

# **Three generations of assessments of basic education\* in Brazil: interfaces with the curriculum in/of the school**

Alicia Bonamino

*Pontifical Catholic University of Rio de Janeiro*

Sandra Zákia Sousa

*University of São Paulo*

## **Abstract**

Three generations of large scale assessments of education are analyzed in this article based on the objectives and designs usual in initiatives of this kind implemented in Brazil. The first generation consists in the diagnostic assessment of the quality of education, without attribution of direct consequences for schools and for school curricula. The other two generations articulate the results of the assessments to accountability policies, with the attribution of symbolic or material consequences for the school agents. Taking as a parameter of analysis the objectives and designs of these assessments, as well as studies and researches that produced results about this theme, possible implications for the school curriculum are explored. On the one hand, a discussion is made of the risks that standardized tests, with evaluations that make reference to accountability policies involving weak and strong consequences, may exacerbate the concern of teachers and principals with the preparation for the tests and for the activities they include, leading to a narrowing of the school curriculum. On the other hand, the text points out the potential of the second and third generation assessments to stimulate an informed discussion of school curriculum in terms of the fundamental abilities of reading and mathematics which have not yet been guaranteed to all pupils.

### **Contact:**

**Alicia Bonamino**  
**Pontifícia Universidade Católica**  
**do Rio de Janeiro**  
**Departamento de Educação**  
**Rua Marquês de São Vicente,**  
**225, 1050 L**  
**22451-900 – Rio de Janeiro/RJ**  
**alicia@puc-rio.br**

## **Keywords**

Assessment of education – Accountability – School curriculum.

\* Translator's note: In the Brazilian education system, basic education comprises early childhood education, primary and secondary school. This paper analyses data on the academic performance of students in the 4th and 8th grade of primary school and 3rd grade of secondary school.

# **Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola**

Alicia Bonamino

*Pontifícia Universidade Católica do Rio de Janeiro*

Sandra Zákia Sousa

*Universidade de São Paulo*

## **Resumo**

*Analisa-se, neste artigo, três gerações de avaliação da educação em larga escala, a partir dos objetivos e desenhos usuais em iniciativas implementadas no Brasil. A primeira geração consiste na avaliação diagnóstica da qualidade da educação, sem atribuição de consequências diretas para as escolas e para o currículo escolar. As outras duas gerações articulam os resultados das avaliações a políticas de responsabilização, com atribuição de consequências simbólicas ou materiais para os agentes escolares. Tomando como parâmetro de análise os objetivos e desenhos dessas avaliações, bem como estudos e pesquisas que produziram evidências sobre o tema, exploram-se possíveis implicações para o currículo escolar. Por um lado, discutem-se os riscos de as provas padronizadas, com avaliações que referenciam políticas de responsabilização envolvendo consequências fracas e fortes, exacerbarem a preocupação de diretores e professores com a preparação para os testes e para as atividades por estes abordadas, levando a um estreitamento do currículo escolar. Por outro lado, aponta-se o potencial das avaliações de segunda e terceira gerações em propiciarem uma discussão informada sobre o currículo escolar, em termos das habilidades fundamentais de leitura e matemática que ainda não têm sido garantidas a todos os alunos.*

## **Palavras-chave**

*Avaliação da educação – Responsabilização – Currículo escolar.*

### **Correspondência:**

**Alicia Bonamino**

**Pontifícia Universidade Católica  
do Rio de Janeiro**

**Departamento de Educação**

**Rua Marquês de São Vicente, 225, 1050 L**

**22451-900 – Rio de Janeiro/RJ**

**alicia@puc-rio.br**

Among the milestones in the preparation and implementation of educational policies in Brazil over the last two decades, the assessments with elements in common with proposals that materialized in other countries stand out, expressing a global agenda. In addition to other objectives, assessment initiatives aim to promote teaching quality by ultimately establishing new parameters for the management of educational systems.

As for the curriculum, in most countries, regardless of the degree of centralization or decentralization of the regulation forms of school curricula, there is a tendency to use centralized assessments to measure student academic performance, under the same curriculum parameters to which one considers all students should have access to.

This more universalist perspective is reinforced by the consensus that seems to exist on a global scale about the small variability of the curriculum proposals, which impacts on the contents of national assessments and on the recent participation of 65 countries in the Programme for International Student Assessment (PISA), based on the idea that each country's curriculum is comparable to that of the other countries involved.

In Brazil, the analysis of the design of ongoing assessments allows identifying three generations of large-scale education assessments, with different consequences for the school curriculum. These generations coexist within school systems, hence the need to use this classification as an analytical tool.

The first generation emphasizes the assessment with a character of diagnosis of the quality of the education offered in Brazil, without direct consequences for schools and the curriculum. At the current phase of large-scale assessment initiatives, two new models of assessment have emerged with the purpose of supporting accountability policies with consequences for school staff in function of student performance. In the literature on this theme, when the consequences of these policies are only symbolic, they are called

low-stakes or weak accountability. When such consequences are serious, they are called high-stakes or strong accountability (CARNOY, LOEB, 2002; BROOKE, 2006). Such assessments are identified in the text as second and third generation assessments respectively.

In Brazil, first generation assessments are those whose purpose is to monitor the evolution of the education quality. In general, these assessments do not give feedback to schools, but disclose results on the Internet for public consultation, or use the media or other forms of dissemination.

In addition to public disclosure, second generation assessments give feedback to schools without establishing material consequences. In this case, the consequences are symbolic and result from the dissemination and appropriation of information on each school's results by parents and society. The assumption of this type of accountability mechanism is that knowledge of the results promotes not only the mobilization of the school staff to improve education but also the pressure from parents and community on the school (ZAPONI; VALENCIA, 2009).

Third generation assessments are those that ground high stakes or strong accountability policies, which include sanctions or rewards related to the results of students and schools. In this case, they include accountability experiments made explicit through norms and compensation mechanisms in function of the goals set. (ZAPONI; VALENCIA, 2009).

This article seeks to characterize the ongoing experiences of assessment of basic education in the country and discusses possible relationships with the school curriculum. Our interest is to answer the following questions: under what conditions does the large-scale assessment of education have consequences for the school curriculum? Is there evidence of the influence of assessment on the school curriculum? For this purpose, we have adopted as an analysis parameter the objectives and designs of such assessments as well as studies and research that have produced evidence on the subject.

This text is organized in four other sections besides this introduction. The first section discusses the main features of the National System of Basic Education Assessment (SAEB) which allow identifying it as a first generation assessment. In its subsections, the following section covers the second and third generation assessments and the relationships between differentiated forms of accountability and school curriculum, taking *Test Brazil* and Sao Paulo and Pernambuco state assessments as examples. The third section presents a review of studies and research on the relationships between such assessments and the school curriculum in different contexts. At the end, we present some conclusions.

### **First generation of large-scale assessment policies: SAEB**

There is evidence that the state was interested in making assessment be part of educational planning since the 1930s. However, not until the late 1980s did assessment gradually integrate government policies and practices of basic education. As Dirce Nei Teixeira de Freitas (2007) stated,

[...] It took about five decades for assessment (external, large-scale, centralized and focused on student achievement and on performance of education systems) to be introduced as a systematic practice in the government of Brazilian basic education. (p. 51)

The use of educational tests has increased since the 1960s<sup>1</sup>. However, the first initiative of organizing a systematic nationwide assessment of primary and secondary education took place in the late 1980s. Since 1991, the Ministry of Education has called it National System of Basic Education Assessment (SAEB)<sup>2</sup>.

**1-** See GATTI, 1987

**2-** Alicia Bonamino (2002) offers a more detailed characterization of the institutionalization of SAEB.

Every two years, SAEB, the main system for assessing the quality of basic education, tests the performance of a sample of 4th and 8th graders in primary education and in the 3rd year of secondary school in public and private schools located in urban and rural areas.

Students take performance tests in conjunction with questionnaires about factors associated with these results, focusing on schools and their staff. By 2009, there were ten assessment cycles.

Since its creation<sup>3</sup>, SAEB has had an appropriate design to diagnose and monitor the quality of basic education in the Brazilian geographic regions and states. In 1995, methodological innovations were introduced in its design, which consolidated its current configuration. Such innovations are: i) inclusion of private schools in the sample, ii) adoption of *Item Response Theory* (IRT)<sup>4</sup>, which allows estimating students skills regardless of the specific set of items answered; iii) option to work with the last grade of each school cycle (4th and 8th grade of primary school and the third grade of secondary school), iv) prioritization of areas of knowledge of Portuguese (focus on reading) and mathematics (focus on problem solving); v) participation of 27 states; vi) questionnaires on student's socio-cultural characteristics and study habits. The introduction of these innovations has allowed comparing the student performance considering grades and years.

**3-** SAEB started years before, but it was formally created in 1994 by Ordinance No. 1795 of December 27.

**4-** The Item Response Theory (IRT) is a mathematical model that estimates the ability of individuals in a particular area or discipline assuming that it is one-dimensional. In other words: it is assumed, for example, that students have an ability or competence in mathematics that defines the probability that a particular student performs adequately the various activities of the question bank. It has some advantages over the classical approach because it allows: putting students and questions in the same scale; making more accurate estimates of changes over time, by equating the scores; estimating a measure of the ability of students which takes into account the difficulty of questions, i.e., the most difficult questions have greater weight in determining individual scores. In IRT scores, SAEB adopts an average of 250 points, which corresponds to the national average of 8th graders in 1997. With the assumptions of IRT, it is possible to construct a single scale of scores for populations of different schooling levels: in the case of Brazil, 4th and 8th-graders and 3rd secondary school. This allows comparing the mean *proficiency* in each discipline between the different levels of education, between regions and between different years, putting all levels in the same scale.

SAEB cognitive tests are prepared based on reference matrices, devised with a *synthesis* of what different municipal, state and national curricular proposals have in common, as well as consultations with teachers and experts in Portuguese language and mathematics and the analysis of the most used textbooks in the systems and grades assessed.

Although the development of tests leads to the definition of what should be considered essential in terms of school learning and, therefore, of what all students should know and be able to do at the end of certain cycles of schooling, SAEB has a low level of interference in schools and in the curriculum because it has a sample basis.

Its design is adequate to diagnose and monitor the evolution of basic education quality, but it does not allow measuring the evolution of the performance of individual students or schools. Their results are reported fairly aggregated and therefore cannot support the introduction of policies of accountability of teachers, principals and managers for quality improvements at schools.

In addition, while the Ministry of Education developed a sample basis assessment of basic education, states and municipalities felt the need to implement assessments that reached all schools. Such need has led several states to adopt their own assessment systems. Minas Gerais state, for example, created the System for the Assessment of Public Education (SIMAVE) in 1991 and Ceará state created the Permanent System for the Assessment of the Basic Education (SPAECE) in 1992. Several other states and municipalities have taken similar initiatives since then. In 2007, 14 out of 27 states had their own assessment systems (LOPES, 2007).

The coexistence of SAEB with statewide assessments, and years later, with *Test Brazil* makes the initial emphasis on diagnostic purposes in the use of assessment results lose strength in face of the tendency to see this use as information for accountability policies, which

leads to the recognition of two new generations of assessment of basic education in Brazil. As we will see in the following sections, such assessments involve the dissemination of test results by systems and / or schools and, in the case of the third generation, the establishment of awards linked to student results.

## **Educational assessment and accountability**

In the educational field, the assessments that support accountability policies have increasingly operated within a framework that combines democratic management of education, assessment and accountability. The underlying definition of democracy here is based on two guiding principles. On the one hand, there is the participation which happens, to a large extent but not exclusively, through the election process and party system. To this end, every citizen should have basic political rights: freedom of expression, association, freedom to vote and run for public office. On the other hand, there is public contestation among various political actors, not only because of political competition, but also and especially as control of the rulers by the ruled. That is, rulers (as agents of popular sovereignty) are accountable for their acts and omissions in the exercise of public power” (CENEVIVA, 2005, p. 12).

These two ideals of democratic systems – participation and public contestation – correspond to two basic forms of accountability. The first one is the electoral process, which expresses vertical control over rulers. It is an instrument of political participation, a guarantee of popular sovereignty. Through periodic elections, it ensures the expression of the people’s preferences through mandates. The second form of accountability is the institutional control over the mandates, which guarantees public contestation and the continuous monitoring of the political representatives, elected or not, in the exercise of public power (CENEVIVA, 2005).

## Second generation of education assessments – accountability e curriculum: *Test Brazil*

Driven by the realization that the democratization of public power should go beyond voting, this form of accountability has been considered in the perspective of the improvement of state institutions, which involves at the same time the improvement of policies and government programs and an increase in transparency and accountability of public policy actions.

In recent years two relationships have gained importance: the relationship between the quality of government actions and the controls and incentives rulers and bureaucracy face; and the relationship between the strengthening of accountability mechanisms and the improvement of administrative practices.

Two mechanisms in particular have been indicated in the establishment of new forms of participation and social control over state actions: social control and the control of results. The introduction of mechanisms of social control and accountability of public administration for the performance of government policies and programs appears as a promise to replace a model in which prevails the bureaucratic control based on compliance with norms and procedures, without citizen participation with another model which establishes *a posteriori* control of the results of government actions, with the participation of society. This mechanism may also involve setting performance goals and indicators, as well as the direct assessment of the public goods and services offered (CENEVIVA, 2005).

Thus, the assessment of public policies and programs is given a prominent place as a means to measure their performance and be accountable to society. From this perspective, assessments appear directly linked to the performance of public administration, to the promotion of greater transparency and to the design of accountability mechanisms.

In the next sections, we shall examine how the evolution of educational assessments articulates with this perspective, within which we can identify the emergence of the second and third generations of large-scale assessments.

*Test Brazil* was implemented in 2005 in order to increase the information content of the assessment and its consequences on schools. Such test allows aggregating the notion of accountability to the diagnostic perspective (FERNANDES, GREMAUD, 2009). The justification for its implementation indicated the limitations of SAEB's sampling design to depict the specifics of municipalities and schools and to induce state and municipal public officials to design policies to improve teaching.

*Test Brazil* is a biennial assessment designed to produce information about the education offered by municipality and school, with the objective of assisting governments to make decisions about the allocation of technical and financial resources and the establishment of goals and implementation of pedagogical and administrative actions aimed at improving education quality. On the other hand, it is considered that this assessment can favor parents' and carers' pressure to improve the quality of education for their children, given that after the dissemination of results, they can claim measures for the school improvement.

The introduction of *Test Brazil* in 2005 and its repetition every two years allow comparing primary education schools over time. In its first edition, it assessed more than 3 million students in approximately 45,000 urban schools in 5398 municipalities. Therefore, it went much further than SAEB, which assesses a sample of 300,000 students on average.

*Test Brazil* was a census for urban schools in 2005 and 2007. In 2007, the minimum number of students in the grade assessed changed. It fell from thirty to twenty. This change was made to include about four hundred municipalities which had not participated in the assessment's first edition. In 2009, in the third edition, the universe evaluated expanded to include all rural schools that had at least twenty students in the grades assessed.

The results of *Test Brazil* 2007 became part of the *Indicator of Basic Education Development* (IDEB<sup>5</sup>), a reference for the definition of goals to be achieved gradually by public school systems by 2021. The basic principle of such indicator is that the quality of education involves that students learn and are promoted to the following grade. With IDEB, performance started being measured by *Test Brazil* and promotion by the *School Census*. Pass rates allow taking into account the average number of years students take to complete a grade.

The results of *Test Brazil* began to be widely disseminated and IDEB is currently the main indicator adopted by the Federal Government to establish educational goals to be achieved by schools and state and municipal systems. The central idea of the system of goals was to obtain greater commitment of systems and schools with the aim of improving educational indicators. It is assumed that a system of goals agreed upon between the Ministry of Education and state and municipal education departments serves to increase society mobilization in defense of quality education. From this perspective, it is noteworthy that Brazil has a decentralized educational system, with more than 5,000 school systems with autonomy to manage their schools (FERNANDES; GREMAUD, 2009).

The results of the first edition of *Test Brazil* were divulged in July 2006 through the main media and a newsletter made available on the Internet and sent to each participating school. Among other information, this newsletter showed the schools' results on a performance scale and the scores by school in municipal, state and federal systems.

While the media divulged rankings of schools, especially the best and worst results, the websites of the Ministry of Education and

INEP, emphasized that the innovation of *Test Brazil* was the feedback to schools on the results in order to collaborate with the planning of pedagogical actions (OLIVEIRA, 2011).

In 2009, months before the third edition of *Test Brazil*, INEP and the Ministry of Education distributed two publications to all public schools: *The Reference Matrix of Test Brazil and SAEB – Primary School* and *The Reference Matrix of SAEB – Primary and Secondary School*. Both publications commented on examples of items from previous editions. In the case of *Test Brazil*, however, for administrative reasons, the results were not disclosed to schools. They could be found only in tables of scores by school in IDEB 2009 (OLIVEIRA, 2011).

Since the Federal Government disclosed the national results of *Test Brazil* and state governments took initiatives in the same direction – for example, Minas Gerais, Ceará and Rio Grande do Sul – we have had experiences of second generation education assessment, characterized by innovations which incorporate the disclosure of results to allow comparisons not only between systems but also between schools.

The strategy of disseminating rankings through the media, albeit unofficial, along with the distribution to schools of the matrix of content and skills used in the preparation of Portuguese and mathematics tests, introduced real prospects of more direct influence on *what* schools do and *how* they do it.

In terms of accountability, however, *Test Brazil* and the use of its results for the composition of IDEB integrate a *lenient accountability* policy, since they merely set goals and disseminate student results by school and school system but do not couple awards or penalties with these results, as it is typical of *strong accountability* policies (HANUSHEK, 2004; HANUSHEK; RAYMOND, 2005). Such policies and their relationships with third generation educational assessments will be addressed in the next section.

**5-** The National Institute of Education Study and Research Anísio Teixeira (INEP), an organ linked to the Ministry of Education, created IDEB. For more detailed information, see <<http://www.inep.gov.br>> and FERNANDES, 2007.

### Third generation of education assessments – accountability and curriculum: Sao Paulo and Pernambuco state assessments

Several state and municipal basic education systems have developed their own assessment proposals – usually of a census of their schools – through biennial tests taken by students in the 4th and 8th grades of primary school and 3rd grade of secondary school. In comparison with the current proposals, there is great similarity in the designs adopted by the assessment systems, which tend to use SAEB and *Test Brazil* reference matrix in the preparation of test items<sup>6</sup>.

However, there are specifics in the educational assessments and the use of their results which illustrate the characteristics of the relationships between third-generation assessments, accountability policies and school curriculum.

#### Recent developments of Sao Paulo state assessment

Sao Paulo System for Assessment of Educational Achievement (SARESP) was established in 1996, with the following objectives:

- To support the Education Department decision making on educational policies;
- To monitor the performance of basic education students to provide information to all levels of the education system that support the training of education human resources; the reorientation of the pedagogical proposal of schools in order to improve it; the feasibility of linking assessment results with school planning, training and goal setting for every school project. (SAO PAULO, 1996, p. 7)

These explicit objectives indicate that the assessment had double orientation: to serve as a reference for policy making by the Education

Department, and to guide the construction of schools' pedagogical proposals and planning. Linking the assessment to the improvement of teaching quality, the implementation document shows that such quality depends on the one hand on the commitment of the education system managers and, on the other hand, on schools, and that these are held particularly accountable for student performance.

Targeted to teachers and other education professionals, the notion of accountability materialized in 2000 with the establishment of *Merit Bonus*, whose distribution took into account the results of the large-scale assessment<sup>7</sup>.

In 2007, the announcement of the *Plan of Goals* by Maria Helena Guimarães Castro, the Secretary of Education, and Governor Jose Serra highlighted the importance that the large-scale assessment would have for this management. The *Plan's* 5th goal was to increase the performance levels of primary and secondary education in national and state assessments by 10%. By setting this goal, the Secretary indicated the continuity of SARESP. And among the *10 Goals for a Better School* two goals stressed the role of large-scale assessment in the development of Sao Paulo state educational policy. The goals were divulged with the following wording:

#### Goal 8 - Assessment Systems:

- The external assessment of state schools (compulsory) and municipal (optional) will allow the comparison of SARESP results with the national assessments (SAEB and Test Brazil), and will serve as a criterion for monitoring the goals to be achieved by schools.

<sup>7</sup>- The *Merit Bonus* was established in the Government of Mario Covas (1999-2001) by Supplementary Law 891/00, and kept during the administrations of Alckmin (2001-2002 and 2003-2006). In the management of Jose Serra, the State Education Department instituted a new *Merit Bonus*, whose calculation is based on Sao Paulo State Education Development Index (IDESP). One of IDESP's criteria is student performance in SARESP Portuguese and mathematics tests. Such link strengthens the relationship between the bonus payment and the large-scale assessment. To find more about the reward for results, see Supplementary Law No. 1.078/08 and Resolutions SEE No. 21/09, 22/09, 23/09 and 26/09. For information on IDESP, see <<http://idesp.edunet.sp.gov.br>>.

<sup>6</sup>- See LOPES, 2007; SOUSA; OLIVEIRA, 2007.

- Participation of the entire system in *Test Brazil* (November 2007).
- Training teachers to use the results of SARESP in pedagogical planning of schools in February 2008.
- Dissemination of SARESP 2007 results to all schools, teachers, parents and students in March 2008.

#### Goal 9 - Management of results and Policy of Incentives:

- Implementation of incentives for good school management valuing school teams.
- SARESP 2005 and promotion rates in 2006 will be the basis for the goals set by school.
- Indicators such as teacher attendance and the stability of teams in schools will also be considered.
- Each school's goals will be set based on its reality, and schools should improve in comparison to themselves.
- Schools with poor performance will receive intensive educational support and special incentives to improve their results.
- The school teams who meet the new goals will earn incentives in the remuneration of professionals<sup>8</sup>.

These goals illustrate the importance attached to the results of large-scale assessments in the managements of Secretary Maria Helena Guimarães Castro (2007-2009) and Secretary Paulo Renato de Souza (April 2009-2010). Goals 8 and 9 support the conclusion that the goals set for SARESP in 1996 have remained to this day, showing that the assessment should serve both to be used by system managers and to guide planning and pedagogical work in schools.

The policy begun in the management of the Secretary Maria Helena has continued and shows concern with the appropriation of results by the system management agencies and by schools. One of the measures taken in such

management is the implementation of a unified curriculum, presented as a guide of teaching organization, guiding the assessment parameters. This curriculum unification is directly related to the changes implemented in SARESP since 2007, especially the adoption of IRT.

The analysis of the official curriculum and SARESP matrices shows the correspondence between the curriculum, the matrices and teaching materials available to teachers (since 2008) and students (since 2009), called Teacher Notebooks and Student Notebooks. These materials exhibit curriculum-based learning situations designed to guide and support teaching in the classroom.

#### Recent developments in Pernambuco state assessment

The introduction of an educational accountability system by Governor Eduardo Campos and by the Secretary of Education Danilo Cabral was a central aspect of Pernambuco state education policy. This system includes an annual Pernambuco System of Educational Assessment (SAEPE), the broad dissemination of its results, bimonthly assessments of students by grades and bimonthly monitoring of educational indicators of every state school through a computerized system.

SAEP was first performed in 2000 and repeated in 2005. But its results were not consolidated and released until 2007. Since 2008, it has been performed annually and its results have been included in the Pernambuco Basic Education Development Index (IDEPE).

The system also collects information on the socioeconomic and cultural background of students, teachers and management teams. Its main objectives are:

- to produce information on the degree of student mastery of skills and competencies considered essential, in each schooling period, not only for further study, but also for life in society;

8 - Available at: <[www.educacao.sp.gov.br](http://www.educacao.sp.gov.br)>. Accessed on Jan 15, 2010.

- to monitor student performance over time, as a way of continually assessing the pedagogical project of each school, allowing the implementation of corrective measures when necessary;
- to contribute directly to the adaptation of teaching practices to student needs diagnosed by means of the assessment tools;
- to link assessment results to incentive policies in order to reduce inequalities and increase the degree of school effectiveness;
- to compose IDEPE along with the promotion rates recorded by the *School Census*.

The central feature of this policy is to establish goals for each school and to grant the *Educational Performance Bonus* (BDE) to the schools which meet their goals. This way, students' grades in SAEPE proficiency tests are used along with IDEPE to define goals.

IDEPE considers both the scores of students in 4<sup>th</sup> and 8<sup>th</sup> grades of primary school and in the third year of secondary school in SAEPE assessment in Portuguese and mathematics and the average student promotion rate measured by the *School Census*. Thus, to raise their IDEPE index, schools have to improve promotion rates and student proficiency scores in SAEPE.

The goals to be achieved by students are agreed on by the Education Department and schools. Each school has its own goal, calculated according to its particularities. Moreover, goals are consistent with the phase of the school: for each grade assessed, a goal is set for Portuguese language and another for mathematics. The difference between the IDEPE index used as a reference and the one expected is the goal for each discipline and grade assessed, and the average score actually achieved shows the percentage achieved by the school in relation to its goals.

2008 goals were set so that schools reached the midpoint between their initial

IDEPE in 2005 and the 2009 goal. These goals vary between groups with *poor performance*, *intermediate performance* and *high performance*. However, since 2009, goals have been the same for all schools within a given group.

SAEPE is a third generation assessment which supports strong accountability mechanisms whose most consistent expression is BDE. The bonus ranges from 50% to 100%: state schools which achieve an overall index below 50% do not receive the bonus; schools which reach 50% of the goal receive half the bonus; from there, the value is proportional to the percentage of goal achieved.

The Education Department also encourages the recognition of the teachers who remain in a single school. To this end, the bonus calculation takes into account the proportionality of goal achieved starting at 50% and the teacher's assignment to the same school for at least six months of the reference year. BDE is collective, since all employees assigned to and working in a given school are entitled to the bonus. It is also proportional to the salary and the percentage of goal reached.

In Pernambuco, the maximum value earned by each public servant is not defined *a priori*. While in Sao Paulo this value reaches a maximum at 2.4 salaries, in Pernambuco only the total amount allocated by the government for bonus payment is fixed. Also, the condition for public servants to have access to the bonus is that the school has one of the grades tested by SAEPE (4<sup>th</sup> and 8<sup>th</sup> grades of primary school and 3<sup>rd</sup> grade of secondary school). Thus, schools that do not offer any of the types of education tested do not have access to the bonus. However, teachers who work in the untested types of education, such as youth and adult or early childhood education, have access to the bonus if the school offers any of the grades tested. For schools that do not reach goals, the legislation provides technical, pedagogical and structural support, so that they meet the BDE criteria in the following school year.

Besides school scores, the assessment website presents the matrices used to prepare the test items, with detailed explanations of the descriptors, knowledge and competencies expected from each grade of primary and secondary education, and result newsletters containing contextual analysis with nationwide information and information by state and municipality.

Therefore, SAEPE is an assessment that can guide the curriculum and what students should learn at each phase of the school cycle.

### **Large-scale assessment and curriculum: What does research say?**

Second and third generation assessments, associated with the introduction of accountability policies based on symbolic and material consequences, are intended to create incentives for teachers to strive to make students learn. However, national and international evidence shows that especially the use of results of third-generation assessments to inform strong accountability policies may involve risks to the school curriculum. One of them is the condition known as *teaching for the test*, which occurs when teachers focus their efforts mainly on the topics that are evaluated and overlook important aspects of the curriculum, including non-cognitive ones.

It is difficult to disagree with the claim that large-scale assessments deal with a narrow view of the school curriculum in face of what schools propose as objectives for the education of their students. Also complex is the use of standardized tests to measure school objectives related to non-cognitive aspects.

The problem stems from the fact that curricula have multiple goals, while the result measures used by large-scale assessments typically aim at cognitive objectives related to reading and mathematics. This is not exactly a limitation of the assessment, but it demands attention to risks related to the narrowing of the curriculum, which can arise

when there is a misreading of the pedagogical meaning of assessment results.

The studies on this issue in Brazil are still limited and very recent. Even so, they collaborate to understand how schools and education departments interpret and articulate the relationships between three generations of large-scale assessments and the curriculum.

The results of a study conducted by John Luiz Horta Neto (2006) on the use of data from SAEB 2000 by the Federal District Education Department to guide the planning of the school system help illustrate the low influence of first generation assessments on the context of educational administration. The survey indicated that although the managers of the Education Department defend the importance of SAEB, they have little knowledge of and hardly use the data produced by the assessment in the management processes, mainly due to limitations to understand the results produced. While this research demonstrates SAEB's poor capacity to affect the educational management and school activities, it contributes to understanding the emergence of a second generation of educational assessments which, like *Test Brazil*, allow schools to see themselves in the results produced.

Ana Paula Oliveira (2011) investigated to what extent the Federal District Education Department has used the results of *Test Brazil* 2007 to support the management of the school system. This issue was also investigated in two schools of the Federal District, which had the highest and lowest IDEB indexes. On the one hand, education managers showed that they know *Test Brazil*, also from its technical point of view, and see it as a tool that establishes standards of education quality that should be met by schools. From this perspective, managers stated that the best results obtained by some schools in the assessments are seen by other schools of the system as an indication of the quality of the work they are developing. On the other hand, the Education Department teams see *Test Brazil* as an initiative that allows

unifying the teaching-learning process because it discloses what is being taught and learned in all schools throughout the country.

In addition, teachers from both schools point out that *Test Brazil* contributes to unifying teaching in their institutions. This is seen as a positive fact, since the curriculum unification could contribute to make students across the country have access to the same knowledge, regardless of where they live and attend school. Thus, teachers and managers perceive *Test Brazil* as a reference for the possible implementation of a national curriculum.

Coordinators and teachers of the schools surveyed by Oliveira (2011) also stated that they seek to redefine the content in order to teach what is assessed by *Test Brazil*. Because the assessment is usually performed before the end of the school year, schools accelerate teaching to enable students to answer the tests, in order to ensure a good performance result for the school. Still in view of what the literature calls *teaching for the test*, teachers say they have incorporated the practice of preparing students to get used to the texts, the commands and the length of *Test Brazil* reading tests.

Studies that examined the assessment implementation in Sao Paulo state highlight elements which illustrate the changes in the curriculum<sup>9</sup>. The earlier studies tend to indicate little effect of SARESP on school everyday life, although they highlight reactions of mistrust and resistance to the system by education professionals (OLIVEIRA, 1998; ESTEVES, 1998; FELIPE, 1999; KAWAUCHI, 2001). These studies, which focused on the possible impact of SARESP on schools, considering in particular teachers' opinions and reactions, tend not to identify any influence of its results on the school curriculum.

The studies conducted since the late 2000s have found evidence of the effects of SARESP in the school setting. In the findings of a survey conducted in a state school,

Lilian Rose Freire (2008) points out some uses of SARESP results, such as: i) use in the calculation of students' bimonthly grades; ii) reproduction of questions in the unified test created by the school in order to *train* students for the assessment; iii) Portuguese teachers' use of SARESP guidelines of essay correction to guide students in school writings, which may lead to the improvement of existing practices; iv) encouraging the participation of students in SARESP tests, through grading that is considered in the bimonthly grade.

The information presented in this research indicates that the meaning given to SARESP and its results in the context of that school is not associated with the idea of an assessment that brings support and guidance to redesign schoolwork. It seems that, except for the use of criteria for essay correction, the interaction with the SARESP is more instrumental in the sense that teachers implement initiatives that may help students achieve better results, such as teaching how to fill out templates and using tests with questions similar to those of SARESP tests.

A study conducted by Paulo Henrique Arcas (2009) focused on the possible impact of SARESP on school assessment, seeking to identify assessment patterns and trends after its implementation. To this end, he carried out a study in a regional directorate of education in Sao Paulo metropolitan area, seeking the views of teacher coordinators. Through questionnaires and interviews, he analyzed how they saw SARESP and how they gradually built their opinions about it until 2007. Such views allowed identifying how the large-scale assessment had impacted on evaluations and on the school curriculum.

The discourse of the teachers-coordinators interviewed evidenced a tendency of acceptance of SARESP, although the system was initially viewed with suspicion. There is evidence that the data produced in the assessment are analyzed and discussed during school planning at the beginning of the year and during its redesign early in the second

<sup>9</sup>- The considerations on SARESP are based on SOUSA, ARCAS, 2010.

semester. The results of the school and its classes are analyzed and guide the schoolwork, defining the skills, competencies and content to be taught.

Therefore, it can be stated that SARESP has been increasingly more present, has influenced practices, set goals, established directions and guided the pedagogical work. Another important revelation about the implications of such an assessment in the school context is that it has impacted on the assessment practices developed at school. Evidence of the research mentioned above has shown that the evaluation of learning conducted at school uses the large-scale assessment as a reference.

However, what we did find is that SARESP, by serving as a reference for the assessment practices undertaken in schools, ended up reinforcing traditional practices of learning assessment. [...] SARESP strengthens the application of tests, in most cases in order to simulate the application of the external assessment. It is assumed that this way students are being prepared. (ARCAS, 2009, p.120)

The centrality SARESP has acquired in the organization of schoolwork by guiding assessment practices allows stating that schools have increasingly appropriated this large-scale assessment. In this sense, by guiding assessment procedures, SARESP has induced an emphasis on the application of mock tests and examinations as means to prepare students to do well on the state assessment.

### **Final considerations**

This paper has aimed to characterize designs, objectives and uses of the results of experiences of assessment of the current basic education in the country, in view of its relationships with the school curriculum. The study of three generations of assessments of basic education allowed identifying the

second and third generation assessments – i.e., which articulate, respectively to weak and strong accountability policies – as those with the most significant consequences for the school curriculum.

Although it is fairly early to make more consistent statements on these assessments and their impact on the school curriculum, early studies show that the new design introduced by the second generation produced results that serve as indicators of the curriculum components which are reaching students, and of those which are not. From this perspective, in schools and education departments, such type of assessment seems to be strengthening the alignment of the curriculum taught with the curriculum assessed.

In fact, the brief review of research on the topic presented here has brought contributions, sometimes recurring ones, to the understanding of the influence of assessment on the curriculum. What these studies evidence together is the importance that second and third generation assessments have gained in the design of education policies and, consequently, their potential to direct *what, how* and *what* for to teach.

According to the findings of this study, the use of standardized tests in the context of assessments related to accountability policies with weak and strong consequences for schools – particularly the strong ones – exacerbates the concerns of principals and teachers with preparing students for tests and for the kind of activity present in them.

Moreover, studies have also shown that first generation large-scale assessments, i.e., assessments *without consequences*, minimize these problems, because principals and teachers find themselves less *threatened* by the assessment and can use it or not with greater freedom. In this context, however, these professionals rarely feel compelled to give an account of the results of their work or have motivation to learn about assessment results and to take them into account in their educational and pedagogical work.

In summary, this study has discussed the risks and potential of second and third generation assessments to the school curriculum. On the one hand, it has pointed out that assessments related to accountability policies risk exacerbating the concerns of principals and teachers with preparing their students for tests, which leads to a narrowing of the school curriculum. It has also pointed out the implications for learning evaluation when schools begin to organize it in function of the type of test used by large-scale assessments.

On the other hand, this study has indicated the potential of second and third generation assessments to provide a more informed discussion about the school curriculum, in terms of the essential reading

and mathematics skills that have not been guaranteed to all primary and secondary school students yet.

Thus, the challenge seems to be the compatibility of the objectives, design and uses of the results of three generations of large-scale assessments in order to provide an informed discussion on the specifics of Portuguese and mathematics that should be learned by all students, as well as a clearer definition of what these students should have learned at the end of each cycle in these two areas of school knowledge. That is why there is a need to increase the body of research that contributes to understanding the impact of the new generations of educational assessments on the school curriculum.

## References

- ARCAS, Paulo Henrique. **Implicações da progressão continuada e do SARESP na avaliação escolar: tensões, dilemas e tendências.** Tese (Doutorado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2009.
- BONAMINO, Alicia. **Tempos de avaliação educacional: o SAEB, seus agentes, referências e tendências.** Rio de Janeiro: Quartet, 2002.
- BROOKE, Nigel. O futuro das políticas de responsabilização no Brasil. **Cadernos de Pesquisa**, São Paulo, v. 36, n. 128, p. 377-401, 2006.
- CARNOY, Martin; LOEB, Susanna. Does external accountability affect student outcomes? A cross-state analysis. **Educational Evaluation and Policy Analysis**, v. 24, n. 4, p. 305-331, 2002.
- CENEVIVA, Ricardo. **Democracia, accountability e avaliação: a avaliação de políticas públicas como mecanismo de controle democrático.** Dissertação (Mestrado) – Fundação Getúlio Vargas, São Paulo, 2005.
- ESTEVES, Maria Eunice de P. **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP: uma ação planejada.** Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1998.
- FELIPE, Jesse Pereira. **Uma análise crítica do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo: SARESP.** Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1999.
- FERNANDES, Reynaldo. **Índice de Desenvolvimento da Educação Básica (Ideb).** Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2007.
- FERNANDES, Reynaldo; GREMAUD, Amaury. Qualidade da educação básica: avaliação, indicadores e metas. In: VELOSO, Fernando et al. (Orgs.). **Educação básica no Brasil: construindo o país do futuro.** Rio de Janeiro: Elsevier, 2009.
- FREIRE, Lilian Rose S. Carvalho. **SARESP 2005: as vicissitudes da avaliação em uma escola da rede estadual.** Dissertação (Mestrado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2008.
- FREITAS, Dirce Nei Teixeira de. **A avaliação da educação básica no Brasil: dimensão normativa, pedagógica e educativa.** Campinas: Autores Associados, 2007.
- GATTI, Bernadete A. Testes e avaliações do ensino no Brasil. **Educação e Seleção**, São Paulo, n. 16, p. 12-21, 1987.
- HANUSHEK, Eric A. United States lessons about school accountability. **CESifo DICE Report 4**, 2004.
- HANUSHEK, Eric A.; RAYMOND, Margaret E. Does school accountability lead to improved student performance? **Journal of Policy Analysis and Management**, v. 24, n. 2, p. 297-327, 2005.
- HORTA NETO, João Luiz. **Avaliação externa: a utilização dos resultados do SAEB 2003 na gestão do sistema público de Ensino Fundamental no Distrito Federal.** Dissertação (Mestrado) – Faculdade de Educação, Universidade de Brasília, Brasília, DF, 2006.
- KAWAUCHI, Mary. **SARESP e ensino de história: algumas questões.** Dissertação (Mestrado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2001.
- LOPES, Valéria Virgínia. **Cartografia da avaliação educacional no Brasil.** Tese (Doutorado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2007.
- OLIVEIRA, Ana Paula M. **A Prova Brasil como política de regulação da rede pública do Distrito Federal.** Dissertação (Mestrado) – Faculdade de Educação, Universidade de Brasília, Brasília, DF, 2011.
- OLIVEIRA, Duzolina Alfredo F. de. **Uma avaliação política do SARESP.** Dissertação (Mestrado) – Universidade Estadual de Campinas, Campinas, 1998.
- SÃO PAULO. Secretaria de Estado de Educação. **SARESP: documento de implantação.** São Paulo: FDE, 1996.

SOUSA, Sandra Zákia; ARCAS, Paulo Henrique. Implicações da avaliação em larga escala no currículo: revelações de escolas estaduais de São Paulo. **Educação: Teoria e Prática**, Rio Claro, v. 20, n. 35, p. 181-199, jul./dez. 2010.

SOUSA, Sandra Zákia; OLIVEIRA, Romualdo Portela de. **Sistemas de avaliação educacional no Brasil**: características, tendências e uso dos resultados. Relatório de pesquisa apresentado à FAPESP, São Paulo, jul. 2007.

ZAPONI, Margareth; VALENÇA, Epifânia. **Política de responsabilização educacional**: a experiência de Pernambuco. Abr. 2009. Disponível em: <www.abave.org.br>.

*Received on August 8, 2011*

*Approved on September 20, 2011*

**Alicia Bonamino** is an associate professor at Pontifical Catholic University of Rio de Janeiro and coordinator of the Laboratory of Education Assessment (LAEd).

**Sandra Zákia Sousa** holds a Ph.D. degree in education, is a professor at the Faculty of Education, University of Sao Paulo, where she works in the graduate studies program –state, society and education. She also works at education Master's program of the University of Sao Paulo City. She has publications mainly in the education assessment field.