators at Simon Bolivar School lack concrete guidance in the creation and use of alternative assessments. Finally, another teacher commented that she had concerns about executing the assessment system. She argued, “We have an assessment system on paper. The challenge is to put this system into practice.”

**Limitations in Their Autonomy to Design an Assessment System.** Even though the majority of the school community felt the new law gave them autonomy to design their own assessment system, they believe this autonomy has several boundaries or limitations. Several teachers thought they could improve their current assessment practices, but were, to an extent, pressured to show improvement based on the results of the national standardized tests. One of the teachers stated that, “we have the autonomy to do whatever we want, but we still have to align our teaching to the standards and to the tests, whether we want to do it or not.” The school principal also worries about this autonomy in the sense that many schools are not equipped to make sound decisions, such as determining the most effective way to assess their students and the best way to use these assessments. In the planning meetings, everyone agreed that they should take advantage of this autonomy to design an assessment system that meets their needs and to change the negative perception that many school stakeholders have about assessment. They also agreed that perhaps they are not ready to make the right decisions because they lack proper training in assessment. They want the National Ministry of Education and the local educational agencies to provide a lot of support.

**Setting Adequate Promotion Criteria.** Prior to the passing of Decree 1290, schools and teachers complained about having to promote at least 95% of the students each year, regardless of their performances. Now that schools have the opportunity to set their own promotion criteria, teachers are concerned about the impact that their decisions could have on teachers, administrators, students, parents and the educational system. For instance, one teacher explained that if they retain too many students they could create overcrowded classrooms the following school year. Likewise, in one of the planning meetings, a few teachers were concerned about this responsibility. They felt that school desertion rates are currently very high in public schools in Bogota. If more students are retained, the number of students who quit school could increase dramatically.

### **Opportunities for the School**

Even though Decree 1290 poses many challenges for schools, administrators and teachers, it also gives them many opportunities to improve their assessment practices. In this study, I found that now this school has an opportunity to have a democratic assessment system, to align their assessment to their instruction, and to use assessment instruments that meet students’ needs. Below I discuss each of these opportunities.

**Make Assessment a Democratic Process.** Decree 1290 has given Simon Bolivar School the opportunity to involve all stakeholders (i.e. administrators, teachers, parents and students) in the design of their institutional assessment system. The principal stated that “this is the first time we

meet, as a school, to talk about assessment.” Prior to this new law, the National Ministry of Education in Colombia simply mandated what to assess and how to assess it, without involving teachers, parents or students. Now, the entire school participated in the design of the assessment system at Simon Bolivar School. Twenty teachers and five school administrators met six times, for two hours, during the second semester of the academic year to plan a strategy to design the assessment system. From these meetings, several tasks were assigned that required all teachers from different grades and areas to engage in a discussion about what to assess, how to assess, how to score the assessments, and how to interpret and use these assessments. One of the teachers commented that these tasks allowed them to reflect about their teaching practices.

**Align Assessment to Instruction.** All the participants agreed that Decree 1290 is an invaluable opportunity to improve the quality of education in Colombia. Teachers feel that now they can improve their assessment practices. In the past, assessment was more summative in nature as it was conducted primarily at the end of the teaching-learning process and it was used simply to assign grades. As one teacher stated, “We have been criticizing our assessment system for so long. This is our only chance to improve it.” In the planning meetings, teachers stated that their assessment instruments were not aligned to their instruction. Now they have the opportunity to align their assessment activities to what they do in class. This will allow teachers to make more valid inferences about what their students are learning. One teacher explained, “This is our opportunity to make assessment an integral part of the teaching-learning process.”

**Use Assessment Practices that Meet Students’ Needs.** Decree 1290 also gives schools the chance to design an assessment system that meets their needs. The school principal pointed out that Simon Bolivar School places students with special needs in regular classes. This is an opportunity for the school to take the needs of these students into account and design an assessment system that will allow them to demonstrate what they are learning. Similarly, one of the teachers stated that there were many factors (e.g. large classes, short class hours, lack of materials) that did not allow them to have a formative assessment system in the past. With these meetings and the tasks that were assigned, the teachers are now able to come up with strategies that will help them overcome these factors. This is an opportunity for the school to have an assessment system that reflects their needs and realities and allows the teachers to use it to improve the quality of education at Simon Bolivar School.

## CONCLUSION

Decree 1290 has many challenges, but it also gives schools many opportunities to improve the teaching-learning process. Even though I present the results from a single case study, a school in Bogota, most of the findings represent the realities of many public schools in Colombia. The new law is allowing schools to design their own assessment systems. This is good because schools have the opportunity to align their institutional assessment system to their specific context and needs. This will allow schools to make valid inferences about what their students are learning and eventually use assessment to improve the quality of the education they provide. But it is important that schools get a lot of support to design these assessment systems. Many stakeholders, including teachers and administrators in Colombia schools, lack appropriate training in assessment. This lack of training could lead them to make wrong decisions. Thus, it is imperative that the National Ministry of Education promotes and supports more teacher education and teacher training programs focusing on classroom assessment. As a result of this, teachers could improve their assessment practices and start using more authentic ways to assess student learning. More research is needed to examine how schools are reacting to Decree 1290 and to understand how these assessment systems are improving the teaching-learning process.

## REFERENCES

- Bogdan, R. C. and Biklen, S. K. (1998) *Qualitative research for education: an introduction to theory and methods*, Boston, MA: Allyn and Bacon.
- Bond, L. A., Braskamp, D., & Roeber, E. (1996). *The status report of the assessment programs in the United States*. Washington, DC: The Council of Chief State School Officers. Oak Brook, IL: North Central Regional Education Laboratory.
- Chalhoub-Deville, M. (2008). *The social and educational impact of standards-based assessment in the USA*, paper presented at ALTE Third International Conference, Cambridge.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Falk, B. (2002). Standards-based reforms: Problems and possibilities. *Phi Delta Kappan*, 83 (8), 612-620.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Washington, DC: Falmer Press.
- Gottlieb, M. (2003). Large-scale assessment of English language learners: Addressing educational accountability in K-12 settings. *Professional Paper #6*. Alexandria, VA: TESOL.
- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12(4), 20-25.
- Hill, R. K., and DePascale, C. A. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice*, 22(3), 12-20.
- Ligon, G. D. (2000). Trouble with a capital T. *The School Administrator*, 57 (11), 40-44.
- Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry*, Beverly Hills, CA: Sage Publications, Inc.
- Linn, R. L. (2000). Assessment and accountability. *Educational Researcher*, 29, 4-16.
- Lopez, A. A. (2009). El Decreto 1290: Oportunidades y retos. *Temas: Revista de Coyuntura Instituto para la Investigación Educativa y el Desarrollo Pedagógico – IDEP*, 1, 25-26.
- Mertens, D. M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods (2<sup>nd</sup> edition)*. Thousand Oaks, CA: SAGE Publications.
- Patton, M. (1990). *Qualitative evaluation and research methods (2nd ed.)*. Thousand Oaks, CA: Sage Publications.
- Schwandt, T. A. (2001). *Dictionary of qualitative inquiry (2<sup>nd</sup> ed.)*. Thousand Oaks, CA: Sage
- Shepard, L. A. (1990). Inflated test score gains: is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9, 15-22.
- Stevenson, D. K. (1981). Language testing and academic accountability: On redefining the role of language testing in language teaching. *International Review of Applied Linguistics*, 19, 15-30.
- Straus, A. and Corbin, J. (1994). Grounded theory methodology: an overview, in Denzin, N K and Lincoln, Y S (Eds), *Handbook of Qualitative Research*, Thousand Oaks, CA: Sage, 273-85.
- Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 82, 358-62.
- Vogler, K. E., and Kennedy, R. (2003). A view from the bottom: What happens when your school is ranked last on the state's high-stakes testing program. *Phi Delta Kappan*, 84 (6), 446-448.
- Yin, R. (1994). *Case study research: Design and methods (2nd ed.)*. Thousand Oaks, CA: Sage Publishing.